RESEARCH TRAJECTORIES IN MATHEMATICS EDUCATION

A Dissertation

by

JOSEPH S. WEBSTER

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Roger Howe
Committee Members,	Mary Margaret Capraro
	Yeping Li
	Christopher Thompson
Head of Department,	Claire Katz

May 2023

Major Subject: Curriculum and Instruction

ABSTRACT

The main part of this dissertation study consists of a survey of the types of articles (e.g., articles based on qualitative research, articles based on quantitative research, theoretical articles, etc.) that appear between 2016 and 2021 in five of the leading mathematics education journals (N = 1,317). In addition, for the articles that carry out quantitative research (N = 115), and the studies examined in one of the articles that reports on a meta-analysis (N = 22), the author examines the variety of variables studied, along with the measures used to study these variables and the methods of analysis used. The author hopes that the results of this work can shed some light on the current nature and progress of the domain of Mathematics Education Research (MER). In pursuit of this goal, a theoretical framework labeled Action Domain Theory (ADT) is developed. The basic premise of ADT is that, to any kind of organized effort (referred to as a domain), such as MER, there is a goal (or set of goals) that drives all actions taken in the domain. In the context of the domain of MER, the author hopes that these explorations can provide some clues pointing towards the goals of the domain, and perhaps also the extent to which the current research is moving towards some of these goals.

DEDICATION

This dissertation is dedicated to my father, without whom I would never have made it through many of the obstacles which life has presented. It was his passion for mathematics and truth which first sent me on this journey, and it was his unwavering belief in me which kept me on the path. The insights, lessons, and truths that he has taught me throughout my life have both inspired this work and made it possible to complete.

This work is also dedicated to Annie, whose passing I have mourned deeply during my Ph.D. studies and whose memory I cherish fondly as I complete this journey. It is her continual love which I have held to so dearly from a very young age.

I would also like to dedicate this work to my chair, Dr. Howe. It was Dr. Howe’s unfailing compassion and support which have allowed my work to progress and come to completion. His belief in me, not only as a doctoral student, but as a human being, has sustained me and has shown me what it truly means to be a teacher.

Finally, this work is dedicated to my family and friends who have supported me both in my work and my life; to my committee members for their continued compassion, patience, and support; and to God who is with me always, in every valley and on every mountaintop.

ACKNOWLEDGEMENTS

I thank Dr. Howe for the many phone conversations (even nights, weekends, and times around the holidays), continual insights, compassionate guidance, and for always believing.

I thank Dr. Yeping Li for continually guiding me in the right directions and for his dedication to my producing a high quality product. His guidance has helped me to learn a great deal about the domain of mathematics education and he often assisted me in finding the proper literature relevant to my work. Additionally, I am eternally grateful for the compassion he has shown me in recent years.

I thank Dr. Mary Margaret Capraro and Dr. Robert Capraro for their compassion and dedication to guiding me to acquire the skills required of a researcher. Their help and teachings have greatly increased my knowledge and understanding of education, learning, research, and teaching.

I also thank Dr. Thompson for his encouragements, swift reviews of my work, and for helping me to increase my knowledge and understanding of statistics and meta-analytic methods.

I acknowledge the contributions of the many authors cited in this dissertation, all of whose efforts I was inspired by and without whom my work would not be possible.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Howe, Capraro, and Li of the Department of Teaching, Learning, & Culture and Professor Thompson of the Department of Educational Psychology.

The inter-rater reliability data coding was provided by Professor Howe.

All other work conducted for this dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University and a graduate assistantship through the Department of Teaching, Learning, & Culture as advised by Professors Howe and Li.

NOMENCLATURE

Throughout this dissertation the author will frequently use the following abbreviations:

ADT	Action Domain Theory
EG	Explicit Goal
IG	Implicit Goal
KOM	Kompetencer Og Matematik
LT	Learning Trajectory
MEL	Mathematics Education Learning
MER	Mathematics Education Research
MET	Mathematics Education Teaching
NAEP	National Assessment of Educational Progress
NCTM	National Council for Teachers of Mathematics
NGA	National Governors Association
NMAP	National Mathematics Advisory Panel
NRC	National Research Council
OECD	Organization of Economic Cooperation and Development
ORQ	Overarching Research Question
PISA	Programme for International Student Assessment
RQ	Research Question
RT	Research Trajectory
TIMSS	Trends in International Mathematics and Science Study
USDOE	United States Department of Education

DEFINITIONS

Throughout this dissertation the author will frequently use the following terms:

Action	A physical or mental motion performed with intention.
Article	A peer-reviewed *paper*.
Domain	An organized effort, consisting of an individual or a group of individuals along with all *actions* taken by these individuals.
Empirical Article	An *article* which contains the results of a *study*.
Explicit Goal	*Goals* which are explicitly stated and communicated (Kawada et al., 2004; Klinger, 1977; Schultheiss & Brunstein, 2010).
Goal	An object of a *domain's actions*; that which is causally linked to the intention of *actions.*
Implicit Goal	*Goals* which are not clearly stated and communicated, but instead can be hypothesized by examining the *actions* within a *domain* (Kawada et al., 2004; Klinger, 1977; Schultheiss & Brunstein, 2010).
Paper	An item published in a journal and given a title.
Study	Reported in an *article* and characterized by an analysis of collected data.
Trajectory	The path that a *domain* traverses in its efforts to reach a *goal*, consisting of all *actions* taken during these efforts.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

INTRODUCTORY OVERVIEW

> In particular, because there are no definitive answers, one should certainly be wary of anyone who offers them. More generally, the main goal for the decades to come is to continue building a corpus of theory and methods that will allow research in mathematics education to become an ever more robust basic and applied field.
>
> —Schoenfeld, 2000, p. 649

This chapter serves as a brief, introductory overview of this dissertation. Presented within this chapter are introductory descriptions of the problem, the theoretical framework, the problem statement, purpose of the study, a discussion of what is *not* the purpose of the study, the research questions, the methodology of the study, the significance of the study, the limitations of the study, and the organization of the study. The background literature and theoretical framework are described in more detail in Chapter II, the methodology is described in more detail in Chapter III, the results are presented Chapter IV, and the results are discussed and interpreted in Chapter V.

Background of the Problem

A number of projects have been conducted in pursuit of the goals of mathematics education. The goals of Mathematics Education Learning (MEL) have been examined in recent decades by TIMSS (Mullis et al., 1997), PISA (OECD, 1999), the Danish KOM Project (Niss, 1999), the NCTM Principles and Standards for School Mathematics (NCTM, 2000), the NRC Study Committee that produced Adding It Up (NRC, 2001), and the NMAP report (USDOE, 2008). The goals of Mathematics Education Teaching (MET) has been examined extensively by Deborah Ball and Heather Hill in their ongoing Mathematical Knowledge for Teaching project (Ball et al., 2008). The goals of Mathematics Education Research (MER) have been discussed by Alan Schoenfeld, who described them as comprising both pure and applied pursuits of knowledge of mathematics teaching and learning (Schoenfeld, 2000).

Reviews have examined the landscape of MER in the last several decades (Begle, 1979; Hart et al., 2009; Inglis & Foster, 2018; Li et al., 2020). In 1979, Edward Begle conducted a survey of the literature in MER. Begle (1979) examined various aspects including: the goals of mathematics education; the variables measured, measures used, and analytic methods employed in MER; and even the populations studied. Other reviews have examined the prevalence of research methodologies in MER (Hart et al., 2009; Li et al., 2020). However, there have not been attempts to examine the progress of research in MER in the context of the goals of MER; neither has a theoretical framework been developed for conducting such studies.

Theoretical Framework

The theoretical framework *implemented* in this study is also *developed* in this study. The framework, called Action Domain Theory (ADT), can be conceptualized as a generalization of Learning Trajectory (LT) theory. In summary, ADT posits that, given an organized effort (referred to as a *domain*), there is a *goal* (or set of goals) that drives all actions taken in the domain. Furthermore, ADT posits that goals and measurement are strongly linked, such that a measurement implicitly characterizes the goal behind the measurement. Moreover, ADT presents the notion of *trajectories*, that are aimed at some goal(s) of the domain of focus. ADT applied to MER puts forward Research Trajectory (RT) theory, which can be thought of as LT theory applied to MER rather than MEL. In short, an RT is a trajectory of some particular line of research (comprised of individual studies) toward some Implicit Goals (IGs). These IGs may agree with some of the Explicit Goals (EGs) of MER or may not. ADT posits that the progress of a domain towards its goals (whether explicit or implicit) can be determined by summarily considering the progress of the individual trajectories within the domain; therefore, this dissertation study is an attempt to explore the current RTs in MER in order to gain some insight into the progress of MER toward some of its EGs.

Problem Statement

There has been, and is currently, a gap in the research literature in MER regarding studies that seek to explore (and in some sense measure) the trajectories, and progress of research in the context of explicit goals of MER, as well as a gap regarding a suitable theoretical framework for conducting these kinds of explorations.

Purpose of the Study

The purpose of this study is two-fold:

1. to explore, and attempt to measure the progress of, RTs currently present in MER in the context of some explicit goals of MER; and
2. to develop and demonstrate a theoretical framework for conducting such an exploration.

What is Not the Purpose

This dissertation is *not* an attempt to determine the specific RTs currently present in MER, nor is it an attempt to determine specific details of the RTs currently present in MER; such tasks, as well as their feasibility, are reserved for future research. Rather, this dissertation is an attempt to explore (from a bird's eye view) what the current RTs in MER might be and whether they are moving:

1. toward some of the explicit goals of MER; and
2. in an organized manner toward these goals.

Research Questions

The overarching research question guiding this study is: What are some goals of the domain of MER and based on publications between 2016-2021 in five leading mathematics education journals, what progress is being made toward those goals? From this overarching question, a number of sub-questions are posed:

RQ1. What are the percentages of the types of articles published within the domain of MER?

- RQ1a. Are there significant changes in these percentages across the selected time period?
- RQ1b. Are there significant changes in these percentages across the selected journals?
- RQ1c. Applying ADT, what might these percentages indicate about the RTs in the domain of MER?

RQ2. Focusing specifically on articles implementing quantitative methodologies: what variables are measured, what measures are used, and what analytic methods are used?

RQ2a. Do the EGs of the researchers match their IGs?

RQ2b. Do any EGs of the domain of MER match the EGs or IGs of the researchers?

RQ2c. Applying ADT, what possible conclusions can be drawn about the RTs in the domain of MER?

RQ3. Focusing specifically on articles implementing meta-analytic methodologies: what variables are measured, what measures are used, and what analytic methods are used across research studies selected for inclusion in the meta-analyses?

RQ3a. Do the EGs of the researchers match their IGs?

RQ3b. Do any EGs of the domain of MER match the EGs or IGs of the researchers?

RQ3c. Applying ADT, what possible conclusions can be drawn about the RTs in the domain of MER?

RQ3d. How do these results compare with the results of Analysis 2?

The methodology, analyses, and organization of this dissertation were all designed around these questions and with the intent to provide evidence that might help provide some insight into what the answers to these questions might be.

Methods of the Study

This dissertation study is primarily exploratory in nature. RQ1 is addressed using a mixed-methods approach by coding categorical data, calculating counts and percentages, and running statistical analyses on these data. RQ2 and RQ3 are addressed using an exploratory, qualitative approach by coding for themes according to a coding scheme derived from the recent literature in MER. Each of these and the ORQ are examined in light of the novel theoretical framework, called ADT, developed in Chapter II for the purposes of studies of the kind within this dissertation.

Significance of the Study

This study can make several claims to significance, including its provision of the following to the domain of MER:

1. a consolidation, statement, and discussion of some of the consensus EGs of MER, accomplished via several substantial searches of the literature;
2. the development and demonstration of a novel theoretical framework that fills a gap in the research literature;
3. a robust survey of the literature from the most recent six years, across five of the leading journals in MER, that corroborates, builds on, and extends previous surveys of a similar type; and
4. possibly the first exploration of its kind, examining and attempting to measure the current progress of research in MER in the context of the aforementioned EGs.

Limitations of the Study

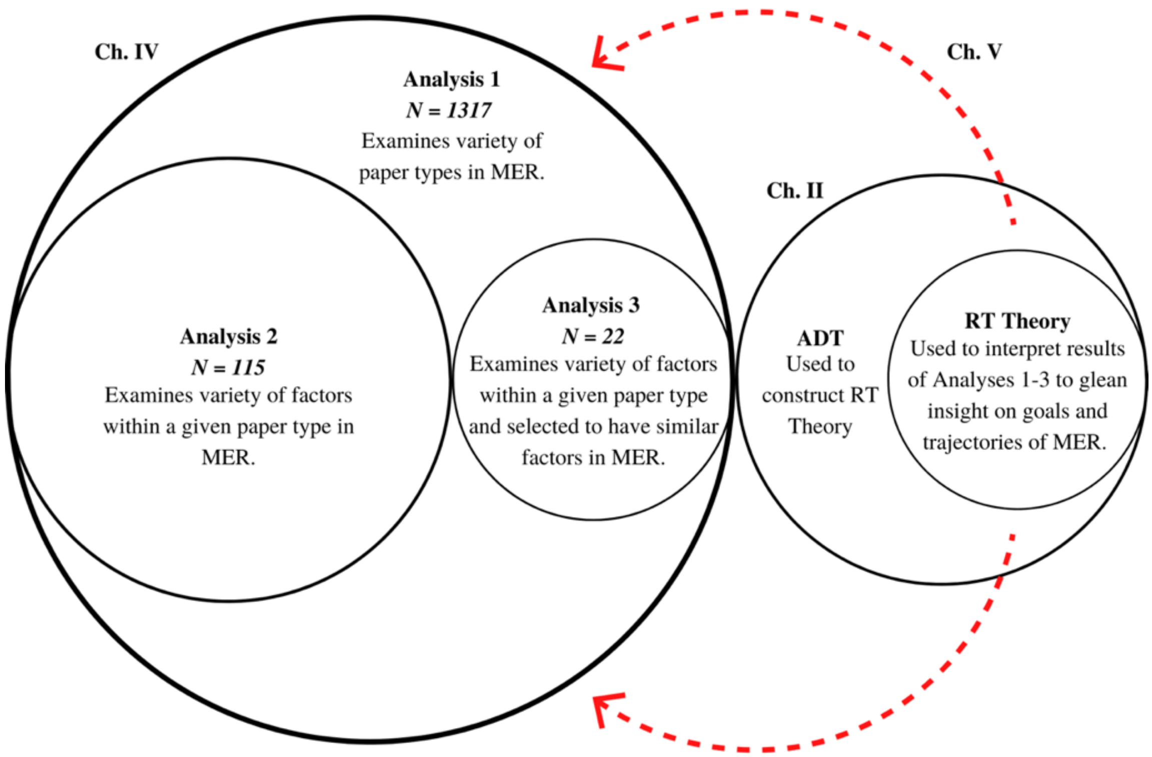
This study is limited insofar as it is a pioneer within the topic of measuring the progress of research in MER. Due to the exploratory nature of this study, and due in part to its novelty, it is limited to a "bird's eye view" perspective in regard to examining the current RTs in MER and in regard to attempting to determine the IGs of these RTs. The study is also marginally limited due to the relatively short date range chosen for data collection, which can be attributed to feasibility within the scope of this dissertation. The choice of journals to survey also constitutes a limitation because each journal has its own focus and selection criteria. Even the restriction of the survey to journal articles constitutes a limitation because books, theses and other literature may deal with issues not well investigated in journal articles.

Organization of the Study

Chapter II presents the pertinent background literature on the topics of domains, goals, research methods, and prior related reviews of MER; Chapter II also presents and develops the theoretical framework of ADT and its application to MER, called RT theory. Chapter III presents the methodology implemented for approaching RQ1, RQ2, RQ3, and the ORQ that are addressed by Analysis 1, Analysis 2, and Analysis 3. Chapter IV presents the results of Analysis 1, Analysis 2, and Analysis 3. Lastly, Chapter V addresses the research questions in light of a discussion of the results from Chapter IV; Chapter V ends with some final considerations for researchers.

Figure 1.

Diagram: overview of the organization of this study.



More specifically, Analysis 1, 2, and 3 explore the variety of features currently present in the research of MER and examines the variety of RTs currently present in MER. The results of these explorations are then examined in Chapter V using ADT (in particular, RT theory) as a lens for attempting to derive insight into the RTs of MER. Figure 1 provides a diagram demonstrating the flow of the study and some reasons for the various stages and developments involved throughout this study.

CHAPTER II

BACKGROUND AND THEORY

> First of all, however, the theoretician should be the conscience of the team by watching its goals even if they have not been made explicit and were only formulated as a result of watchfulness and as a means of warning.
>
> —Freudenthal, 1977, p. 177

Background Literature

In this section, the author examines the goals of Mathematics Education Learning (MEL), Mathematics Education Teaching (MET), and Mathematics Education Research (MER); the concepts of measurement and research methodologies in MER; and prior reviews of research in MER. First, the goals of MEL, MET, and MER are gathered from the literature, described, compiled, and relationships between them are discussed (see Table 1 and Figure 2). Next, the two main research methodologies in MER (qualitative and quantitative) are reviewed, as well as the combination of the two (mixed-methods). Finally, the prior reviews that have been conducted that examine and explore the landscape of MER are discussed.

Goals

There are two types of goals that the author will examine in this dissertation study. The first type of goal is the Explicit Goal (EG). These are goals that are explicitly stated and communicated. Klinger (1977), Kawada et al. (2004), and Schultheiss and Brunstein (2010) provide discussions of this type of goal. Such goals are called "explicit" because they are clearly stated and communicated, and they are fairly specific.

The second type of goal is the Implicit Goal (IG; sometimes referred to as an "implicit motive"). These are goals that are not clearly stated and communicated, but instead can be inferred or hypothesized by examining the actions taken by organized groups, or by individuals (Kawada et al., 2004; Klinger, 1977; Schultheiss & Brunstein, 2010). In the following sections, the author will gather the EGs of MEL, MET, and MER from the literature and organize them for use throughout this dissertation study. The results are summarized in Table 1.

Table 1.

Some of the EGs of MEL, MET, and MER.

Learning	Teaching	Research
L1 Symbols and Formalism	Content Knowledge	Basic/Pure
L2 Procedural Fluency	T1 Common Content	R1 Not directly Relevant
L3 Conceptual Understanding	T2 Horizon Content	R2 Directly Relevant
L4 Thinking Mathematically	T3 Specialized Content	R3 School Relevant
L5 Problem Solving	Pedagogical Knowledge	Applied
L6 Problem Posing	T4 Content and Students	R4 Experimental
L7 Logic and Reasoning	T5 Content and Teaching	R5 Normal Setting
L8 Communication	T6 Content and Curriculum	R6 Advocacy and Adoption
L9 Productive Disposition		
L10 Aids and Tools		

Goals of MEL

Recent attempts to uncover what goals are important in the development of student learning of mathematics include the NCTM Standards (NCTM, 2000), The Danish KOM Project (Niss, 1999), PISA (OECD, 1999), the NRC Study Committee that produced Adding It Up (NRC, 2001), the NMAP (USDOE, 2008), and TIMSS (Mullis et al., 1997). It is worth noting that the Common Core State Standards for Mathematics (CCSSM) produced by the National Governors Association (NGA) combined the work of NCTM and the NRC Study Committee in an effort to produce standards that could be applied to all U.S. states (NGA, 2010). The NGA acknowledged at that time that the standards varied greatly across the U.S. states and intended CCSSM to contribute to reducing that variation. These projects were all multifaceted, but each addressed the goals of the domain of MEL. Table 1 summarizes the results of these projects.

The NCTM *Principles and Standards of School Mathematics* (NCTM, 1989; NCTM, 2000) presented the following to be the general goals of MEL: communication, representation, connection, problem solving, and reasoning and proof. The Danish KOM Project (Niss, 1999) presented the following to be the important general goals of the domain of MEL: communication, representation, symbols and formalism, modeling, aids and tools, mathematical thinking, problem handling, and reasoning. PISA (OECD, 1999) presented the following to be the general goals of MEL: mathematical thinking; mathematical argumentation; modelling; problem posing and solving; representation; symbolic, formal, and technical; communication; and aids and tools. Adding It Up (NRC, 2001) presented the following to be the general goals of MEL: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. TIMSS (Mullis et al., 1997) presented the following to be the general goals of MEL: knowing, using routine procedures, investigating and problem solving, mathematical reasoning, and communicating. The NMAP (USDOE, 2008) presented the following to be the general goals of MEL: conceptual understanding, computational (procedural) fluency, and problem solving.

By combining and synthesizing these various goals, it is possible to compile a list of the EGs of MEL. Many of these frameworks have commonalities. Leveraging these commonalities, the following themes emerge that the author will use in this dissertation as the EGs of MEL: symbols and formalism (Niss, 1999; OECD, 1999); procedural fluency (Mullis et al., 1997; NRC, 2001; USDOE, 2008); conceptual understanding (Mullis et al., 1997; NRC, 2001; USDOE, 2008) which also contains representations (NCTM, 2000; Niss, 1999; OECD, 1999), modelling (Niss, 1999; OECD, 1999) and connection (NCTM, 2000); thinking mathematically (Niss, 1999; OECD, 1999); problem solving (NCTM, 2000; Niss, 1999; OECD, 1999; USDOE, 2008); problem posing (Niss, 1999; OECD, 1999); logic and reasoning (Mullis et al., 1997; NCTM, 2000; Niss, 1999; NRC, 2001; OECD, 1999); communication (NCTM, 2000; Niss, 1999; OECD, 1999); productive disposition (NRC, 2001); and aids and tools (Niss, 1999; OECD, 1999).

Goals of MET

The Mathematical Knowledge for Teaching (MKT) project of Deborah Ball and Heather Hill has become a very prominent effort to uncover the goals of MET (Ball et al., 2005; Ball et al., 2008; Hill et al., 2004; Hill et al., 2005). The MKT project and the associated studies have been cited extensively as a source of information on the goals of MET. In particular, the MKT project extended the categories of Shulman (1986, 1987). Shulman initially described the goals of MET to be content knowledge and pedagogical content knowledge. Ball et al. (2008) extended these categories by splitting each into three. Content knowledge was split into: common content knowledge, specialized content knowledge, and horizon content knowledge. Pedagogical content knowledge was split into: knowledge of content and students, knowledge of content and teaching, and knowledge of content and curriculum.

Within the content knowledge category, *common content knowledge* is described as "the mathematical knowledge and skill used in settings other than teaching" (Ball et al., 2008, p. 399), *specialized content knowledge* is described as "the mathematical knowledge and skill unique to teaching" (p. 400), and *horizon content knowledge* is described as "an awareness of how mathematical topics are related over the span of mathematics included in the curriculum" (p. 403).

Within the pedagogical knowledge category, *knowledge of content and students* is described as "knowledge that combines knowing about students and knowing about mathematics" (Ball et al., 2008, p. 401), *knowledge of content and teaching* is described as "knowledge that combines knowing about teaching and knowing about mathematics" (p. 401), and *knowledge of content and curriculum* is described as "represented by the full range of programs designed for the teaching of particular subjects and topics at a given level, the variety of instructional materials available in relation to those programs, and the set of characteristics that serve as both the indications and contraindications for the use of particular curriculum or program materials in particular circumstances" (Shulman, 1986, p. 10).

Goals of MER

Schoenfeld (2000) described the goals of MER to be the basic/pure goal of understanding the nature of mathematical teaching and learning, and the applied goal of using this understanding to improve mathematics instruction. These goals are strongly linked to the goals of MEL and MET. The goals of MEL and MET are pursued through the goals of MER, but they also might be shaped or determined by MER to some extent (see Figure 2 for a visual representation of these relationships). Throughout this dissertation study, the author will treat the goals of MEL and MET as the goals of MER. For example, if a study examines procedural fluency, then that study will be said to have the goal of L2 as seen in Table 1. Although more formally, the goal of that study is not to *obtain* procedural fluency, but rather to *understand* the obtaining of procedural fluency.

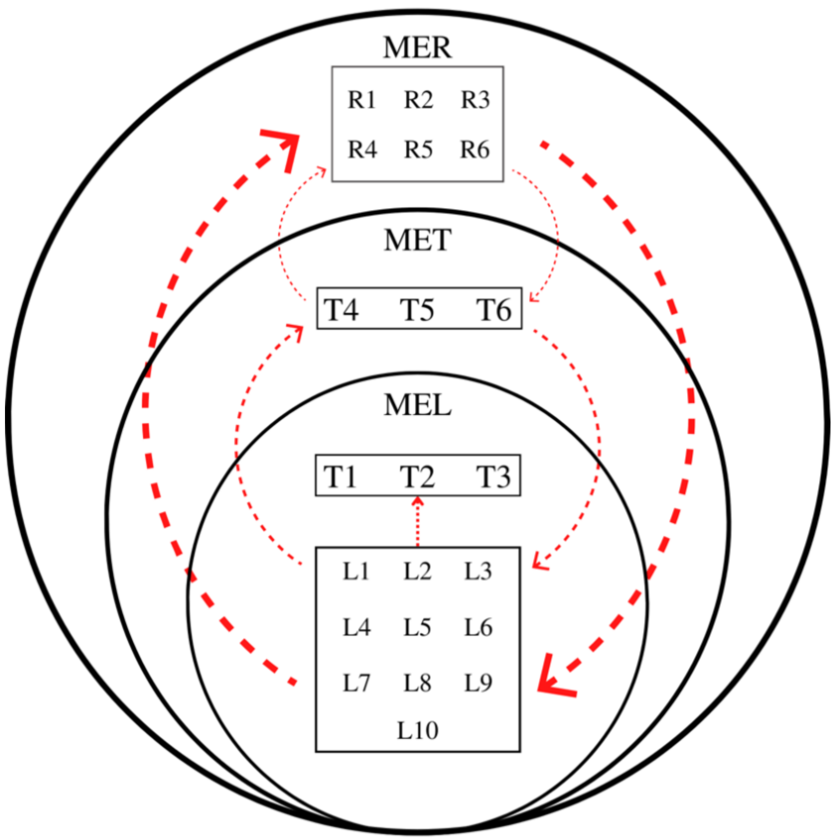
Begle and Gibb (1980) suggested that the goals of MER "are to find out how and why something works and then to see what works in practice" (p. 8) and that researchers and teachers share the common goal of "improving the teaching and learning of mathematics" (p. 3). Begle and Gibb discuss a further categorization of the pure and applied categories of research, first suggested by Hilgard (1964). The pure category was split into: not directly relevant, relevant subjects or topics, and school-relevant subjects or topics. The applied category was split into: laboratory, classroom, and special teacher; tryout in a "normal" classroom; and advocacy and adoption.

Within the pure category, not directly relevant refers to studies that examine topics adjacent to learning (e.g., a study of brain scans to determine which part of the brain is responsible for learning), directly relevant refers to studies that examine topics concerning the learning process itself (e.g., a study of student conceptual understanding), and school-relevant subjects and topics refers to studies that examine topics relevant to academic learning (e.g., a study of mathematics conceptual understanding in algebra).

Within the applied category, laboratory, classroom, and special teacher refers to studies that examine some kind of experimental conditions (e.g., a teaching program entirely designed by the researchers), tryout in a "normal" classroom refers to studies that apply some theory from the research literature to a classroom environment within a school, and advocacy and adoption refers to considerations of policy, curriculum, and implementation (e.g., writing an algebra textbook based on the research literature and advocating to implement it into classrooms).

Figure 2.

Diagram: the domains of MEL, MET, and MER; some of their EGs; and relationships among those goals.



Note: Please refer to Table 1 for the goals corresponding to the codes in this figure.

Research Methods

The methodology of a research article is primarily determined by the kind of data collected and analyzed, and in turn the data collected and analyzed is primarily determined by the measure used by the researchers. In this section, the author discusses the two main research methodologies, as well as the methodology that utilizes both of them.

Qualitative research involves collecting and analyzing non-numerical data. A working definition is provided by Creswell (2013):

> Qualitative research begins with assumptions and the use of interpretive/theoretical frameworks that inform the study of research problems addressing the meaning individuals or groups ascribe to a social or human problem. To study this problem, qualitative researchers use an emerging qualitative approach to inquiry, the collection of data in a natural setting sensitive to the people and places under study, and data analysis that is both inductive and deductive and establishes patterns or themes. The final written report or statement includes the voices of participants, the reflexivity of the researcher, a complex description and interpretation of the problem, and its contribution to the literature or a call for change. (p. 44)

Quantitative research involves collecting and analyzing numerical data. A working definition is provided by Creswell (2014):

> Quantitative research is an approach for testing objective theories by examining the relationship among variables. These variables, in turn, can be measured,

> typically on instruments, so that numbered data can be analyzed using statistical procedures. (p. 32)

Mixed-methods research combines these two approaches. Creswell and Plano Clark (2014) define that in mixed-methods, the researcher:

1. collects and analyzes both qualitative and quantitative research;
2. integrates the two forms of data and their results;
3. organizes these procedures into specific research designs; and
4. frames these procedures within theory and philosophy.

Choy (2014) examined the strengths and weaknesses of qualitative and quantitative methods and found that quantitative methods are advantageous primarily because they are quick and easy to administer compared to qualitative methods and that these methods allow comparison between groups. Choy found that qualitative research is advantageous because it allows researchers "to probe underlying values, beliefs, and assumptions" (p. 102) and it also is a broad and open-ended inquiry as compared to the narrower exploration of quantitative research. By nature, quantitative methods enable researchers to gather very specific data on a particular topic and particular group, while qualitative methods enable researchers to explore a range of topics on a range of groups. This can be summarized as saying that quantitative research holds the advantage of providing a depth of knowledge while qualitative research holds the advantage of providing a breadth of knowledge.

Mixed-methods research was described as advantageous for five reasons by Greene et al. (1989):

1. *triangulation* which means drawing from results of both methods to inform conclusions about a phenomenon;
2. *complementary* which means that the two methods can sometimes inform each other;
3. *development* which means using the results of one method to build on the other;
4. *initiation* which means the two combined methods can help uncover issues that may require the research question to be rethought; and
5. *expansion* which means combining the two methods can increase the breadth of inquiry.

Madey (1982) listed six ways that qualitative methods can inform quantitative methods and three ways that quantitative methods can inform qualitative methods. Madey noted that qualitative methods can inform quantitative methods by improving:

1. sampling framework;
2. evaluation design;
3. index construction;
4. external validation;
5. case study illustration; and
6. clarification.

He noted that quantitative methods can inform qualitative methods by improving:

1. leads;
2. overlooked respondents; and
3. correction of the elite bias.

Onwuegbuzie and Leech (2005) described the debate between the two main types of research methodology as so divisive that graduating students intending to enter academia "are left with the impression that they have to pledge allegiance to one research school of thought or the other" (p. 376). They then go on to make several claims pertaining to the often not discussed similarities between the two methodologies, a few of which are that:

1. both use observations to address research questions;
2. both use techniques that are analogous to an extent;
3. both attempt to use techniques that derive the maximal meaning from their data; and
4. both verify their data.

They then continue by arguing that researchers should become what they refer to as "pragmatic researchers" who incorporate both qualitative and quantitative methodologies into their line of research. The argument is that this helps researchers to "address a range of research questions that arise" (p. 383) and also to "delve further into a dataset to understand its meaning and to use one method to verify findings from the other method" (p. 384).

Prior Reviews of MER

One of the first major studies of the landscape of Mathematics Education Research (MER) was conducted by Begle (1979). In this study, Begle conducted a survey of the literature in MER in order to determine what he referred to as the "critical variables" of MER. The survey examined the research literature in MER in the time period of 1960 to 1976. A total of 33 journals were investigated, two of which (Educational Studies in Mathematics and Journal for Research in Mathematics Education) are also examined in this dissertation study for the modern time period. Begle found a great variety of variables being studied in MER across several categories (or levels), including teachers, curriculum, students, environments, instructional, tests, and problem solving. In each of these categories, Begle identified a great number of variables. In summarizing his observations, Begle noted that studies rarely used the same measuring instrument or examined the same kind of student. In addition, he noted "even when studies of a single narrow topic are reviewed, the rationales for the various studies rarely, if ever, fit together as parts of a single theoretical structure" (p. 155).

A more recent report (USDOE, 2008) published by the National Mathematics Advisory Panel (NMAP) reported a systematic review of the literature in MER and found,

> The dearth of relevant rigorous research in the field is a concern. First, the number of experimental studies in education that can provide answers to questions of cause and effect is currently small. Although the number of such studies has grown in recent years due to changes in policies and priorities at federal agencies, these studies are only beginning to yield findings that can inform educational policy and practice. Second, in educational research over the past two decades, the pendulum has swung sharply away from quantitative analyses that permit inferences from samples to populations. Third, there is a need for a stronger emphasis on such aspects of scientific rigor as operational

> definitions of constructs, basic research to clarify phenomena and constructs, and disconfirmation of hypotheses. Therefore, debates about issues of national importance, which mainly concern cause and effect, have devolved into matters of personal opinion rather than scientific evidence. (p. 63)

The NMAP's findings seem to echo the observations of Begle (1979). Notably, both identify a need for greater scientific rigor in MER. In addition, the NMAP concluded that more research is needed in MER which identifies:

1. effective instructional practices and materials;
2. mechanisms of learning;
3. ways to enhance teachers' effectiveness, including teacher education that focus on learning processes and outcomes; and
4. item and test features that improve the assessment of mathematical knowledge. (USDOE, 2008, p. 63)

In Chapter IV of this dissertation, the author provides evidence for the percentage of empirical research articles in the recent literature in MER, which corroborates the findings of both Begle and the NMAP. Additionally, in Chapter V, the author provides a discussion regarding item 4 just above as identified by NMAP.

A study published by Hart et al. (2009) examined the percentages of types of research methodologies implemented in the literature published in MER between the years 1995 and 2005. Hart et al. selected six total journals within this time-period, three of which were specifically mathematics education journals, and three of which were education journals in general (but not necessarily mathematics education specifically).

Of the three mathematics education journals selected, there was Journal for Research in Mathematics Education (JRME), Educational Studies in Mathematics (ESM), and Journal of Mathematics Teacher Education (JMTE). The first two of these journals are also examined in this dissertation. The total dataset had a size of 1,636 and the dataset selected for analysis had a size of 710. The Hart et al. study reports the percentages of types of research articles in the domain of MER, where the types are quantitative, qualitative, and mixed-methods. The researchers reported that these percentages across the 1995-2005 time-period are 66% qualitative, 21% quantitative, and 13% mixed-methods.

There was also another, smaller (albeit more thorough) study conducted more recently by Inglis and Foster (2018). In this study, the researchers examined two mathematics education journals in-depth: Journal for Research in Mathematics Education (JRME) and Educational Studies in Mathematics (ESM). The time-period was 1968 to 2015. These two journals are also examined in this dissertation. The study included an examination of two main aspects of the mathematics education literature: (1) the percentages of the specific mathematical topics of focus across the time-period examined, and (2) the percentages of studies implementing experimental design research methodologies across the time-period examined. Inglis and Foster found evidence that across the time-period, the percentages of specific mathematical topics changed (for several topics including proof, school algebra, and analysis; some increasing in prevalence and some decreasing) and also that the percentage of studies published in JRME and ESM which utilize experimental designs decreased significantly.

Theoretical Framework

Action Domain Theory (ADT) is an abstract generalization of the well-known theoretical construct in educational research of Learning Trajectories (LTs). In order to describe and define ADT, a close look at the theoretical framework of LTs will be helpful.

First developed in the 1990s by Simon (1995), LTs were developed to examine the goals of student learning and what paths are available to students in meeting such goals. The root of the theory of LTs, and by proxy ADT, goes back to the constructivist theories of John Dewey and Jean Piaget.

Constructivism

Constructivism builds on the work of Piaget (1970) in his theories of learning as an adaptive mechanism. This theory posits that we (human beings) are limited to observing the world only through the lens of our experiences and perceptions. Rather than learn from the world through passive observation, we must construct (hence the name) knowledge of the world around us. As Simon (1995) defines it:

> Constructivism derives from a philosophical position that we as human beings have no access to an objective reality, i.e., a reality independent of our way of knowing it. Rather, we construct our knowledge of our world from our perceptions and experiences which are themselves mediated through our previous knowledge. Learning is the process by which human beings adapt to their experiential world. (p. 5)

In the context of formal education, this theory posits that knowledge cannot simply be passed to students with the expectation that they can directly absorb it. Instead, they must be put in situations that impel them to construct their own knowledge. Work in recent years has urged the mathematics education community to give students *opportunities* to construct knowledge and for educators to move away from attempts at passive transference of knowledge. One such example is the development of Project-Based Learning (PBL). This is an approach to teaching which gives students projects to complete, with the teacher acting as a guide as the students construct (sometimes in a literal as well as figurative sense) the knowledge that the teacher has chosen as a goal for the students to learn (Capraro et al., 2013; Carpenter et al., 1996).

Learning Trajectory Theory

Simon (1995) built upon the theory of constructivism to develop the notion of LTs. Or as he initially named them, “hypothetical learning trajectories.” There are three primary components to LTs which Simon (1995) identified in his original study:

1. the learning goal;
2. the learning activities; and
3. the thinking and learning in which students might engage.

Each of the three components requires careful consideration in the theoretical framework of LTs. More recently, Clements and Sarama (2014) describe the three components of LTs as:

1. the goal;
2. the developmental progression; and

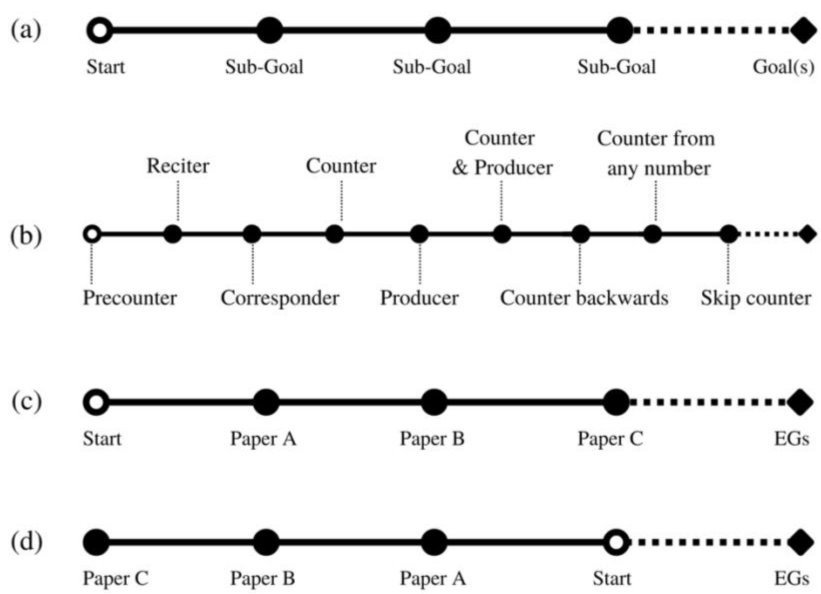
3. the instructional tasks.

They emphasize that the *goal* of an LT is the most important component to the trajectory, but that it is the least discussed. In discussing the developmental progression, they describe levels (or competencies) that students must meet along their way to the goal. These may be viewed as "sub-goals." A great deal of work has been done in an attempt to uncover what exactly these goals (and also the sub-goals or competencies) should be, as discussed previously in this chapter.

Sarama and Clements (2009, pp. 73-79) provide examples of LTs for young children in elementary mathematics. One is of a child learning to count (see Figure 3b).

Figure 3.

Diagram: (a) general trajectory in ADT, (b) adapted from Sarama and Clements (2009, pp. 73-79), (c) RT with matching IGs and EGs, and (d) RT with differing IGs and EGs.



Several stages or sub-goals are provided and given names: pre-counter, chanter, reciter, corresponder, counter, produce, and counter and producer. These goals correspond to age levels as well, and each sub-goal is provided tasks designed to facilitate learning of that level. For example, the reciter is given count-along games such as counting along to a song or to a computer game. This explicates the goal (counting), the sub-goals/competencies (such as reciter) comprising the developmental progression, and the instructional tasks (i.e., count-along games). These describe one key piece to the puzzle when it comes to LTs. Another key piece is how to actually determine whether a student has achieved the sub-goals and, finally, the main goal.

An important consideration in any LT model is the consideration of assessment (or measurement). This is because, (1) a teacher needs to know when a student has achieved the goal, and (2) a teacher needs to know when a student is ready to move from one level of the developmental progression to the next. Likewise, a teacher needs to know when a student has achieved a sub-goal, and the teacher needs to know when a student is ready to move from one sub-goal to the next. Scriven (1967) describes two kinds of assessment: formative and summative. The distinction between these two kinds of assessment has become a standard idea in the domain of mathematics education. Summative assessments have been described as "a judgement which encapsulates all the evidence up to a given point" (Taras, 2005, p. 467). One of the most common summative assessments used are standardized tests, but there is substantial debate about their usefulness (Goslin, 1963; Kohn, 2000; Neill & Medina, 1989; Phelps, 2005). In the context of LTs, standardized tests may be useful in determining whether a student has met the overall goal of the trajectory. However, LTs also require measurement of the sub-goals comprising the developmental progression. For this purpose, formative assessments may be employed. This type of assessment has been described as “an active and intentional learning process that partners the teacher and the students to continuously and systematically gather evidence of learning with the express goal of improving student achievement” (Moss & Brookhart, 2019, p. 6). To summarize, in order for LTs to be properly implemented, teachers require four things:

1. an overall learning goal;
2. a developmental progression of sub-goals;
3. instructional tasks meant to develop these sub-goals; and
4. the ability to measure where a student is along the developmental progression of sub-goals (via formative assessment), including the ability to measure whether the overall learning goal has been achieved (via summative assessment).

Action Domain Theory

Now that we have formulated the four necessary components of Learning Trajectories (LTs), it becomes possible to generalize the theoretical framework of LTs to build Action Domain Theory (ADT). In reframing, it is useful to relabel the four components simply as:

1. the goal;
2. the trajectory to the goal;
3. required actions to traverse the trajectory to the goal; and

4. the ability to take measurements along (and at the end of) the trajectory to the goal.

What this relabeling affords is the ability to abstract the core theoretical constructs present in LTs to other domains. We will limit our discussion to academic domains. However, it is worth noting that ADT can be applied outside the realm of education and research. A general trajectory in ADT can be seen in Figure 3a.

Early accounts of what constitutes an academic domain can be found in Biglan (1973) who described two ways of classifying academic domains: hard vs soft and pure vs applied. Glaser et al. (1987) describe an academic domain as having three primary defining characteristics:

1. a paradigm;
2. concern for practical application; and
3. concern for life systems.

It is clear that mathematics falls into Biglan's hard category and pure category. However, mathematics education is a domain which seems to overlap with the hard, soft, pure, and applied categories because there is concern for both mathematics and for education. Furthermore, it is reasonable to claim that mathematics as a discipline all its own, the learning of mathematics, the teaching of mathematics, and the research of these pursuits are all somehow distinct domains. If we distinguish between the domains of Mathematics Education Learning (MEL), Mathematics Education Teaching (MET), and Mathematics Education Research (MER), then the existence of the theory of LTs poses the question: if the domain of MEL has LTs as a framework, then what might be the analogous construct in the domain of MER?

ADT posits that given any domain within which actions are taken, there is an overarching *goal*. Whenever a goal is established, whether implicitly or explicitly, there is necessarily an action domain within which that goal resides. That is to say, a goal necessarily defines an action domain. ADT also posits that any action taken in an action domain is necessarily done with intention to move toward the goal. If an individual acts without intention (whether conscious or unconscious) to move toward the goal, then the individual has acted in a different action domain in that instance. The action domain itself can be *defined* by the overarching goal, but it is *comprised of* (composed and made entirely of) actions which move an individual toward that goal. We say that a *trajectory* is a set of actions that leads to the goal. It is important to note that the concept of a *measure* in the context of ADT is an instrument which allows for the measurement of the distance to the goal. This distance we may define to be the smallest number of required actions to reach the goal. However, it is also important to note that a goal does not necessarily have to be *measurable*. ADT allows the abstraction of LTs to the domain of MER.

Applying ADT to MER: Research Trajectory Theory

In the domain of MER, there is currently no analog to the LTs of the domain of MEL. In other words, there is no such concept as a “research trajectory”. However, by taking the four components of LTs and transplanting them into the domain of MER (see

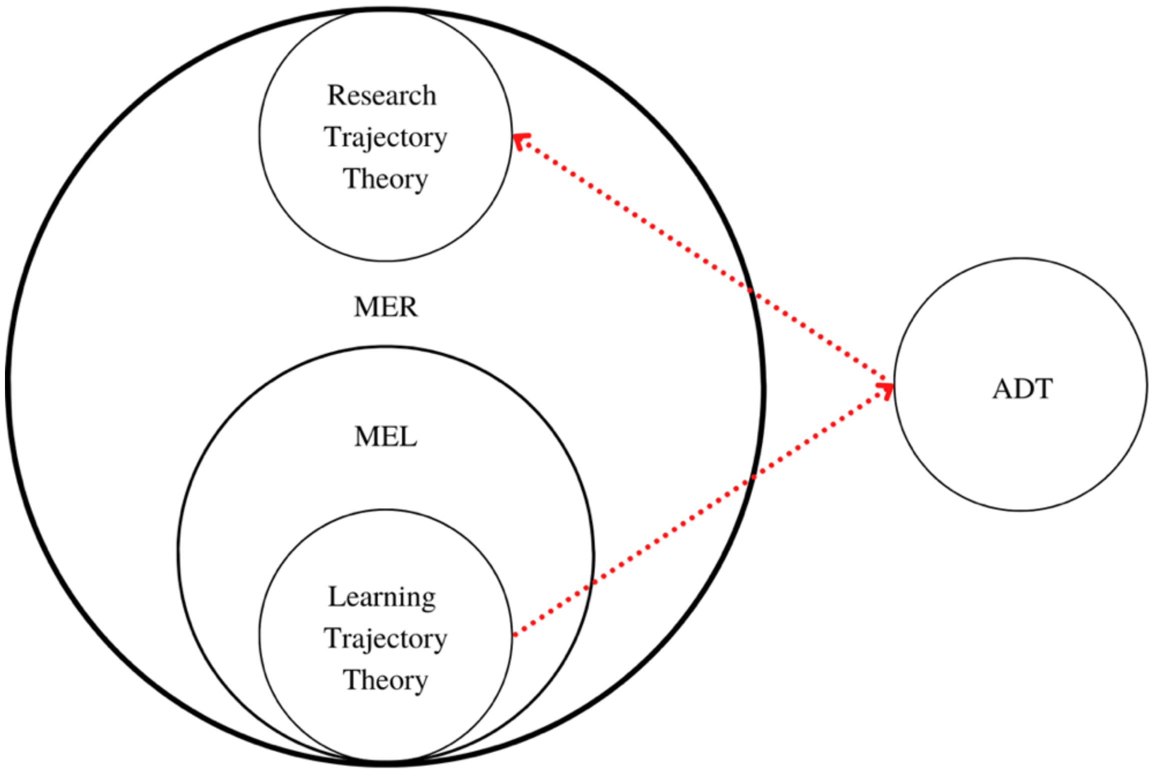
Figure 4) it is possible to discuss such a concept. Suppose that we relabel the four components as:

1. the goal of some group of mathematics education researchers (or possibly a single researcher);
2. the progression of research required to achieve that goal (which may be unknown a priori and so the hope is that ADT can at least shed light on whether the trajectory is moving toward the goal or not);
3. the specific research projects required for that progression (this also may be unknown a priori); and
4. the ability to measure progress (this is a potential contribution of ADT).

Taken together, these components describe what we will refer to as a *Research Trajectory (RT)*. As an example, let us suppose that we have the first component in the form of the overarching research question: does PBL improve student math achievement? Then answering this question is our goal as mathematics education researchers. The second component would then consist of sub-questions that altogether would help answer the overarching research question. The third component would be the studies conducted to answer those sub-questions. Finally, the fourth component is the broad combination of the domain's either agreement or disagreement with the results and methodologies of all of those studies. This new theoretical construct of RTs is an example of ADT. However, ADT is one step further in abstraction and it is ADT which provides the conceptual formulation of domains. This contribution of ADT is what will enable discussion of the domains of MEL, MET, and MER.

Figure 4.

Diagram: constructing ADT from LT Theory in MEL and applying ADT to MER to construct RT Theory.



RTs are a unique application of ADT because the RTs require measures of progress under the framework of ADT yet the RTs themselves make progress toward the respective goal(s) by the use of measures. In other words, measurement and collection of data is what drives progress along the trajectory of an RT. This yields two levels of measurement to consider in RTs: (1) the measurement of progress toward the goal(s) along an RT, and (2) the measurements taken within an RT which produce the data from which the research projects along the RT are conducted, thus driving progress toward the goal(s). In this study, the second level of measurement is examined in order to derive some insight concerning the first level of measurement.

Another consideration on the topic of measurement is measuring the progress of the RTs of MER as a whole. In an attempt to pursue this kind of measurement, there will be two measurement indicators considered:

1. whether or not RTs are moving toward the Explicit Goals (EGs) of MER; and
2. whether or not the RTs are compatible in the sense that the results of studies within different RTs can be consolidated to derive more general conclusions.

The first indicator is somewhat clear in its reasoning: for the RTs to make progress toward the goals of MER, then surely the RTs as a whole move toward the EGs of MER. This indicates that the Implicit Goals (IGs) of the RTs match the EGs of MER, as can be seen in Figure 3c. If the IGs of the RTs do not match the EGs of MER, then the RTs may move "away" from the EGs of MER, as can be seen in Figure 3d.

The second indicator is not as clear in its reasoning: if the collection of RTs in MER are incompatible along some dimension, say the topic of interest, then the results cannot be meaningfully consolidated. As an example, suppose there are two studies, one of which examines the topic of algebra, and the other which examines the topic of geometry. Suppose that both studies use the same population, measure the same variables, use the same measures, and employ the same analytic methods. In other words, the studies are compatible along every dimension except for the topic of interest. Then both studies might move MER toward the same EGs of MER, yet they each move toward different *aspects* of these EGs. Further progress toward either of these different aspects, though they may belong to the same EGs, would require further study of these specific aspects. For this reason, the second indicator of progress of the RTs as a whole toward the EGs of MER is examined in this dissertation.

CHAPTER III

METHODOLOGY

> If we knew what kind of knowledge mathematics education aims at, we would be better equipped for answering the question of methods of validation.
>
> —Sierpinska et al., 1993, p. 278

The methodology of this dissertation study was exploratory in nature. For Analysis 1, a mixed-methods methodology was implemented to survey the types of articles appearing in the selected sample. For Analysis 2 and Analysis 3, a qualitative inquiry was implemented to explore the variety of features in the quantitative article type found in Analysis 1 and in the selected sample of studies analyzed in one of the meta-analytic articles found in Analysis 1. Due to the exploratory nature of this dissertation study, hypotheses were not formulated for testing, though statistical tests were performed whenever appropriate and relevant to the research questions. The goal for this design was to determine the variety of articles in general in Mathematics Education Research (MER; article types in Analysis 1), then to narrow the sample to only one specific article type to determine the variety of features within a specific type (quantitative articles in Analysis 2), and finally to examine articles selected to have a low variety of features (the sample of studies included in one of the meta-analysis articles in Analysis 3).

Pilot Investigations

The motivation for initial investigations was to gain some insight into:

1. the types of features there might be in the landscape of MER;

2. what some good methods might be to categorize and distinguish articles in MER; and
3. what a viable approach could be for this type of exploration.

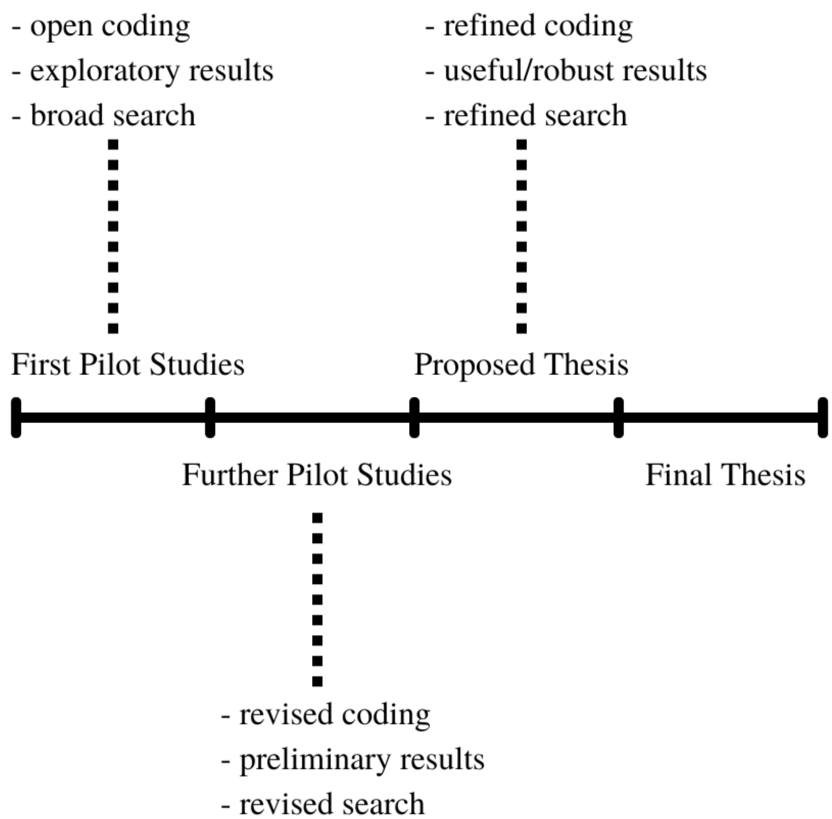
The first pilot study was conducted beginning in 2020 by gathering papers using a large number of databases as searched by keywords such as “math education.” Papers were then coded using open, axial, and selective coding (Glaser & Strauss, 1967). During this initial study, the codes were simply intended to distinguish articles that implemented a qualitative methodology, quantitative methodology, or neither. However, during the coding it was discovered that there are many more types of articles than these, and it was also found that many of the articles did not belong to the domain of MER. It was concluded that the search methodology was flawed and that the coding scheme needed to be expanded to account for the various types of articles found in MER.

Subsequent pilot studies, beginning in early 2021, were attempts to refine the flawed search methodology, and the coding scheme was expanded to include categories such as “meta-analytic” and “mixed-methods.” These changes yielded more promising outcomes, yet the search methodology itself was found to still be insufficient for filtering out articles not belonging to the domain of MER. After explorations of possible alternative search methodologies, the decision was made to select journals belonging specifically to the domain of MER for further studies.

After the selection of journals, a number of volumes, issues, and papers between the years 2010 and 2020 were examined. This served as a means to test and further refine the coding scheme. Eventually, the coding scheme utilized in this dissertation was decided upon. Figure 5 demonstrates the progression from the initial investigation to this dissertation and what considerations were included at each step of the process.

Figure 5.

Diagram: the progression from pilot investigations to this study.



Data Sources

The data for this study originates from articles published between 2016 and 2021 (inclusive) in five leading journals in the domain of MER. For Analysis 1 and Analysis 2, the articles are directly from the selected date range and journals. For Analysis 3, the articles are included in a meta-analytic study, and the meta-analysis is an article from the selected date range and journals.

Selecting the Date Range

The date range of 2016 to 2021 (inclusive) was selected for several reasons. One is that the date range needed to be broad enough to allow for a sufficient sample size of articles in order for the subsequent analyses to be robust. Another is that selecting a short date range would make it more difficult to identify any possible trends over time in the data. Another third reason is that the date range of 2016 to 2021 captures papers published recently, papers published prior to the 2020 COVID pandemic, and papers published after the 2020 COVID pandemic.

Selecting the Journals

The five journals selected are Educational Studies in Mathematics (ESM), Journal for Research in Mathematics Education (JRME), Mathematics Education Research Journal (MERJ), Research in Mathematics Education (RME), and ZDM Mathematics Education (ZDM). The author of this dissertation study selected journals as the modality for acquiring published research in MER because the largest volume of research is published through this modality. However, it is an important consideration that the inclusion of only this modality of research might impose a limitation on this dissertation study by way of a publication bias. This publication bias might be due to journals emphasizing certain perspectives and deemphasizing others that may be better represented in other modalities such as books, theses, or other types of publication. The extent to which this is the case is not clear and it would be an interesting future exploration to examine this possible limitation more closely.

The five journals were selected based on citation ratings and based on prior reviews of MER. The Web of Science citation index named the Social Science Citation Index (SSCI) was initially considered as a basis for selection of leading journals in MER. However, not all of the years in the selected date range were indexed for many journals in the SSCI. For this reason, the SSCI in addition to rankings provided by prior publications were used as a basis for the journal selections. In particular, Williams and Leatham (2017) recently ranked a number of journals in MER by quality, of which the journals selected for the data sources in this dissertation were ranked 1 (ESM), 2 (JRME), 7 (ZDM), 8 (MERJ), and 15 (RME).

Another important consideration for journal selection was availability and access to volumes and issues published in the selected date range. The host university's access for journals in this date range partially limited journal selection, and in addition to the considerations listed above, the university's access meant that the five selected journals were optimal for this dissertation. Of the journals selected and within the chosen date range, only two issues were inaccessible, both of which were issues from MERJ. Specifically, Volume 33, Issue 4 was inaccessible due to a one-year access delay at the host university, and Volume 32, Issue 3 was inaccessible for unknown reasons.

Each of the five selected journals has its own perspective, focus, and scope. These factors influence, to some extent, the articles that a journal's editors select for publication. The journal editors often state these as the "aims and scope," or otherwise provide a brief description of the journal's purposes, on the journal's website. For each of the descriptions subsequently discussed, the author of this dissertation study retrieved the descriptions from each journal's website in December of 2022. The "aims and scope" of the editors of ESM are stated as,

> *Educational Studies in Mathematics* presents new ideas and developments of major importance to those working in the field of mathematics education. It seeks to reflect both the variety of research concerns within this field and the range of methods used to study them. It deals with methodological, pedagogical/didactical, political and socio-cultural aspects of teaching and learning of mathematics, rather than with specific programmes for teaching mathematics. Within this range, Educational Studies in Mathematics is open to all research approaches. The emphasis is on high-level articles which are of more than local or national interest. All contributions to this journal are peer reviewed.

This description appears to indicate that ESM contains a wide assortment of different articles but does not deal specifically with particular curricular prescriptions.

The editors of JRME describe their journal as,

> An official journal of the National Council of Teachers of Mathematics (NCTM), JRME is the premier research journal in mathematics education and is devoted to the interests of teachers and researchers at all levels--preschool through college. JRME presents a variety of viewpoints. The views expressed or implied in JRME are not the official position of the Council unless otherwise noted.

Unlike the description of ESM, the description of JRME does not specifically discourage the submission of particular kinds of articles. The editors of JRME appear to describe their journal as one that is open to all kinds of research.

The editors of MERJ describe the "aims and scope" of their journal as,

> The Mathematics Education Research Journal seeks to promote high quality research that is of interest to the international community.
>
> The Mathematics Education Research Journal seeks to present research that promotes new knowledge, ideas, methodologies and epistemologies in the field of mathematics education.
>
> The Mathematics Education Research Journal actively seeks to promote research from the Australasian region either as research conducted in the region; conducted by researchers from the region and/or draws on research from the region. The Mathematics Education Research Journal accepts papers from authors from all regions internationally but authors must draw on the extensive research that has been produced in the Australasian region.
>
> The Mathematics Education Research Journal normally does not encourage publication of teacher education programs or courses. These are more suited for the other MERGA journal, Mathematics Teacher Education and Development.

The editors of MERJ appear to dissuade authors from submitting articles dealing with particular teaching programs, in agreement with the "aims and scope" of ESM. It is worth noting that both MERJ and ESM are Springer journals. Additionally, the editors of MERJ emphasize their journal being comprised of research of an international scope, whereas ESM emphasizes a local and national scope. JRME, on the other hand, is operated by NCTM (a U.S. organization), yet its editors do not make mention of the geographical scope of the journal.

The editors of RME describe the "aims and scope" of their journal as,

> Research in Mathematics Education is an international English language journal, publishing original refereed articles on all aspects of mathematics education. Papers should address the central issues in terms which are of relevance across educational systems and informed by wider thinking in the field. The journal has three sections, covering research papers, book reviews, and current reports.

As was true for MERJ, the editors of RME emphasize their journal as international in scope. The editors also emphasize that the publications within RME should take into consideration the relevance of the educational systems across international lines.

The editors of ZDM describe the "aims and scope" of their journal as,

> ZDM – Mathematics Education is one of the oldest mathematics education research journals. The papers appearing in the seven themed issues per year are strictly by invitation only followed by internal peer review by the guest-editors and external review by invited experts. The journal exists to survey, discuss and extend current research-based and theoretical perspectives as well as to create a forum for critical analyses of issues within mathematics education. The audience is predominantly mathematics education researchers around the world interested in current developments in the field.

Similarly to MERJ and RME, the editors of ZDM emphasize that their journal is of an international scope. However, unique to the description of ZDM is that it is an invitation-only journal and that each issue is themed. All of the different perspectives, foci, aims, and scopes presented by the editors of the five journals that the author of the current dissertation study has included for analysis might influence the kinds of articles contained in each of the journals.

Analysis 1

Analysis 1 is a survey of article types. All articles within the selected date range and within the selected journals were coded using the refined coding scheme developed in the pilot investigations. All papers were added to an Excel spreadsheet and sorted by journal, volume, and issue. The full text of each paper was read carefully and assigned a code using the coding scheme outlined in Table 2. After coding, frequencies were counted for each article type, both total and by journal. Additionally, frequencies per year for each article type were counted. These counts were used for the remainder of the analysis process.

Table 2.

Coding scheme used for article type determinations.

Code	Criterion
Quantitative (Quant)	The article includes results of some statistical test: p-values, effect sizes, correlations, etc.; only the calculation and reporting of means, standard deviations, and/or reporting of counts does not suffice as a statistical test under this criterion because these alone do not provide a decision on statistical significance. The data type reported is numerical.
Qualitative (Qual)	The article includes results of an analysis but does not report any numeric data or results except perhaps for counts or percentages. The main data type reported is categorical. The method of data collection may be in the form of interviews, observation, surveys, coding for themes, narrative data, etc.
Mixed-Methods (Mixed)	The article transforms qualitative data into quantitative data for a quantitative analysis; OR, the article includes the results of both a quantitative study AND a qualitative study as defined above.
Meta-Analytic (Meta)	The article is a meta-analysis of EITHER qualitative or quantitative research; OR otherwise, the article is a synthesis of research.

Code	Criterion
Theoretical (Theo)	The article does not include participants AND includes EITHER: results of the formulation of a new theory, results of testing of the validity/reliability of some instrument, or some other discussion of theoretical considerations.
Commentary (Comm)	The article is a commentary on some previously published work(s).
Other	The article does not fit into any of the other categories.
Excluded (Ex)	The paper is not a refereed article, such as an editorial, announcement, correction, erratum, etc.

Difficult Cases to Code

Some of the items were more difficult to code than others. Some of these cases and how they were handled are as follows:

If the article was a book review, it was classified as "commentary."

If the article collected quantitative data (the raw data were numerical) but did not report statistical results (*p*-values, correlations, or things of that sort) then it was classified it as "other."

To distinguish between "commentary" articles and "theoretical" articles, the theoretical type had to contribute some sort of theoretical consideration, a review of articles was not sufficient, but rather a synthesis must have been provided with the intent to contribute to the theoretical base of knowledge.

Some articles seem able to be classified as either "meta-analytic" or "commentary," so to distinguish these, it was not sufficient for an article to only provide a review of literature in order to be classified as "meta-analytic."

For the "meta-analytic" type to be chosen, the article had to report some kind of systematic search of the literature and a subsequent analysis and/or synthesis of results.

A key identifying characteristic of the "mixed-methods" type was for the article to report an inter-rater reliability, effectively meaning there was a coding of qualitative (categorical) data at some point in the methodology, in addition to reporting of statistical results as mentioned prior.

If the article included a transformation of qualitative data to quantitative data (via the coding and inter-rater reliability process) but did not subsequently report any statistical results, then it was classified it as "other."

Quantitative articles had to only collect numerical data (coding of data would immediately classify the article as "mixed-methods") and report statistical results.

Inter-Rater Reliability

Due to the statistical analyses conducted on the counts obtained from the coding performed in Analysis 1, the coding scheme was tested for inter-rater reliability. The author selected a random sample of about 10% of the included articles. This was accomplished by randomly sorting all of the included articles using Excel's RAND() function and then selecting the first 10% of the randomly sorted list of articles. This random sample was given to Professor Howe along with the coding scheme. The percent agreement was reported to be about 73%. This should be considered a lower bound on agreement because Dr. Howe did not have sufficient time to carefully determine the category of the more difficult-to-code articles. By consensus, a reliability greater than or equal to 70% is considered to be acceptable reliability (Stemler & Tsai, 2008). The popular Cohen's Kappa statistic was not considered due to various unfavorable mathematical properties which can often render interpretations to be flawed (Pontius Jr. & Millones, 2011).

Comparing Journals

Using the counts of article types by journal, possible differences between journals were investigated. First, line charts were created using Excel's built-in plotting tools to search for any possible trends in the percentages of article types by journal, both in all included articles and within only the empirical article types (quantitative, qualitative, mixed-methods, and meta-analytic). These line charts are reported in Chapter IV.

Next, Kendall's τ (rank) correlation was calculated to test for any possible correlations in the percentages of article types (Kendall, 1938). The percentages of article types by journal were used as points, yielding five points per article type to test sets of five pairs for possible correlations between article types. Kendall's τ is suitable for testing correlations when the sample size is small and when the distribution of the variables tested may not be normal. The correlations were calculated by loading the data in the R studio statistical software using R version 4.2.2. The cor() function was then used to run the tests. The results are reported in Chapter IV.

To compare the percentages of article types between journals, two-proportions *z*-tests were conducted (Wilson, 1927). This was conducted on the sample of all included articles and also separately on the empirical articles. In the case that only two groups are compared, a *z*-test is precisely equivalent to the χ^2-test (Wallis, 2013). The *z*-tests were performed by adding the data into the R Studio statistical software using R version 4.2.2.

The prop.test() function was then used to run the tests. R automatically applies Yates' (1934) continuity correction where applicable when running the prop.test() function. The results are reported in chapter IV.

Trends Over Time

The year-by-year counts of article types were tabulated to check for possible trends over the full six-year period. Using Excel's built-in plotting tools, the author created line charts to search for trends, both among all article types, and also among the empirical (qualitative, quantitative, mixed-methods, and meta-analytic) article types. These line charts are included in Chapter IV.

In addition, Mann-Kendall (Kendall, 1976) times-series trend tests were performed to search for any statistically significant evidence of trends over time. This particular test seemed appropriate because it is useful for checking time-series data for monotonic trends, even with a small sample size, such as the six-year data of this dissertation study. In order to run the Mann-Kendall test, the data was loaded into the R Studio statistical software using R version 4.2.2. The Kendall library was loaded in order to use the MannKendall() function. For each paper type, the corresponding time-series was tested using this function. The raw counts were used for the data. The *p*-values were calculated and are reported in Chapter IV.

Analysis 2 and Analysis 3

Analysis 2 and Analysis 3 are purely qualitative and exploratory in nature. Unlike in Analysis 1, where determining a given article's type is relatively well-agreed upon, Analysis 2 and Analysis 3 in part examine the Implicit Goals (IGs) and Explicit

Goals (EGs) of particular articles. There is currently no agreed-upon method for conducting such explorations and this study serves as a sort of first example of its kind. For this reason, the decision was made to not run statistical tests on the data gathered in Analysis 2 and Analysis 3, and inter-rater reliability was not considered.

The author of this dissertation study obtained the data for Analysis 2 from the articles coded as quantitative in Analysis 1. The data for Analysis 3 was obtained by selecting one of the articles coded as meta-analytic in Analysis 1 and then using the sample of studies included in that article's analysis.

Observational Survey

For both Analysis 2 and Analysis 3, the author examined the articles and initial observations were noted. In particular, some of the populations included in the analyses of the studies and some of the topics of interest of the studies were noted and lists of these observations were made. The author grouped some of these items under common themes, such as "mathematics topics." This piece of the analysis is intended to provide insight to the variety of populations from which data is collected and the variety of topics of interest. Additionally, the author noted observations of some of the variables measured, measures used, and analytic methods implemented. The Appendix provides tables of these observations.

Extracting EGs and IGs

For both Analysis 2 and Analysis 3, the EGs and IGs were extracted from the sample of articles and collected for data. The EGs were found by extracting explicit statements of goals from the articles. The three forms that these statements take are research questions, hypotheses, and statements (of goals, aims, or purposes). Research questions are EGs in the form of questions. Hypotheses are EGs in the form of hypothetical claims. Statements are EGs in the form of statements of purposes, aims, or goals. Calculations of counts for the number of articles with a certain type of EG statement form are provided in the analyses.

The EGs of each article were read and coded according to the codes provided in Table 1. A best determination was made in order to identify which, if any, of the EGs of Mathematics Education Learning (MEL) and/or Mathematics Education Teaching (MET) each EG statement seems to point to. The two goals of Mathematics Education Research (MER) were not used for coding because these two goals are strongly linked to the EGs of MEL and MET as discussed in Chapter II (Schoenfeld, 2000). The IGs of each article were identified by reading the methodology and results sections of each article thoroughly. The measures implemented and variables measured were examined in order to code which of the EGs of MEL and/or MET were being measured, which was used for the IG coding determination.

Calculations of counts for the number of EG/IG matching pairs (pairs of EGs and IGs which match within a given article) are provided. Calculations of counts for the number of articles with EGs or IGs that directly link (to one of the goal codes in Table 1) and the number of articles with EGs or IGs that do not directly link are also provided. The former is used to draw insight regarding whether or not author EGs match their IGs, and the latter is used to draw insight regarding whether or not author EGs or IGs match the EGs of MER. For additional insight, counts of the appearance of each of the 10 EGs of MEL and each of the 6 EGs of MET in Table 1 are calculated and provided.

It is not always readily clear which, if any, of the EGs of MEL and/or MET a particular EG statement should be linked to, nor is it always readily clear which of the EGs of MEL and/or MET the measures and variables measured should be linked to. For this reason, Analysis 2 and Analysis 3 should be considered much more exploratory in nature than Analysis 1. It should also be considered that the measures and variables measured were the standard by which IG determinations were made because Action Domain Theory (ADT) posits that goals and measures are strongly linked, such that a measure can imply IGs of a given trajectory, as discussed in Chapter II.

As a demonstration of the EG/IG coding process, we may consider four possibilities:

1. the article has an EG statement that *can* be coded to the goal codes listed in Table 1;
2. the article has an EG statement that *cannot* be coded to the goal codes listed in Table 1;
3. the article has IGs that *can* be coded to the goal codes listed in Table 1; and
4. the article has IGs that *cannot* be coded to the goal codes listed in Table 1.

For each of the four cases, an article included in Analysis 2 will be used to exemplify the process of EG/IG coding.

For case 1 and 3 of the EG/IG coding process, we can consider the quantitative article published by Strohmaier et al. (2019). The researchers list three research questions that are treated as the EG statements: (RQ1) "Do eye movements associated with cognitive processes during word problem solving differ between students solving word problems in German and in Chinese?" (p. 48), (RQ2) "What reading patterns can be distinguished by clustering eye movement patterns in the two groups? How similar are these reading patterns and how are they distributed?" (p. 49), and (RQ3) "How are reading patterns associated with mathematical performance, flow experience, mathematical self-concept, and mathematical anxiety?" (p. 49).

In RQ1, the terminology of "solving word problems" was coded to L5 (Problem Solving) and in RQ3, the terminology of both "mathematical self-concept" and "mathematical anxiety" appeared to describe aspects of L9 (Productive Disposition). The terminology of "eye movement" and "flow experience" did not appear to clearly link to any of the goals listed in Table 1. The overall result was that the EGs of this article were coded as L5 and L9.

To code the IGs, the author of the current dissertation study examined the variables measured and measures used (p. 50). The variables measured and measures used agreed with the EG statements (RQ1-RQ3). The measures (adapted from measures used in the PISA studies (OECD, 1999)) were described as a measure of word problem solving ability and measures of both math anxiety and self-concept, but the descriptions of these measures did not reveal a link to any further goals listed in Table 1, so the IGs were also coded as L5 and L9.

In some of the articles included in Analysis 2 of this dissertation, terminology such as "mathematical performance" in RQ3 of Strohmaier et al. did not appear to clearly link to any of the goals listed in Table 1, but upon examination of the measures of mathematical performance that the researchers used, the descriptions appeared to link to some of the goals listed in Table 1. This happened when the researchers used general terms to describe mathematics performance, achievement, etc., but their measures captured information about only some of the specific goals associated with mathematical performance in general.

For case 2 and 4 of the EG/IG coding process, we can consider the quantitative article published by Morton and Riegle-Crumb (2019). The researchers include the following research question that is treated as the EG statement,

> To what extent do racial disparities in prior opportunities to learn (as evidenced by grades, test scores, and level of prior mathematics course) explain subsequent inequality in access to eighth-grade algebra in two different school contexts of racially integrated middle schools and predominantly Hispanic middle schools? (p. 531)

None of the terms in this EG statement appear to link directly to any of the goals listed in Table 1. The terms "grades," "test scores," and "level of prior mathematics course" seem to describe mathematics performance (or achievement) in general, but as discussed above, these kinds of terms do not directly link to specific goals that comprise mathematics performance. For this reason, the EG statement cannot be coded as any of the goals listed in Table 1.

To code the IGs, the author of the current dissertation study examined the variables measured and measures used in order to determine whether these provide further insight into what aspects of mathematics performance were measured. However, the measures indicate that even when mathematics performance was measured by the researchers, the description did not provide insight that linked the measures to any of the goals listed in Table 1. For example, the measure of "test scores" was described as "a standardized version of students' seventh-grade mathematics test score on the state accountability exam" (p. 538). This indicates that some of the goals listed in Table 1 might have been measured by some of the items on this exam, but Morton and Riegle-Crumb did not provide enough information in the description of the measure to be able to distinguish which of the goals were measured. Therefore, the IGs could not be coded to any of the goals listed in Table 1.

CHAPTER IV

RESULTS

The author found a total of 1,514 individual items across the five journals between 2016 and 2021. Of these, 197 were either errata, editorials, or non-refereed notes, and so were excluded from this study. This left 1,317 refereed articles that were included in the analyses reported here. Of these, 285 were classified as “theoretical”, “commentary”, or “other” and 1,032 were of the empirical article types (qualitative, quantitative, mixed-methods, or meta-analytic). The total number of quantitative articles was 115, and these were included in Analysis 2. There were 31 total articles coded as meta-analytic, and of these, the one chosen for use in Analysis 3 was Rittle-Johnson et al. (2017). Their study included 22 articles, which the author included in Analysis 3 of the current dissertation study. Counts of the paper types are given in Table 3 by journal and Table 4 by year.

Table 3.

Counts of paper types by journal.

Journal	Quant	Qual	Mixed	Meta	Theo	Comm	Other	Ex	Total
ESM	23	215	63	12	38	40	2	47	440
JRME	15	45	24	5	2	41	1	48	181
MERJ	17	94	25	1	6	2	11	10	166
RME	8	52	21	1	11	28	0	36	157
ZDM	52	268	79	12	35	57	11	56	570
Total	115	674	212	31	92	168	25	197	1514

Table 4.

Counts of paper types by year.

Journal	Quant	Qual	Mixed	Meta	Theo	Comm	Other	Ex	Total
2016	23	102	23	4	12	41	1	36	242

Journal	Quant	Qual	Mixed	Meta	Theo	Comm	Other	Ex	Total
2017	17	90	33	6	20	30	2	31	229
2018	13	124	26	2	8	35	9	26	243
2019	20	118	27	6	17	23	5	30	246
2020	25	100	59	7	12	19	3	45	270
2021	17	140	44	6	23	20	5	29	284
Total	115	674	212	31	92	168	25	197	1514

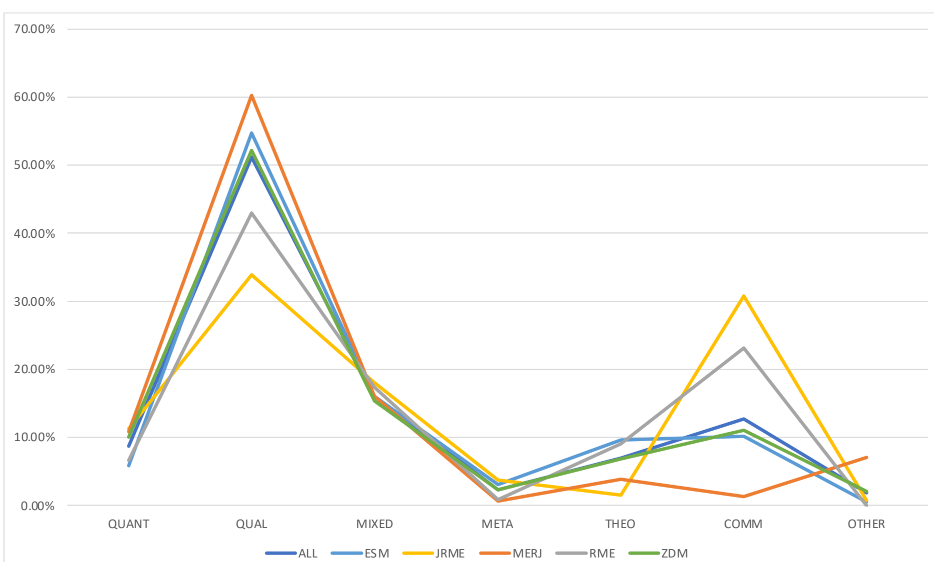
Analysis 1

All Included Articles (N = 1,317)

Of the 1,317 included articles, the number of the articles types was 115 quantitative (~9%), 674 qualitative (~51%), 212 mixed-methods (~16%), 31 meta-analytic (~2%), 92 theoretical (~7%), 168 commentary (~13%), and 25 other (~2%). Figure 6 displays a line chart of the percentages of article types by journal. It appears that the greatest variation across journals occurs in the qualitative and commentary types, but for all remaining types there appears to be little variation across the journals.

Figure 6.

Line chart: percentages of article types by journal.



The Kendall's τ correlations displayed in Table 5 seem to corroborate this observation, as the correlation between the percentages of qualitative articles and commentary articles was estimated to be about -1. It should be noted that this was the rounded-up estimate provided by the R software. Other correlations with $|\tau| > 0.5$ include mixed-methods and qualitative, theoretical and quantitative, as well as commentary and mixed-methods. This metric was used because a correlation greater than 0.5 is said to be significant (Akoglu, 2018).

Table 5.

Kendall's τ correlation matrix comparing the percentages of article types between journals.

	Quant	Qual	Mixed	Meta	Theo	Comm	Other
Quant	1	-0.2	0.2	0	-1*	0.2	0.4
Qual	-	1	-0.6*	-0.4	0.2	-1*	0.4
Mixed	-	-	1	0.4	-0.2	0.6*	-0.4
Meta	-	-	-	1	0	0.4	-0.2
Theo	-	-	-	-	1	-0.2	-0.4
Comm	-	-	-	-	-	1	-0.4
Other	-	-	-	-	-	-	1

*$|\tau|>0.5$

Comparison tests between journals indicate statistically significant differences in the percentages of quantitative, qualitative, theoretical, commentary, and other article types (see Table 6). The qualitative and commentary article types display the greatest variation in percentages across journals according to the results of the *z*-tests, which confirms the variation observable in Figure 6. In particular, there is statistically significant evidence for a difference in percentages of the qualitative article type between JRME and ESM, JRME and MERJ, RME and ESM, RME and MERJ, and between ZDM and JRME. There is also statistically significant evidence for a difference in percentages of the commentary article type between JRME and ESM, JRME and MERJ, ESM and MERJ, ESM and RME, RME and MERJ, ZDM and JRME, ZDM and MERJ, and between ZDM and RME.

Table 6.

Results from two-proportions z-tests displayed as matrices comparing the percentages of article types between journals.

	ESM	JRME	MERJ	RME	ZDM	ESM	JRME	MERJ	RME	ZDM
	Quantitative					Qualitative				
ESM	-	X	X	X	*	-	**	X	*	X
JRME	-	-	X	X	X	-	-	**	X	**
MERJ	-	-	-	X	X	-	-	-	**	X
RME	-	-	-	-	X	-	-	-	-	X
ZDM	-	-	-	-	-	-	-	-	-	-
	Mixed-Methods					Meta-Analytic				
ESM	-	X	X	X	X	-	I	I	I	X
JRME	-	-	X	X	X	-	-	I	I	I
MERJ	-	-	-	X	X	-	-	-	I	I
RME	-	-	-	-	X	-	-	-	-	I
ZDM	-	-	-	-	-	-	-	-	-	-
	Theoretical					Commentary				
ESM	-	**	*	X	X	-	**	**	**	X
JRME	-	-	I	*	*	-	-	**	X	**
MERJ	-	-	-	X	X	-	-	-	**	**
RME	-	-	-	-	X	-	-	-	-	**
ZDM	-	-	-	-	-	-	-	-	-	-
	Other									
ESM	-	I	I	I	X					
JRME	-	-	*	I	I					
MERJ	-	-	-	I	**					
RME	-	-	-	-	I					
ZDM	-	-	-	-	-					

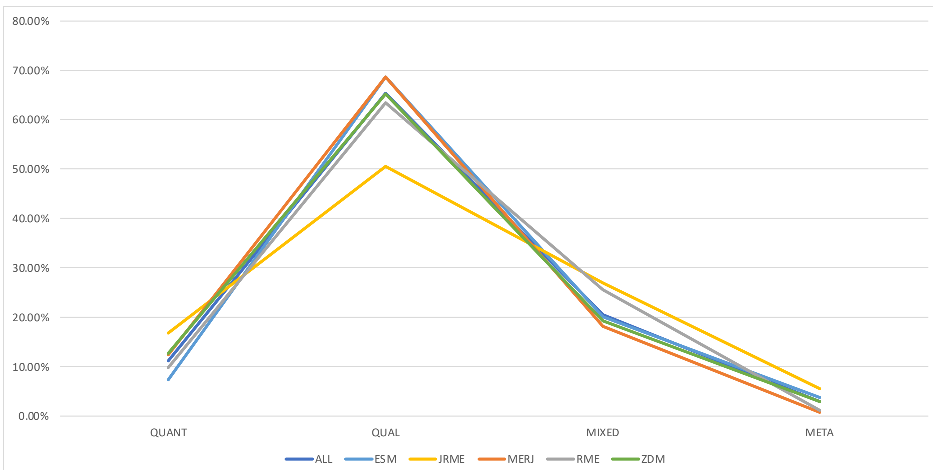
*$p<0.05$, **$p<0.01$, X = not significant, I = insufficient conditions to test

Empirical Articles (N = 1,032)

Of the 1,032 empirical articles, the numbers of the article types was 115 quantitative (~11%), 674 qualitative (~65%), 212 mixed-methods (~21%), and 31 meta-analytic (~3%). Figure 7 displays a line chart of the percentages of empirical article types by journal. As in Figure 7, we can see noticeable variation in the percentages of qualitative articles. Also, the variation in percentages of mixed-methods articles is somewhat more prominent in Figure 7 than in Figure 6. For the remaining types, there appears to be little variation across the journals.

Figure 7.

Line chart: percentages of empirical article types by journal.



The Kendall's τ correlations seen in Table 7 seem to corroborate this observation, as the correlation between the percentages of qualitative article types and mixed-methods article types was estimated to be about -0.6. Other correlations with | τ | > 0.5 include qualitative and quantitative, as well as meta-analytic and mixed-methods.

Table 7.

Kendall's τ correlation matrix comparing the percentages of empirical article types between journals.

	Quant	Qual	Mixed	Meta
Quant	1	-0.6*	0.2	0.2
Qual	-	1	-0.6*	-0.2
Mixed	-	-	1	0.6*
Meta	-	-	-	1

*| τ |>0.5

Comparison tests between journals (see Table 8) indicate statistically significant differences in the percentages of article types, and qualitative articles display the greatest variation in percentages across journals according to the results of the *z*-tests, which confirms the variation observable in Figure 7. In particular, there is statistically significant evidence for a difference in percentages of the qualitative article type between JRME and ESM, JRME and MERJ, and between ZDM and JRME. There is also statistically significant evidence for a difference in percentages of the quantitative article type between JRME and ESM, and between ZDM and ESM.

Table 8.

Results from two-proportions z-tests displayed as matrices comparing the percentages of empirical article types between journals.

	ESM	JRME	MERJ	RME	ZDM	ESM	JRME	MERJ	RME	ZDM
	Quantitative					Qualitative				
ESM	-	*	X	X	*	-	**	X	X	X
JRME	-	-	X	X	X	-	-	**	X	*
MERJ	-	-	-	X	X	-	-	-	X	X
RME	-	-	-	-	X	-	-	-	-	X
ZDM	-	-	-	-	-	-	-	-	-	-
	Mixed-Methods					Meta-Analytic				
ESM	-	X	X	X	X	-	I	I	I	X
JRME	-	-	X	X	X	-	-	I	I	I
MERJ	-	-	-	X	X	-	-	-	I	I
RME	-	-	-	-	X	-	-	-	-	I
ZDM	-	-	-	-	-	-	-	-	-	-

**p*<0.05, ***p*<0.01, X = not significant, I = insufficient conditions to test

Trends Over Time

The percentages of paper types across years displays some variation year-to-year, yet there do not appear to be any strong, monotonic trends present. Figure 8 seems to indicate a possible monotonic trend in the percentage of the commentary article type year-to-year, but the results of the Mann-Kendall time-series test in Table 9 confirms that there are no statistically significant monotonic trends year-to-year for the percentages of any of the paper types. Figure 9 displays a line chart of the percentages of only the empirical article types year-to-year.

Table 9.

Results from Mann-Kendall time-series trend tests displayed as p-values examining possible year-to-year trends in paper types.

Quant	Qual	Mixed	Meta	Theo	Comm	Other	Ex
1	0.45	0.13	0.31	0.57	0.06	0.34	0.7

Figure 8.

Line chart: percentages of article types by year.

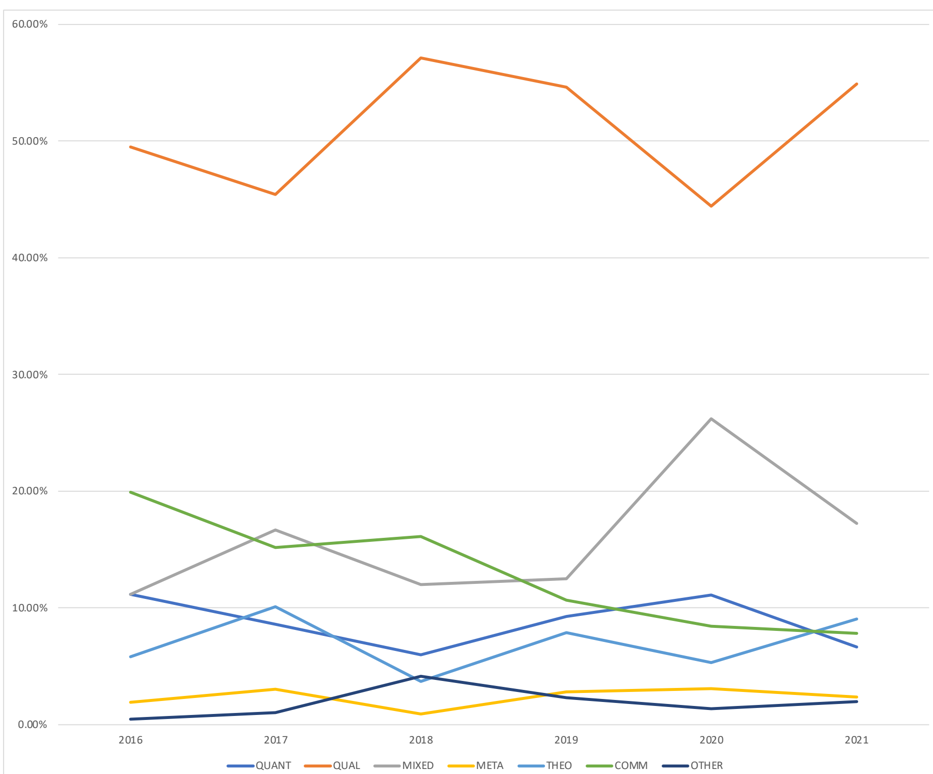
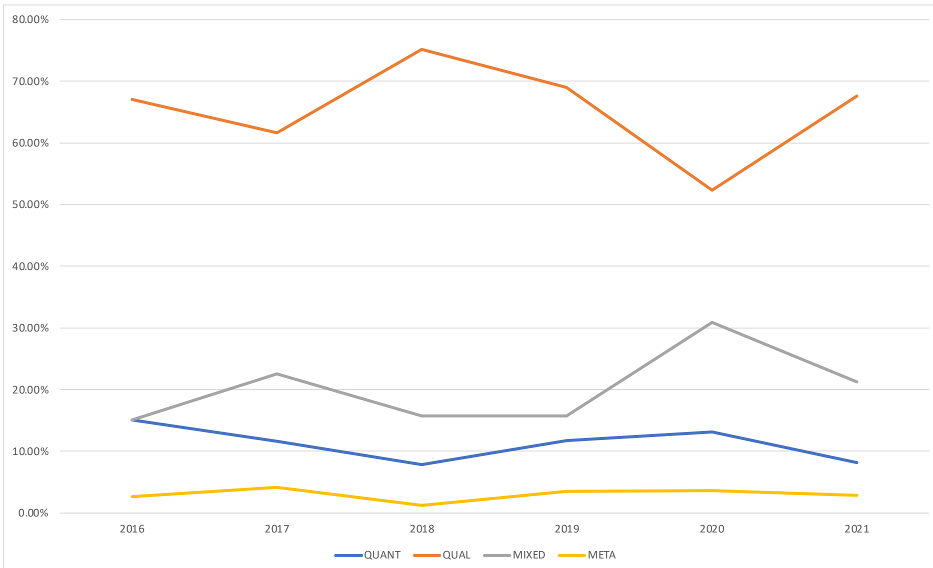


Figure 9.

Line chart: percentages of empirical article types by year.



Analysis 2 (*N* = 115)

As discussed in Chapter III, the author counted the Explicit Goal (EG) statement formats which appeared in the quantitative articles (research question, hypothesis, or statement). Of the 115 quantitative articles, the numbers of EG statement formats was found to be 24 hypothesis (~21%), 43 research question (~37.5%), 35 both hypothesis and research question (~30.5%), and 13 statement (~11%). These counts can be seen in Table 10.

Table 10.

Counts of EG statement formats in the sample of 115 quantitative articles.

Hypothesis	Research Question	Both H and RQ	Statement
24	43	35	13

There are 98 total EGs and 148 total Implicit Goals (IGs) coded in the sample of 115 quantitative articles. The number of matching pairs (see Chapter III for a description) across these articles was 91. This means about 93% of the 98 total EGs and about 61% of the 148 total IGs belong to a matching pair within the same study. The number of unmatched EGs is 7 and the number of unmatched IGs is 57. The counts can be seen in Table 11.

Table 11.

Counts of EG/IG matches in the sample of 115 quantitative articles.

	EGs	IGs
Matched	91	91
Unmatched	7	57
Total	98	148

From Table 12, the number of articles with directly linked EGs is 78 (~68%) and with directly linked IGs is 100 (~87%). This means that these articles had EGs and IGs that could be coded. In this context, "coded" means to be directly linked to the goals in Table 1. The number of articles without any directly linked EGs is 37 (~32%) and without any directly linked IGs is 15 (~13%). This indicates that these articles did not have any EGs or IGs which could be coded.

Table 12.

Counts of articles with directly or not directly linked goals in the sample of 115 quantitative articles.

	EGs	IGs
Directly Linked	78	100
Not Directly Linked	37	15

As can be seen in Table 13, the top three most frequently appearing EGs are L9 (Productive Disposition, 28, ~29% of total EGs), L2 (Procedural Fluency, 14, ~14% of total EGs), and L5 (Problem Solving, 10, ~10% of total EGs). The top three most frequently appearing IGs are also L9 (Productive Disposition, 35, ~24% of total IGs), L2 (Procedural Fluency, 32, ~22% of total IGs), and L5 (Problem Solving, 19, ~13% of total IGs).

Table 13.

Counts of EGs and IGs in the sample of 115 quantitative articles.

MEL	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	Total
Explicit	4	14	8	2	10	0	3	1	28	5	75
Implicit	4	32	9	5	19	1	3	2	35	6	116
Total	8	46	17	7	29	1	6	3	63	9	189
MET	T1	T2	T3	T4	T5	T6					Total
Explicit	1	1	2	9	6	4					23
Implicit	1	0	4	11	9	7					32
Total	2	1	6	20	15	11					55

Analysis 3 (*N* = 22)

As discussed in Chapter III, the author counted the Explicit Goal (EG) statement formats which appeared in the quantitative articles (research question, hypothesis, or statement). Of the 22 articles examined by Rittle-Johnson et al. (2017), the frequency of EG statement formats was found to be 12 hypothesis (~55%), 2 research question (~9%), 6 both hypothesis and research question (~27%), and 2 statement (~9%). The counts can be seen in Table 14.

Table 14.

Counts of EG statement formats in the sample of 22 articles examined by Rittle-Johnson et al. (2017).

Hypothesis	Research Question	Both H and RQ	Statement
12	2	6	2

There are 51 total EGs and 65 total Implicit Goals (IGs) coded in the sample of 22 articles examined by Rittle-Johnson et al. (2017). The number of matching pairs (see Chapter III for a description) across these is 46. This means that about 90% of the 51 total EGs and about 71% of the 65 total IGs belong to a matching pair within the same study. The number of unmatched EGs is 5 and the number of unmatched IGs is 19. The counts can be seen in Table 15.

Table 15.

Counts of EG/IG matches in the sample of 22 articles examined by Rittle-Johnson et al. (2017).

	EGs	IGs
Matched	46	46
Unmatched	5	19
Total	51	65

From Table 16, the number of articles with directly linked EGs is 21 (~95%) and with directly linked IGs is 22 (100%). This means that these articles had EGs and IGs which could be coded. In this context, “coded” means to be directly linked to the goals in Table 1. The number of articles without any directly linked EGs is 1 (~5%) and without any directly linked IGs is 0 (0%). This indicated that these articles did not have any EGs or IGs which could be coded.

Table 16.

Counts of articles with directly or not directly linked goals in the sample of 22 articles examined by Rittle-Johnson et al. (2017).

	EGs	IGs
Directly Linked	21	22
Not Directly Linked	1	0

As can be seen in Table 17, the top three most frequently appearing EGs are L8 (Communication, 17, ~33% of total EGs), L3 (Conceptual Understanding, 11, ~22% of total EGs), and L2 (Procedural Fluency, 10, ~20% of total EGs). The top three most frequently appearing IGs are also L8 (Communication, 17, ~26% of total IGs), L2 (Procedural Fluency, 17, ~26% of total IGs), and L3 (Conceptual Understanding, 14, ~22% of total IGs). Additionally, it is worth noting that none of the articles had EGs or IGs directly linked to the EGs of Mathematics Education Teaching (MET).

Table 17.

Counts of EGs and IGs in the sample of 22 articles examined by Rittle-Johnson et al. (2017).

MEL	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	Total
Explicit	0	10	11	0	5	0	4	17	1	3	51
Implicit	0	17	14	0	7	0	6	17	2	2	65
Total	0	27	25	0	12	0	10	34	3	5	116

MET	T1	T2	T3	T4	T5	T6	Total
Explicit	0	0	0	0	0	0	0
Implicit	0	0	0	0	0	0	0
Total	0	0	0	0	0	0	0

CHAPTER V

DISCUSSION AND CONCLUSION

> Nothing can be done about the past, but I would hope that future research would be planned to include more coordination and thus an increased efficiency for the whole research endeavor.
>
> —Begle, 1979, p. 156

Discussion

Research Question 1

Analysis 1 shows that Mathematics Education Research (MER) includes articles of several quite different types. The most striking feature of the enumeration of article types reported in this dissertation study is that, among all of the articles in the survey sample, the qualitative articles comprise over one-half of the research output. Among the empirical articles, the qualitative articles comprise nearly two-thirds. If taken together, the articles that report results of some kind of statistical analysis (quantitative, mixed-methods, and meta-analytic) constitute only a little over one-fourth of the total sample, or slightly over one-third of the empirical papers. The disparity in percentage between empirical articles in which statistical analysis is conducted and empirical articles in which statistical analysis is not conducted could in part be due to the nature of the domain of MER itself. Measurement is not always such an easy task, particularly measurement which results in quantitative data, and especially in a domain in which measurement often implies psychometrics. As it turns out, psychometrics can often be a difficult task to accomplish with a high degree of validity and reliability as compared to measurement of physical phenomena (Humphry, 2017; Salzberger, 2013).

To address RQ1a, the results indicate that there are statistically significant differences in the percentages of article types between journals, both in the total survey sample and limited to only the empirical articles. In the total sample of included articles, percentages of the qualitative article type differed significantly between all but five of the journal comparisons, while the commentary article type differed significantly between all but three of the comparisons. Furthermore, the results of the Kendall's τ correlations seem to indicate that the variation in percentages of the qualitative and commentary article types are very strongly negatively correlated, such that when the percentage of the qualitative article type goes down, the percentage of the commentary article type goes up by nearly the same amount. The same can be said for the percentages of the quantitative and theoretical article types. It is not entirely clear why this might occur, but one hypothesis is that there may be a great amount of overlap between researchers who publish the qualitative and commentary article types. Future studies could examine this hypothesis. Likewise, there may be a great amount of overlap between researchers who publish the quantitative and theoretical article types. This appears to support the "two camps" hypothesis that most researchers in MER fall into either the "qualitative camp" or the "quantitative camp." As discussed in Chapter II, Onwuegbuzie (2005) has made several compelling arguments that MER should move toward a merging of these two "camps." This would, in theory, increase the percentage of the mixed-methods article type.

To address RQ1b, in this dissertation study there were no statistically significant trends found in the percentages of paper types across time. This begs a question: what occurs in the percentages of article types across larger time periods? Hart et al. (2009) found percentages of article types across the 1995 to 2005 time period. Specifically, these researchers found percentages of 66% qualitative, 21% quantitative, and 13% mixed-methods. Or, to say this another way, they found percentages of 66% qualitative, and 34% which include results of some statistical analysis (quantitative, mixed-methods, or meta-analytic). The sample of empirical articles surveyed for this dissertation study includes 65% qualitative, and 35% which include results of some statistical analysis (quantitative, mixed-methods, or meta-analytic). It appears that the percentages have not changed much between the 1995-2005 time period and the 2016-2021 time period. What has changed is that the percentages of quantitative and mixed-methods article types have "traded off": there are fewer quantitative articles located during the present time period and more mixed-methods articles were found. Additionally, as discussed in Chapter II, Inglis and Foster (2018) reported a decrease in experimental studies in the 1968-2015 time period, and the NMAP report (USDOE, 2008) also reported a decrease in experimental studies in recent decades.

To address RQ1c, the percentages of article types might provide some indication of the Research Trajectories (RTs) in MER from a top-down perspective. The NMAP (USDOE, 2008) discusses the need for studies which allow the generalization of results in order to answer questions about cause and effect relationships. However, in order to generalize (i.e., to summarize, synthesize, or otherwise combine) results between two studies, it is desirable that there be some similarity in the data collected, measures implemented, and analytic methods employed. Also, the populations and topics of interest should be comparable. In some sense, the article type of a given article provides some indication as to whether it would be comparable with other articles of the same type. For example, it is probably more likely that two quantitative articles are comparable than it is likely that a quantitative article and qualitative article are comparable in the sense of synthesizing results. To be certain, meta-analytic studies, which select for studies that report results which can be synthesized, do not seem to include articles of dissimilar types for the purposes of the intended analyses.

ADT provides insight into these issues through RT theory. Begle (1979) urged for a greater organization of the research effort to increase the efficiency of the whole research endeavor. This research effort can be thought of through the lens of ADT as the RTs in MER. The percentages reported in Analysis 1 indicate that the RTs in MER might be somewhat disorganized in the sense that Begle discussed. As discussed above, it is not so clear that articles of differing types can be reconciled into the same RT. Furthermore, the NMAP stated the following regarding experimental studies in MER,

> To achieve these goals, the rigor and scale of the federal government's infrastructure for educational research must be dramatically increased. In particular, the nation's research portfolio should be better diversified, increasing experimental research at multiple points along a continuum from smaller-scale (less costly but highly informative) experiments to large field trials that address problems of major national importance. And, to be ready for even small-scale

> experiments, basic research and intervention development studies are needed to bring interventions and models to a point such that studying their efficacy is viable. Both smaller-scale experiments on the basic science of learning and larger-scale randomized experiments examining effective classroom practices are needed to ensure the coherent growth of research addressing important questions in mathematics education. (p. 63)

It is worth noting that experimental studies are a subset of the articles that report the results of some statistical analysis in MER (quantitative, mixed-methods, and meta-analytic), so that the percentage reported by the author in Analysis 1 of this dissertation study (35% of the empirical articles surveyed) is an overestimate of the current percentage of experimental studies in MER. With this in mind, the NMAP's findings and comments appear to still be relevant in the current time period. If the NMAP's comments above are any indication of the organization of RTs in MER that should be aimed for by researchers, then there is room for improvement in increasing the percentage of experimental studies in MER. For future investigations, ADT (via RT theory) can help to frame the discussion of the percentage of article types in MER under a theoretical framework, providing a common language for researchers to discuss these issues moving forward.

Research Question 2

It can be seen in the Appendix that there is a great variety of variables, measures, and analytic methods among the sample of 115 quantitative articles. In particular, 105 variables, 83 measures, and 61 analytic methods were observed across the sample. It should be noted that measures which were author-developed within the same article, or which did not have an acronym, were not noted in these observations. These findings seem to corroborate the findings of Begle (1979), in which he found a similar variety of variables, measures, and analytic methods. Begle noted his observation,

> One thing that struck me very forcefully in the course of reading or even skimming the literature surveyed in the preceding chapters is that the research efforts and empirical studies which have been carried out were, for the most part, uncoordinated. Very seldom have I found two studies of the same variable which used either the same measuring instrument or the same kinds of students. Almost never have I observed experiments being replicated. (pp. 155-156)

While the results of RQ1 appear to indicate that the percentages of article types have remained consistent over the last few decades, the results of RQ2 appear to indicate, along with Dr. Begle's observations, that the variety of populations, variables, and measures have also remained consistent over the last several decades. In addition, the results of RQ2 appear to indicate that calls from the NMAP (USDOE, 2008) to increase scientific rigor in Mathematics Education Research (MER) might be going unanswered.

To address RQ2a, about 93% of the total Explicit Goals (EGs) matched (to some IG of the researchers) while about 61% of the total Implicit Goals (IGs) matched (to some EG of the researchers). This seems to indicate that the researchers were rather efficient at matching their EGs to their IGs, but not as efficient at matching their IGs to their EGs. Another way to say this is that researchers tended to pursue more goals than what they explicitly stated were their goals.

To address RQ2b, about 68% of the articles had EGs which directly linked to the EGs of MER and about 87% of the articles had IGs which directly linked to the EGs of MER. This possibly indicates that about one-third of the time researchers do not explicitly link their intended inquiries to the EGs of MER, yet their implicit actions link to at least some of the EGs of MER substantially more often.

To address RQ2c, the particular EGs of MER which have been examined here are not directly pursued approximately 13%, which is close to one-eighth, of the time. This does not mean that 13% of the RTs do not pursue the EGs of MER. However, it could indicate that some of the RTs in MER do not directly pursue the EGs of MER. This might be attributed to studies which examine variables which *mediate* or *moderate* other variables which do directly link to the EGs of MER. It is also possible that the EGs of MER examined in this dissertation are not fully representative of all of the EGs which may be of interest to MER. However, this 13% figure also allows for the possibility that some of the RTs among the research literature do not pursue the goals of MER but provides no way of knowing to what extent that is the case. Overall, it is a reasonable conclusion that the great majority of RTs in MER do in fact pursue the EGs of MER, and this further seems to provide evidence that the EGs examined in this dissertation comprise many of the "main" EGs of MER, even if they are not the only EGs that may be of interest.

Although the great majority of RTs appear to pursue the EGs of MER according to the results in this dissertation, there is still the question of whether these RTs comprise organized efforts toward the EGs of MER. Pursuit of the EGs of MER is one measure of progress according to the framework of ADT, but speed of progress is another consideration, as discussed in Chapter II. If the RTs are moving in many different directions (toward differing EGs) then this may slow progress. Likewise, if the RTs are incompatible in the sense that two RTs do not collectively contribute to progress, then this also may slow the overall progress toward the EGs of MER. The results and observations gained from this dissertation study seem to indicate a large variety of populations, topics, variables, measures, and analytic methods, even within the same article type (such as quantitative, as examined in Analysis 2). This might point to the second measure of the progress of MER. The RTs in MER might be somewhat varied according to the second measure of progress, yet they satisfy the first measure of progress by mostly pursuing the EGs of MER. The first measure of progress is an important indicator, and it is good that the RTs in MER appear to largely pursue the EGs of MER as listed in Table 1. The second measure of progress is also important, and the results reported by this dissertation study in Analysis 2 appear to corroborate the findings of Begle (1979), that the variety of features in the research literature in MER appears to be large, and the findings of the NMAP (USDOE, 2008), that "the number of experimental studies in [mathematics] education that can provide answers to questions of cause and effect is currently small." (p. 63).

Research Question 3

The Appendix includes a great variety of variables, measures, and analytic methods even in the sample of 22 articles examined by Rittle-Johnson et al. (2017). In particular, there were 26 variables, 4 measures, and 12 analytic methods observed across the sample. It should be noted that measures which were author-developed within the same article, or which did not have an acronym, were not noted in these observations.

To address RQ3a, about 90% of the total Explicit Goals (EGs) matched (to some IG of the researchers) while about 71% of the total Implicit Goals (IGs) matched (to some EG of the researchers). This seems to indicate that the researchers were rather efficient at matching their EGs to their IGs, but not as efficient at matching their IGs to their EGs. Another way to say this, is that researchers tend to carry out studies in the same manner as they intend, but then they tend to do accomplish even more than they explicitly state in their methodology.

To address RQ3b, about 95% of the articles had EGs which directly linked to the EGs of Mathematics Education Research (MER) and 100% of the articles contained IGs which directly linked to the EGs of MER. This appears to indicate that the researchers almost exclusively linked both their EGs and IGs to the EGs of MER.

To address RQ3c, almost all of the Research Trajectories (RTs) among the sample of 22 articles examined by Rittle-Johnson et al. (2017) seem to pursue the EGs of MER. This seems to indicate that the first measurement of progress in MER is satisfied. However, given the great variety of populations, topics, variables, measures, and analytic methods within this sample of 22 articles (see Appendix), it is not clear that the second measure of progress is satisfied in the RTs among this sample of 22 articles. It is an important consideration that this sample of 22 articles was selected as part of a meta-analysis, which is an attempt to synthesize the results of some collection of studies. From another perspective, a meta-analysis is an attempt to consolidate, or combine, various different RTs together. Meta-analysis shows that this can be done in MER, yet even when this is accomplished, some of the features (such as populations, topics, or measures) remain varied, as seen in the results of this dissertation study.

To address RQ3d, the results of RQ2a and RQ3a are largely similar. The results of RQ2b and RQ3b are not so similar, as a much higher percentage of the articles in Analysis 3 had EGs and IGs linked to the EGs of MER than did the articles in Analysis 2. This can potentially be explained by the filtering process that the articles underwent for selection in the meta-analysis of Rittle-Johnson et al. (2017). It is also possible that there is some correlation between articles being linked to the EGs of MER and those articles being suitable for inclusion in a meta-analysis. However, it is important to consider that the author of this dissertation has only examined one meta-analysis, and so more work along these lines should be done before conclusions are drawn.

Overarching Research Question

As has been noted above, most of the articles in both Analysis 2 and Analysis 3 can be directly linked to the EGs of MER given in Table 1. This would support the contention that Table 1 includes the main EGs of MER. However, it is conceivable that there may still be other EGs or even IGs to consider in MER which are not listed there, and so that list needs to be expanded. It is also conceivable that some of the articles published in the journals examined by the author in this dissertation study that seemed not to be linked directly to the EGs of MER but might be found to be indirectly linked to the list of Table 1. For instance, a researcher might examine some variable(s) which indirectly link to some of the EGs of MER, or even variable(s) that mediate or moderate the relationship of some variable which *is directly linked* to some of the EGs of MER.

In this dissertation study, and using ADT as a theoretical lens, the author examined the progress of MER toward the EGs listed in Table 1 with two indicators in mind:

1. whether or not RTs are moving toward the Explicit Goals (EGs) of MER; and
2. whether or not the RTs are compatible in the sense that the results of studies within different RTs can be consolidated to derive more general conclusions.

In regard to indicator 1, the RTs of the articles in both Analysis 2 and Analysis 3 appear to mostly move toward the EGs of MER as listed in Table 1, with 87% and 100% of the IGs, respectively, directly linked to these EGs of MER. In regard to indicator 2, the articles in both Analysis 2 and Analysis 3 appear to display a large variety of populations, topics, variables, measures, and analytic methods. Even within the meta-analysis conducted by Rittle-Johnson et al. (2017), which is an attempt to consolidate articles (and therefore the corresponding RTs of those articles), there is a large variety of these features. It is worth noting again that the author of this dissertation study only examined one meta-analysis, and so the results here are simply a first indication in that regard. With this in mind, these findings might indicate that the RTs of MER, while largely focused on the EGs of MER, approach them slowly and in a way that induces difficulty in generalizing results to infer more general conclusions. This is the difficulty discussed by the NMAP (USDOE, 2008). This might increase the difficulty of making greater steps toward the EGs of MER by reducing the organization of RTs in MER (Begle, 1979).

Final Considerations

Through this dissertation study, the author has provided evidence that the Research Trajectories (RTs) in Mathematics Education Research (MER) are collectively, for the most part, moving toward the Explicit Goals (EGs) of MER described in Table 1. However, the author has also shown evidence for the RTs being largely incompatible in the sense that consolidating studies and generalizing results is fairly difficult. This lack of organization may lower the efficiency of progress toward the EGs of MER listed in Table 1, as argued by Begle (1979) and as the NMAP (USDOE, 2008) warned against. The large variety of populations, topics, variables, measures, and analytic methods may be some of the factors contributing to this variety in the survey sample.

The variety of populations can be attributed to the variety of populations in human beings in general. The main student populations currently researched in MER are elementary students, middle school students, high school students, and college students (including PSTs). The main teacher populations are primary grades (K-5) teachers and secondary grades (6-12) teachers. There are also non-student populations to consider, and within the student populations there are populations involving learning disabilities as well as gifted and talented populations. In addition to these, some student variables are age, gender, and ethnicity to name a few.

The variety of topics can be attributed to the variety of topics included in the domain of mathematics itself. Mathematics is always expanding and growing, and for this reason the variety of topics is always expanding and growing as well. Within even a specific topic such as fractions or algebra, there are many sub-topics.

The variety of variables can be attributed to the many different aspects of Mathematics Education Learning (MEL), Mathematics Education Teaching (MET), and Mathematics Education Research (MER) that can be measured. There are population characteristics which can be measured, for example student variables such as age. There are teaching characteristics which can be measured, such as the years of experience of a teacher. There are also the many psychological phenomena to be measured, such as student performance on a given test of knowledge or self-efficacy. Many of these variables can also be measured differently depending on the population being measured.

The variety of measures can be attributed to two main considerations: (1) the variety of populations, topics, and variables necessarily causes a variety of measures, and (2) researchers often create their own measures specific to their study.

The variety of analytic methods can be attributed to the variety of data types collected and also to the variety of different kinds of author EGs as observed in the results of the current dissertation study. Researchers choose their analytic methods based on which analytic method best fits the data they have collected.

It is unlikely that anything can, or should, be done to reduce the variety of variables, including populations and topics. However, there are some things that can, and probably should, be done to alleviate and reduce some of the current variety in order that future RTs might be more frequently comprised of organized efforts toward the EGs of MER than they currently are. In particular, measures could be developed that apply to a wide variety of populations and topics. If standardized measures are developed, then the variety of variables does not change, but the variety of methods used to collect data on those variables can be reduced. During the course of this dissertation study, the author often observed researchers taking measures from prior studies and then adapting them to their studies. However, there did not appear to be a standardized way to adapt these measures from study to study, and thus the adapted measures behaved much like different measures entirely. It may be fruitful for improving the organization of RTs in MER that researchers examine the development of standardized measures which are widely applicable to a variety of topics and populations in MER. Action Domain Theory (ADT) provides, via Research Trajectory (RT) theory, can provide a useful theoretical perspective for pursuing such developments. The pursuit of developing such measures echoes one of the calls made by the NMAP (USDOE, 2008) to develop "item and test features that improve the assessment of mathematical knowledge" (p. 63). ADT implores that the development of standardized measures be undertaken with the EGs of MER within the immediate field of view. One example of an attempt to accomplish developments of these kinds of measures would be the surveys developed by Hill et al. (2005) to measure the EGs of MET in teacher populations. Another such example might be the NAEP developed by the National Center for Education Statistics (Johnson, 1992) to measure the mathematical proficiency of K-12 students in the United States across a variety of metrics.

Future Studies

Some possible future investigations would be useful. Considering that this dissertation study is the first study that attempts to use ADT, all of the techniques here are subject to improvement and refinement. The most pertinent studies for the immediate future would be to further refine the methodology of linking EGs of MEL, MET, and MER to a given study. It is this coding methodology which allows the progress of MER to be measured, and therefore refinement of this methodology will enable future studies to better measure the progress of MER. Similarly, studies which conduct similar explorations as Analysis 2 of this study, except to instead examine the qualitative article type or the mixed-method article type. Attempts to design and implement specific RTs may also be a fruitful direction for future research. Future studies that examine the development of measures designed to examine the EGs of MER as outlined in this dissertation may also be beneficial.

Closing Remarks

In this dissertation study, the author has provided a sizable survey of the recent literature in MER. It has provided evidence for: (1) the percentages of article types in MER which builds on prior studies such as Hart et al. (2009), and (2) possible differences in the percentages of article types in MER between journals and possible trends across time. It also investigated the populations, topics, variables, measures, and analytic methods within the quantitative articles in the recent literature. It provided a similar investigation of the sample of articles examined in Rittle-Johnson et al. (2017). The theoretical framework of ADT was developed, and used as a lens to analyze these results in order to suggest possible insights regarding the current RTs in MER and to what extent these RTs are moving MER toward some EGs as organized from a review of the literature on goals in MEL, MET, and MER. This dissertation study corroborates the observations of Begle (1979) by indicating that the variety of populations, topics, variables, measures, and analytic methods in MER appears to have remained largely unchanged in the past several decades. In addition, it reverberates the findings of the NMAP report (USDOE, 2008) by indicating that the percentage of empirical studies appears to have remained low in the years since the NMAP report, and by indicating that there appears to remain a need for an increase of scientific rigor in MER.

In conclusion, through this dissertation study, the author has provided some evidence that the RTs in MER may hold variety to such an extent that the progress toward some of the EGs of MER (as listed in Table 1) may be losing efficiency (Begle, 1979). Although the RTs of MER appear to largely move toward these goals, they do so with some lack of organization. As Begle suggested, it may do well for the efficiency of the whole research endeavor to consider ways to increase the organization of RTs in MER moving forward. Recalling the recommendations of NMAP (USDOE, 2008), one possible avenue for this pursuit might be the development of measures which are standardized, developed with replicability in mind, widely applicable to a range of populations and topics, and (using ADT as a theoretical lens) developed with the EGs of MER as a top priority of measurement.

REFERENCES

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91-93.

Ball, D. L., Hill, H. H., & Bass, H. (2005). Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide? *American Educator*, *29*(1), 14-46.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, *59*(5), 389-407.

Begle, E. G. (1979). *Critical variables in mathematics education: Findings from a survey of the empirical literature*. NCTM.

Begle, E. G., & Gibb, E. G. (1980). Why do research? In R. J. Shumway (Ed.), *Research in mathematics education* (pp. 3-19). NCTM.

Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, *57*(3), 195-203.

Capraro, R. M., Capraro, M. M., & Morgan, J. R. (2013). *STEM project-based learning: An integrated Science, Technology, Engineering, and Mathematics (STEM) approach*. Brill.

Carpenter, T. P., Fennema, E., & Franke, M. L. (1996). Cognitively guided instruction: A knowledge base for reform in primary mathematics instruction. *The Elementary School Journal*, *97*(1), 3-20.

Choy, L. T. (2014). The strengths and weaknesses of research methodology: Comparison and complimentary between qualitative and quantitative approaches. *IOSR Journal of Humanities and Social Science, 19*(4), 99-104.

Clements, D. H., & Sarama, J. (2014). Learning trajectories: Foundations for effective, research-based education. In A. P. Maloney, J. Confrey, & K. H. Nguyen (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 1–30). Information Age.

Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among the five approaches* (3rd ed.). Sage.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage.

Creswell, J. W., & Plano Clark, V. L. (2014). *Designing and conducting mixed methods research* (3rd ed.). Sage.

Freudenthal, H. (1977). *Weeding and sowing: Preface to a science of mathematical education*. Springer.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research.* Aldine.

Glaser, R., Lesgold, A., & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing* (pp. 41-85). Erlbaum.

Goslin, D. A. (1963). *The search for ability: Standardized testing in social perspective*. Russell Sage Foundation.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, *11*, 255-274.

Hart, L. C., Smith, S. Z., Swars, S. L., & Smith, M. E. (2009). An examination of research methods in mathematics education (1995-2005). *Journal of Mixed Methods Research, 3*(1), 26-41.

Hilgard, E. H. (1964). A perspective on the relationship between learning theory and educational practices. *Teachers College Record*, *65*(9), 401-414.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, *105*(1), 11-30.

Hill, H. C., Rowan, B., & Ball, D. L. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Education Research Journal*, *42*(2), 371-406.

Humphry, S. M. (2017). Psychological measurement: Theory, paradoxes, and prototypes. *Theory & Psychology*, *27*(3), 407-418.

Inglis, M., & Foster, C. (2018). Five decades of mathematics education research. *Journal for Research in Mathematics Education, 49*(4), 462-500.

Johnson, E. G. (1992). The design of the national assessment of educational progress. *Journal of Educational Measurement*, *29*(2), 95-110.

Kawada, C. L., Oettingen, G., Gollwitzer, P. M., & Bargh, J. A. (2004). The projection of implicit and explicit goals. *Journal of Personality and Social Psychology, 86*(4), 545-559.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*(1/2), 81-93.

Kendall, M. G. (1976). *Rank correlation methods* (4th Ed.). Griffin.

Klinger, E. (1977). *Meaning and void: Inner experience and the incentives in peoples' lives*. University of Minnesota Press.

Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Heinemann.

Li, Y., Wang, K., Xiao, Y., & Froyd, J. E. (2020). Research and trends in STEM education: A systematic review of journal publications. *International Journal of STEM Education, 7*(11), 1-16.

Madey, D. L. (1982). Some benefits of integrating qualitative and quantitative methods in program evaluation, with some illustrations. *Educational Evaluation and Policy Analysis, 4*, 223-236.

Morton, K., & Riegle-Crumb, C. (2019). Who gets in? Examining inequality in eight-grade algebra. *Journal for Research in Mathematics Education, 50*(5), 529-554.

Moss, C. M., & Brookhart, S. M. (2019). *Advancing formative assessment in every classroom: A guide for instructional leaders* (2nd ed.). ASCD.

Mullis, I., Martin, M., Beaton, A., Gonzalez, E., Kelly, D., & Smith, T. (1997). *Mathematics achievement in the primary school years: IEA's Third International*

Mathematics and Science Study (TIMSS). TIMSS International Study Center - Boston College.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics.*

National Governors Association (NGA). (2010). *Common core state standards.*

National Research Council (NRC). (2001). *Adding it up: Helping children learn mathematics*. National Academy Press.

Neill, D. M., & Medina, N. J. (1989). Standardized testing: Harmful to educational health. *Phi Delta Kappan*, *70*(9), 688-97.

Niss, M. (1999). Kompetencer og uddannelsesbeskrivelse. *Uddannelse*, *9*, 21-29.

Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology*, *8*(5), 375-387.

Organization of Economic Cooperation and Development (OECD). (1999). *Measuring student knowledge and skills: A new framework for assessment.*

Phelps, R. (2005). *Defending standardized testing.* Taylor & Francis.

Piaget, J. (1970). *Genetic epistemology*. Columbia University Press.

Pontius Jr., R. G., & Millones, M. (2011). Death to kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, *32*(15), 4407-4429.

Rittle-Johnson, B., Loehr, A. M., & Durkin, K. (2017). Promoting self-explanation to improve mathematics learning: A meta-analysis and instructional design principles. *ZDM Mathematics Education, 49*(4), 599-611.

Salzberger, T. (2013). Attempting measurement of psychological attributes. *Frontiers in Psychology*, *4*(75), 1-4.

Sarama, J., & Clements, D. H. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. Routledge.

Schoenfeld, A. H. (2000). Purposes and methods of research in mathematics education. *Notices of the AMS, 47*(6), 641-649.

Schultheiss, O. C., & Brunstein, J. C. (2010). *Implicit motives*. Oxford University Press.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation Chicago* (pp. 39-83). Rand McNally.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, *57*, 1-22.

Sierpinska, A., Kilpatrick, J., Balacheff, N., Howson, A. G., Sfard, A., & Steinbring, H. (1993). What is research in mathematics education, and what are its results? *Journal for Research in Mathematics Education, 24*(3), 274-278.

Simon, M. A. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114-145.

Stemler, S. E., & Tsai, J. (2008). Best practices in interrater reliability three common approaches. In J. Osborne (Ed.), *Best practices in quantitative methods* (pp. 29-49). Sage.

Strohmaier, A. R., Schiepe-Tiska, A., Chang, Y. P., Müller, F., Lin, F. L., & Reiss, K. M. (2020). Comparing eye movements during mathematical word problem solving in Chinese and German. *ZDM Mathematics Education*, *52*(1), 45-58.

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466-478.

United States Department of Education (USDOE). (2008). *The final report of the national mathematics advisory panel*.

Wallis, S. (2013). Z-squared: The origin and application of χ^2. *Journal of Quantitative Linguistics*, *20*(4), 350-378.

Williams, S. R., & Leatham, K.R. (2017). Journal quality in mathematics education. *Journal for Research in Mathematics Education, 48*(4), 369-396.

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association, 22*, 209-212.

Yates, F. (1934). Contingency tables involving small numbers and the χ^2-test. *Supplement to the Journal of the Royal Statistical Society, 1*(2), 217-235.

APPENDIX

OBSERVATIONS

Table 18.

Some observed populations and topics in the sample of 115 quantitative articles.

Populations	Topics	
Adults	Mathematics Topics	Other Topics
Working Professionals	Abstract algebra	Attitudes
Incarcerated	Algebra	Beliefs
Parents	Arithmetic	EEG
Students	Calculus	Emotions
Pre-School (Pre-K)	Fractions	Eye-tracking
Elementary (Grades K-5)	Geometry	Health
Middle (Grades 6-8)	Math anxiety	Enjoyment
High (Grades 9-12)	Modelling	MRI
College	Numeracy	Representations
Pre-Service Teachers (PSTs)	Pre-arithmetic	Risk level
Primary (Grades K-5)	Probability	Salience
Secondary (Grades 6-12)	Problem posing	Self-concept
Special Needs/Learning Disability	Problem solving	Self-efficacy
Gifted/Talented	Proofs	SES
Teachers	Real analysis	Solution errors
Pre-School (Pre-K)	Statistics	Spatial reasoning
Elementary (Grades K-5)	Teaching Topics	Strategy use
Middle (Grades 6-8)	Content knowledge	Student interest
High (Grades 9-12)	Pedagogy	Student values
College	Teacher interaction	Technology
Mathematicians	Teacher noticing	

Table 19.

Some observed variables, measures, and analytic methods in the sample of 115 quantitative articles.

Variables	Measures	Analytic Methods
Achievement	AC-MT	Akaike Information Criterion
Adjective Ratings	ACT	ANCOVA
Age	AEQ	ANOVA
Algebra Preparedness	AMFT	Autoregressive cross-lagged
Anxiety	ANS	model
Appropriateness	ARS	Bayesian Information Criterion
Attitudes	ASRS	Bayesian t-test
Calculator Use	BAS	Benja- mini–Hochberg
Career interest	BESQ	procedure

Variables	Measures	Analytic Methods
Class size	BIN	Chi-Squared
Classroom management	Broad Math battery	Cluster analysis
Cognitive Ability	BSRA-3	Cohen's d
Comprehension	CDMTA	Confirmatory Factor Analysis
Computational skills	CDRT	Diagnostic classification
Confidence	CDT-MPS	Electrophysiological analysis
Content Knowledge	COEMET	Explanatory Factor Analysis
Course taken	CRVDT	Factor Analysis
Curriculum	DAT-SR	General estimation of equation
Disposition	EAC survey	model
Dwell Time	ECLS-B	General Linear Model
EEG	ECLS-K	Generalized linear mixed model
Enjoyment	EEG Machine	Hierarchical Linear Model
ENL Use	ENL Test	Hierarchical Multiple
Enrollment	Eye Tracking	Regression
Ethnicity	FICSMath Survey	Intra-class correlation
Event related potentials	fMRI	Item Response Theory
Executive function	FSMAS	Kolmogorov-Smirnov
Experience	GTCA	Kolmogorov-Smirnov *z*-tests
Explicit attention to concepts	HSC Math	Kruskal–Wallis test
Final grade	HTKS	Latent change model
First Language	LMT	Latent Growth Model
Gender	MANX	Likelihood ratio test
General Reasoning	MAQ	Linear logistic model
Goals	MARS	Linear Regression
GPA	Metacognitive Questionnaire	Linear ridge regression
Grade	MKT-G	Longitudinal structural equation
Grade Level	MM Survey	model
Health	MPOT	Mann–Whitney U test
Highest Grade	MRT	MANOVA
Income	NAEP Items	Multi-trait multi-method model
Institution Type	NAPLAN	Multilevel Linear Model
Interest	NBQL	Multilevel logistic model
Intervention	NIN	Multilevel Regression Path
IQ	NSC	Model
Item Difficulty	PACSF	Multinomial logistic regression
Item Type	PALS	Multiple Regression
Language Scores	PARCC	Odds ratios
Learning Location	PIAAC	Ordinary Least Squares
Literacy	PISA Measures	Regression
Logical Reasoning	PMI	Partial least squares regression
Math Ability	PPVT-III	model
Math Literacy	PSU	Path Analysis
Mental Rotation	PSVT-R	Pearson Correlation
Metacognition	PSVT-ROT	Principal component analysis
Misconceptions	RAPM	Probit model
Motivation	REMA	Propensity scores
Multiplicative reasoning	RMAT	Random intercept linear model
Number of Courses Taken	RPMT	Rasch Analysis
Number sense	SAT	Social network analysis
Numeracy skill	SAT-M	Spearman correlation
Numerical Ability	SBA	Spearman's rho correlation

Variables	Measures	Analytic Methods
Opportunity to learn	SCM	Standard linear mixed effects
Parent job	SDQ-III	Structural Equation Modelling
Parental Education	SEGDQ	t-test
Patterning	SFOE	Tukey honesty significant difference
Pedagogy	SIMCE	
Perception	SIT-R	Two-level measurement model
Preparation	SMA	VanderWeele model
Statement	SMI eye tracker	Welch t-test
Prior Achievement	SMK-G Assessment	*z*-test
Prior Knowledge	SON-R	
Problem posing	SRI	
Problem Solving	State standardized tests	
Procedural Shifts	Survey	
Proof Type	TEAM	
Proportional Reasoning	TED-M	
Representations	TEDI-MATH	
Response Type	TEDS-FU	
SAT Scores	TIMSS Measures	
SAT/ACT	TTNST	
School Location	WIAT	
School Type	WIPPSI-NL	
Self-regulation	WJ-III	
Self-Efficacy		
SES		
Spatial ability		
Spatial reasoning		
Storyboards		
Strategy use		
Student Approval		
Student errors		
Study habits		
Task Type		
Teacher beliefs		
Teacher expectations		
Teacher knowledge		
Teaching Experience		
Teaching Method		
Time spent		
Topic		
Verbal ability		
Visual difficulty		
Visualization		
Years in college		
Years of Explicit Names		

Table 20.

Some observed populations and topics in the sample of 22 articles examined by Rittle-Johnson et al. (2017).

Populations	Topics	
Adults	Mathematics Topics	Other Topics
Working professionals	Algebra	Cognitive load
Students	Arithmetic	Computer games
Pre-School (Pre-K)	Fractions	Computer tutor
Elementary (Grades K-5)	Geometry	Eye-tracking
Middle (Grades 6-8)	Number theory	Procedural transfer
High (Grades 9-12)	Pre-arithmetic	Representations
College	Probability	Self-explanation
Pre-Service Teachers (PSTs)	Proofs	Solution errors
Primary (Grades K-5)	Statistics	Strategy use
	Teaching Topics	Technology
	Content knowledge	

Table 21.

Some observed variables, measures, and analytic methods in the sample of 22 articles examined by Rittle-Johnson et al. (2017).

Variables	Measures	Analytic Methods
Achievement	MBQ	ANCOVA
Beliefs	PISA measures	ANOVA
Cognitive load	SRA	Chi-squared test
Conceptual understanding	Tobii eye tracker	Cohen's d
Eye tracking		Cohen's f
Game performance		Fisher's Z-transformation
Group feedback		MANCOVA
Item difficulty		Mann-Whitney U test
Knowledge access		MANOVA
Metacognition		Path model
Prior grade		Pearson's r
Problem posing		t-test
Problem solving		
Problem type		
Procedural knowledge		
Procedural transfer		
Prompting		
Proof comprehension		
Propositional knowledge		
Proving skills		
Representations		
Self-explanation		

Variables	Measures	Analytic Methods
Self-management		
Strategy use		
Time spent		
Time to test		