DEVELOPMENT OF THE RESEARCH PARADIGM INVENTORY TO MEASURE VIEWS ABOUT RESEARCH PRACTICES AND BELIEFS

Paul Parker

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
in
Educational Research and Evaluation

Gabriella Belli, Chair
Elizabeth Creamer
Gary Skaggs
Edward Wolfe

February 4, 2009
Falls Church, Virginia

Keywords: Educational Research, Paradigms, Rasch Measurement

Development of the Research Paradigm Inventory
to Measure Views about Research Practices and Beliefs
by
Paul F. Parker

ABSTRACT

The purpose of this study was to develop an instrument to measure the views of educational researchers across the dimensions that comprise research paradigms. The development of the instrument, entitled the *Research Paradigm Inventory* (RPI), was conducted in the hope of providing a mechanism for future research that will enable the examination of prospective links between the ontological, epistemological and methodological orientations of researchers and graduate students who will serve as future inquirers in the field of education.

During its development and validation, various aspects of the RPI were examined in accordance with the validity framework outlined by Samuel Messick (1989). According to Messick (1989), the validity of measure interpretation and corresponding action can be examined in terms of *content, substantive, structural, generalizability, external* and *consequential* forms of evidence. During this study, the content aspect of validity was addressed through the creation of instrument specifications and the development of items that were mapped to those specifications. In addition, the content aspect of validity was addressed by selecting items that were reviewed by experts, pilot tested, field tested and exhibited high technical quality. The substantive aspect of validity was addressed through an analysis of item rating scale functioning, person fit, item difficulty hierarchies and relationships among instrument scale measures. The structural aspect of validity was addressed through a confirmation of the instrument's dimensionality. The generalizability aspect of validity was addressed through an analysis of person reliability, the precision of item/person parameter estimates and item calibration invariance.

The development activities and analyses described above resulted in the creation of six subscales measuring: (1) Realism in Research, (2) Research Objectivism, (3) Quantitative Methodology, (4) Relativism in Research, (5) Research Interpretivism and (6) Qualitative Methodology. Given the evidence collected, these scales appear to provide reasonably reliable and defensible estimates of individuals' attitudes toward various research practices and beliefs, and should be appropriate for future research studies exploring educational research paradigms.

DEDICATION

Dedicated

to

my wife

Jill

and our children

Nolan and Madelyn

ACKNOWLEDGEMENTS

First and foremost, I would like to thank each of the members of my committee. You all are excellent examples of what it truly means to be an educator. Thank you Dr. Belli for your willingness to take a chance and admit me into the program. Thank you also for your guidance, continued honesty and support, and encouragement throughout this process. Thank you Dr. Wolfe for your extraordinary willingness to donate your time, expertise and consultation to me. It was certainly more than I could have ever hoped or expected. Thank you Dr. Creamer for opening my eyes up to newer ways at looking and learning about the world, and for sharing with me ideas that have only furthered my desire to be a lifelong learner. And thank you Dr. Skaggs, for providing me with a welcomed and rewarding introduction to concepts and skills that will hopefully guide me in the rest of my career. I appreciate everything you all have done for me, as all of you have contributed to my life in ways you can never imagine.

To my friends at The Lab, whose countless hours of conversation and training help to remind me that I can still challenge myself in life without losing sight of the things that really matter. Thank you Rob LeBlanc, at an early age you showed me what it really means to be a friend.

To my family, thank you to my father whose means were not great, but who worked every day of his life to ensure a better one for his children. Thank you to my mother, who reminds me that God is with me no matter what. Thank you Dan, you have been my best friend throughout my life and shown me what a brother is supposed to be.

To my wife Jill, you are the love of my life. I cannot begin to express how thankful I am to you for giving me two beautiful children, and for the many sacrifices I know you have made for me. I only hope that in the years to come I can be the husband that you deserve. And finally, thank you Nolan and Madelyn. You are my most precious gifts. You have already given me more happiness than I could have ever dreamed. My only wish is that each of you can grow to be like the people I have thanked here. God Bless you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

The purpose of this study was to develop an instrument to measure the views of educational researchers across the philosophical and methodological dimensions that comprise research paradigms, and provide evidence of validity to support an appropriate interpretation and use of its resulting measures. In this specific context, the term paradigm refers to the fundamental set or system of beliefs that serve to guide and inform the exploration of educational research questions, issues and phenomena (Guba & Lincoln, 1994). The development of the instrument, entitled the *Research Paradigm Inventory* (RPI), was conducted in the hope of providing a mechanism that will enable the examination of prospective links between the philosophical and methodological orientations of current researchers and graduate students who will serve as future inquirers in the field of education.

In recent years, a great deal of attention has been paid to how scientific research in education is defined and conducted (ESRA, 2002; NCLB, 2002), as well as to what constitutes quality in that regard (Hostetler, 2005; Lather & Moss, 2005). In the literature, quality often refers to an inquirer's ability to adhere to specific research traditions, but then also to opinions about what constitutes credible evidence for inquiry-based assertions and the philosophical foundation undergirding various research approaches. The consideration given to questions about quality by educational researchers has helped to catalyze the emergence of alternative perspectives and exploratory archetypes for examining issues, where certain research communities have shifted from a traditional "positivist" and/or "empiricist" position, sometimes referred to as the "received view" (Guba & Lincoln, 1994), to a more milieu-dependent positioning of "interpretivism" (Howe, 1998; Nisbet, 2005; Tsai & Liu, 2005). Consequential to the inauguration and enlargement of these alternative ways of knowing, coupled with efforts to legislatively define scientific research (ESRA, 2002; NCLB, 2002), practitioners of nascent approaches and even those adhering to more "traditionalist" approaches often seek recognition and perhaps hegemony within particular research and academic communities. Investigators with diverse philosophical orientations often view the purpose of their research differently, incorporate the use of different methods and apply an assortment of standards to evaluate the quality of research studies.

Principal differences between philosophical views have given rise to a variety of perspectives for research in education, where the cogitative orientation and tacit assumptions of advocates to any particular approach often vary as much as individual disciplines. In the past, a number of scholars have offered up the concept of paradigmatic distinctions between research communities as a possible source for this diversity (Creswell, 1998; Guba & Lincoln, 1994; Schwandt, 1989). Contemporary discussions surrounding paradigms in educational research typically embrace, but then extend beyond the topic of methods (i.e., research design, data collection and analysis) to include deliberations on specific philosophical propositions, most notably existence, knowledge and values (Bahm, 1974). If one interprets a paradigm to mean a convergence of shared beliefs among a community of researchers (Johnson & Onwuegbuzie, 2004), where "philosophy and methods intersect to determine what kinds of evidence one finds acceptable" (Patton, 2002, p. 571), then it may be useful to assess how distinctions within philosophical belief systems (e.g., ontology and epistemology) relate to other types of beliefs (e.g., methodology) in order to understand the role these important factors have for shaping researchers' perspectives and their individual approaches to inquiry.

Guba and Lincoln (1994, 2005) suggested researchers explore issues in accordance with an "inquiry paradigm" that reflects their underlying philosophical perspectives. Strongly related to, but argumentatively extending beyond the concept of "philosophy of science", Guba and Lincoln (1994) defined inquiry paradigms as "the basic belief systems or worldviews that guide the investigator, not only in choices of method, but in ontologically and epistemologically fundamental ways" (p. 105). Guba and Lincoln (2005) proposed that inquirers' beliefs may be represented by a *positivist*, *post-positivist*, *critical theorist*, *constructivist* or *participatory research* paradigm, suggesting also that the particular paradigm in which a researcher is situated would manifest in the answers he or she is likely to give to three fundamental questions:

1. "What is the form and nature of reality?" (the ontological question);
2. "What is the nature of the relationship between the knower or would-be-knower and what can be known?" (the epistemological question); and,
3. "How can the inquirer go about finding out whatever he or she believes can be known?" (the methodological question) (Guba & Lincoln, 1994, p. 108).

Guba and Lincoln (1994, 2005) did not believe that inquiry paradigms are restricted to methodology, where dissenting perspectives might represent perceived incompatibility between

the quantitative and qualitative traditions. Rather, they contended inquiry paradigms demonstrate interdependency among ontology, epistemology and methodology. Intertwined within each of the aforementioned inquiry dimensions is the concept of *axiology* (the philosophical discipline concerned with value and quality), where moral forces, either intrinsically or extrinsically based, feed into the inquiry process by helping to define what questions are being asked, what paradigmatic choices will serve to guide an investigation, what types of methods will be incorporated and the manner in which research-based assertions are made accessible (e.g., in the presentation of 'voice' in a research report) (Guba & Lincoln, 2005, p. 197).

Guba and Lincoln (2005) noted readers who are familiar with the literature on methods and paradigms "reflect a high interest in ontologies and epistemologies that differ sharply from those undergirding conventional social science" (p. 191). In addition, they claimed, "new young professionals being mentored in graduate schools are asking serious questions about and looking for guidance in qualitatively oriented studies and dissertations" (2005, p. 191), perhaps because these individuals feel disenfranchised by traditionalist approaches. Given the prominence of research in doctoral programs of education, an analysis of its stakeholders' varying perspectives on these issues warrants consideration and attention. Further, and possibly even more salient, is to suggest that the conduct and interpretation of any research should include a firm understanding of the theoretical foundation on which it is based, the methods that are used, as well as the requisite assumptions that one must presumably accept in order to make reasonable assertions about a particular topic.

Guba and Lincoln (1994) asserted, "no inquirer ought to go about the business of inquiry without being clear about just what paradigm informs and guides his or her [own] approach" (p. 116), while others have stressed "researchers must understand the basic ethical, ontological, epistemological and methodological assumptions of each and be able to engage them in dialogue" (Denzin & Lincoln, 2000, p. 162). Schnelker (2006) elaborated this view, arguing that an appreciation of paradigmatic distinctions would provide "a framework for understanding *why* researchers do what they do, resulting in more critical researchers and consumers of research" (p. 43). Although these propositions reflect a high degree of ideology, especially in cases that relate to what researchers 'must' know or how they 'ought' to behave, consideration for them leads to a number of interesting questions, such as:

- What are the philosophical views of researchers?

- To what degree are researchers familiar with their own philosophical orientation?
- What are the relationships between paradigmatic dimensions?
- To what extent do researchers' belief systems and behaviors adhere to hypothesized paradigmatic structures?
- What effect do philosophical predilections have on the development of a researcher?

In order to answer questions such as these, it is essential to have valid and reliable measures of the dimensions that comprise research paradigms. Throughout the literature, instruments exist that purportedly measure beliefs about the nature of science (Nott & Wellington, 1986), attitudes about reality (Unger, Draper & Pendergrass, 1986), epistemological belief systems (Erwin, 1983; Royce and Mos, 1980; Schommer, 1990; Schraw, Bendixen & Dunkle, 2002; Tsai & Liu, 2005), attitudes toward research (Papanastasiou, 2005) and attitudes toward specific methodological traditions such as quantitative methods (Chang, 1996). Unfortunately, in some cases, these instruments either fail to exhibit consistency regarding the composition of important elements within the aforementioned paradigm framework (e.g., epistemology) or fail to tap the dimensions that comprise research paradigms in the appropriate context for the referent populations (i.e., current researchers and graduate students). As such, the creation of the RPI should serve as a necessary first step in the process of exploring research paradigms by providing fundamental measures of the dimensions that comprise them in the appropriate context.

In addition to addressing ideological questions and catalyzing researcher self-reflection, the RPI might serve a more pragmatic role in the edification of graduate students who will serve as future researchers in the field of education. For example, an instrument that delivered valid and reliable measures of the referent concepts could be used to identify philosophical predilections of graduate students and help guide them toward coursework and research projects in which they might flourish. The RPI could also be used as a research tool for examining the views of current researchers and graduate students, possibly by measuring change across the life of an educational program or a career, and/or in the exploration of the variables that contribute to the development of one's research approach altogether. Given these points, and in congruence with segments of the paradigm framework presented by Guba and Lincoln (1994), the RPI was

created to identify the ontological, epistemological, and methodological orientation of researchers and graduate students.

Framework for Instrument Development

To help guide the creation of the RPI, aspects of its development were examined in accordance with the validity framework outlined by Samuel Messick (1989, 1995). Past depictions of validity have focused on three distinct types of substantiation: (a) *content-related* forms – where developers provide evidence that the content of an instrument is representative of the universe from which it was theoretically drawn; (b) *criterion-related* forms – where developers attempt to demonstrate concurrent or predictive relationships between an instrument's measures and external measures of the same or a similar construct; and, (c) *construct* related forms – where developers attempt to provide evidence that theories about a construct adequately account for instrument-based observations (Cronbach, 1971 as cited by Wolfe & Smith, 2007a). In lieu of this fragmented conceptualization, Messick (1989) suggested validity be reconsidered unitarily within the concept of construct validity, which subsumes content-relevance and representativeness, as well as criterion-relatedness (p. 17).

Messick (1995) defined validity as "an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment" (p. 741). His portrayal of validity as a unified concept illustrated a discernable modification to previous notions of the construct and how it has come to be viewed within many academic communities. In his view, unified validity embraced all forms of evidence and integrated content, criteria and consequences into a construct framework that allowed for the empirical testing of "rational hypothesis about score meaning and theoretically relevant relationships" (Messick, 1995, p. 741). When reviewing his comments, it becomes clear that validity is not to be interpreted as a property of any instrument in question, but rather a characteristic of the scores obtained from that instrument, the persons responding, and the context of the assessment itself (Messick, 1995). As such, the most important consideration for building a strong validity argument is how scores from an instrument are interpreted and then used (Messick, 1995).

According to Messick (1995), the validity of measure interpretation and corresponding action should be examined in terms of *content, substantive, structural, generalizability, external*

and *consequential* forms of evidence. Taken collectively, these aspects serve as "general validity criteria or standards for all educational and psychological measurement (Messick, 1995, p. 747). In their review of Messick's (1989) framework, Wolfe and Smith (2007a) interpreted the six aspects of validity as follows:

> The *content* aspect of validity addresses the relevance and representativeness of the content upon which the items are based and the technical quality of those items. The *substantive* aspect refers to the degree to which theoretical rationales relating to both item content and processing models adequately explain the observed consistencies among item responses. The *structural* aspect of validity appraises the fidelity of the scoring structure to the structure of the construct domain. The *generalizability* aspect concerns the degree to which score properties and interpretations generalize to and across population groups, settings, and tasks, as well as the generalization of criterion relationships. The *external* aspect of validity includes what has traditionally been termed convergent and discriminant evidence generated from multitrait-multimethod comparisons, but the external aspect also concerns criterion relevance and the applied utility of the measures. Finally, the *consequential* aspect appraises the value implications of score interpretation as a basis for action as well as the actual and potential consequences of test use, especially in regard to sources of invalidity related to issues of bias, fairness, and distributive justice (p. 205-206).

Through the application of guidelines presented by Wolfe and Smith (2007a, 2007b) for instrument development using a Rasch measurement framework (Rasch, 1960), several of the aforementioned aspects to validity were explored during the course of this study to develop the RPI and provide evidence of validity for the interpretation and use of its resulting measures. This document outlines these steps through the presentation of two research articles and a summary that describes the instrument development process for the RPI. Chapter Two of this document is comprised of a manuscript for a journal article that provides a brief overview of literature on paradigmatic concepts, as well as arguments for how their continued consideration might be used for the betterment of the educational research community. Chapter Three provides a detailed overview of the instrument development process for the RPI, such as how items were constructed, tested and refined for inclusion on a draft form of the instrument. Chapter Four of this document is comprised of a manuscript for a journal article that describes the steps that were taken to develop the RPI in accordance with the validity framework outlined by Messick (1995, 1999). More specifically, this manuscript includes a brief introduction about the enlargement of theory relating to the instrument's content (i.e., a review of key concepts presented in Chapter

Two), an overview of the development process for the items used to operationalize that theory (i.e., a review of concepts presented in Chapter Three), and a presentation of information detailing the psychometric properties of the RPI such as item technical quality, instrument dimensionality, rating scale functioning, reliability, the precision of item/person parameter estimates and evidence of measurement invariance for RPI scales. And finally, Chapter Five of this document provides a discussion about this project, areas in which the RPI might be applied, ways in which the validity of RPI measure use will continue to be explored, and personal reflections about how paradigmatic concepts are currently viewed in contemporary educational research.

CHAPTER 2

MANUSCRIPT 1

An Overview of Literature on Educational Research Paradigms

Abstract

This article provides a brief overview of the literature on educational research paradigms. The article begins by detailing the various ways in which paradigms have been considered in educational research, as well as social science research as a whole. Next, by using an adaptation outlined by Guba and Lincoln (2005), the views of respective research communities are highlighted to illustrate some examples of paradigms that help to structure contemporary social science inquiry. Issues about the compatibility between research paradigms and their respective components are then discussed, as well as the critiques various communities have for alternative positions. Finally, through the use of examples, it is reasoned that both educational and social science research can benefit by continuing to focus on research paradigms, such as in the provision of a framework for understanding different perspectives (including language), comparing researchers from across disciplines and fostering collaboration between members of scholarly communities with similar or divergent views.

Keywords: Educational Research, Paradigms, Methodology

Introduction

In recent years, a great deal of attention has been paid to how educational research is defined and conducted (ESRA, 2002; Eisenhart & Towne, 2003; NCLB, 2002), as well as to what constitutes quality in that regard (Hostetler, 2005; Lather & Moss, 2005). In the literature, quality often refers to the efficacy of research or an inquirer's ability to adhere to specific research traditions, but then also to opinions about what constitutes credible evidence for inquiry-based assertions and the philosophical foundation undergirding various research approaches. Consideration for questions about quality has helped catalyze the emergence of alternative perspectives and exploratory archetypes for examining issues in education, where certain research communities have shifted from a traditional "positivist" position, sometimes referred to as the "received view" (Guba & Lincoln, 1994), to a more milieu-dependent positioning of "interpretivism" (Howe, 1998; Nisbet, 2005; Rabinow & Sullivan, 1979; Tsai & Liu, 2005). Investigators with diverse philosophical orientations often view the purpose of their research differently, ask different research questions, incorporate the use of different methods and apply an assortment of standards to evaluate the quality of research studies.

Following the landmark publication of Thomas Kuhn's *The Structure of Scientific Revolutions* and its utilization of the term "paradigm" (Kuhn, 1962/1996), many scholars began offering up the concept of paradigmatic distinctions between research communities as a possible source for this diversity (Creswell, 1998; Guba & Lincoln, 1994; Morgan, 2007; Schwandt, 1989). Discussions surrounding research paradigms typically embrace, but then extend beyond the topic of method (i.e., research design, data collection and analysis) to include deliberations on specific philosophical propositions, most notably existence, knowledge and values (Bahm, 1974). Contemporary views within the philosophy of science hold that all scientific theories, and by extension researchers themselves, assume a position on these matters (Bahm, 1974; Johnson, Germer, Overton & Overton, 1988), which in turn help to structure the paradigms to which inquirers adhere, either deliberately or tacitly.

Although many researchers "rarely have the time or inclination to assess what they do in philosophical terms" (Guba & Lincoln, 1994, p. 117) and proceed through an investigation without the necessity of long existential consideration (Patton, 2002), Guba and Lincoln (1994) emphasized, "no inquirer ought to go about the business of inquiry without being clear about just what paradigm informs and guides his or her [own] approach" (p. 116). Recent articles (e.g.,

Kim, 2003; Ponterreto, 2005; Schnelker, 2006) appear to uphold this ideology in support of self-awareness through the provision of outlines for understanding and making sense of research paradigms. By presenting a brief overview of literature on paradigmatic concepts, this article is an attempt to preserve this discussion so that researchers can begin to recognize their own positions, as well as the philosophical parameters anchoring the work of others (Ponterreto, 2005), for the purpose of understanding and engaging them in dialogue (Guba & Lincoln, 2005). Given this, the central questions guiding this discussion are as follows:

- What are research paradigms?
- Which paradigms structure contemporary social science inquiry?
- Are various research paradigms and their components compatible?
- How might a continued focus on paradigms help to enhance the efficacy of educational and/or social science research?

Paradigms

Consequential to Kuhn's (1962/1996) popularization of the term paradigm, its concepts have become prevalent as a means to summarize researchers' beliefs about the acquisition, creation and/or preservation of knowledge. In contemporary social science inquiry, the term paradigm has grown to take on several meanings, emblematic of the assortment of ways in which Kuhn applied the idiom throughout his original discourse (Masterman, 1970). Mindful of this multiplicity, Morgan (2007) cited four examples for how paradigms have been conceptualized in the social sciences: (1) worldviews (i.e., ways of experiencing and thinking about the world); (2) epistemological stances (i.e., belief systems about knowledge that influence how research questions are asked and then answered); (3) shared beliefs among members of a specialty area (i.e., a research community's accord about which questions are most meaningful and the procedures for answering those questions); and lastly, (4) exemplars for research (i.e., models for how research is conducted in a given field).

Albeit distinct, Morgan (2007) noted these conceptualizations were not mutually exclusive, but rather nested and representative of a hierarchy of ideals ranging from the specific to the general. For example, exemplars for how to conduct research arise from a community of scholars with shared beliefs about the importance of specific research questions and how to go about studying them. Beliefs about research topics and methods tend to align with specific

commitments to epistemological stances, which summarize researchers' assumptions about what can be known and the position one assumes when going about knowing. And finally, at the broadest level, epistemological assumptions serve as important components to structuring a researcher's worldview (Morgan, 2007). Interestingly enough, the type of hierarchical integration Morgan described more than likely represents what Stake (1993) had in mind when he stated, "there are worlds within worlds…each with its own paradigm" (*personal communication* as cited by Guba and Lincoln, 1994, p. 117).

In the present tense, the term paradigm refers to "a set of interrelated assumptions about the social world that provide a philosophical and conceptual framework for the organized study of that world" (Filstead, 1979, p. 334). At this most general level, research paradigms represent an ideological nexus, or the point at which philosophy and methods intersect to determine the kinds of evidence one finds acceptable (Patton, 2002, p. 571). Notable among the various schema (Burrell & Morgan, 1979; Creswell, 1998; Guba, 1990; Morris, 1999) that have been used to explain research paradigms is the framework outlined by Guba and Lincoln (1994, 2005), which has been discussed at length (Morgan, 2007; Morris, 1999; Paul & Marfo, 2001; Ponterreto, 2005; Schnelker, 2006), most likely due to its brevity and the manageability of its content, as well as for its utility in translating viewpoints across different perspectives.

Inquiry Paradigms

Guba and Lincoln (1994, 2005) suggested researchers explore issues in accordance with an "inquiry paradigm" that reflects their underlying philosophical perspectives. Strongly related to, but argumentatively extending beyond the philosophy of science, Guba and Lincoln (1994) defined inquiry paradigms as "the basic belief systems or worldviews that guide the investigator, not only in choices of method, but in ontologically and epistemologically fundamental ways" (p. 105). Guba and Lincoln (1994) claimed inquiry paradigms demonstrate an inherent link between the views one has about reality and existence (ontology), the relationship between the knower and what can be known (epistemology) and the ways one goes about gathering and analyzing evidence (methodology). Intertwined within each of these are axiological considerations (the philosophical discipline concerned with values and quality), where moral forces, either intrinsically or extrinsically based, feed into the inquiry process by helping to define what questions are being asked, what paradigmatic choices serve to guide an investigation, what types

of methods are incorporated and the format in which research assertions are made accessible (Guba and Lincoln, 2005).

Guba and Lincoln (1994) argued that inquiry paradigms defined for an inquirer "what it is they are about, and what falls within the limits of legitimate inquiry" (p. 108). Although strongly indicative of views within the major branches of philosophical belief, Guba and Lincoln (1994) did not believe inquiry paradigms should be dismissed as such. In their view, inquiry paradigms encapsulated and then transcended philosophical beliefs to have "important consequences for the practical conduct of inquiry, as well as the interpretation of findings and policy choices" (Guba & Lincoln, 1994, p. 112).

Guba and Lincoln (2005) presented a list of paradigms in which they believed researchers may be deliberately or tacitly positioned: *positivism, post-positivism, critical theory, constructivism,* and *participatory research*. Conspicuously absent from the framework they originally outlined were those axioms associated with participatory research (Guba & Lincoln, 1994). Guba and Lincoln (2005) elected to include participatory research in a revised framework following the important contributions of Heron and Reason (1997). They stated this decision was an excellent example of the "hermeneutic elaboration so embedded in [their] own view, constructivism" (Guba & Lincoln, 2005, p. 192). Their decision was interesting in that participatory research is rarely featured in the quantitative tradition and "at the most general level, four major interpretive paradigms structure qualitative research; positivist, post-positivist, constructivist-interpretivist, and critical (Marxist, emancipatory) and feminist-post structural" (Denzin & Lincoln, 2005, p. 22). Because of this, some would argue that participatory research reflects more of a methodological line of reasoning, with philosophical underpinnings sufficiently captured by alternative paradigms. However, in the spirit of encompassment, participatory research is mentioned here briefly, but will not be described in any greater detail, so that readers of this article can review its elements at a future time and judge for themselves the veracity of its proponents' claims to being a distinct paradigm in its own right (e.g., Heron & Reason, 1997).

Scientific Revolutions and Paradigmatic Shifts

Paradigms help research communities establish the necessary parameters for their traditions, which in turn help to define their avenues for inquiry and create the mechanisms by

which those positioned within a respective paradigm become and then remain informed (e.g., textbooks, research journals, training programs). In the original version of *The Structure of Scientific Revolutions*, Kuhn (1962) pondered whether the social sciences had attained such paradigms (p. 15). Although contemporary views about paradigms in the social sciences differ from what Kuhn had originally envisioned, the abundance of literature on the topic helps to illustrate the answer to this question is a resounding *yes.*

In his discourse, Kuhn (1962/1996) outlined the circumstances under which "scientific revolutions" take place and stimulate "paradigmatic shifts," which result in the emergence of newer paradigms. Kuhn (1962/1996) described these revolutions as, "tradition-shattering complements to the tradition-bound activity of normal science" (p. 6). As portrayed, scientific revolutions occur when an anomaly in an established paradigm interrupts and then challenges the conventions for current practice (e.g., Albert Einstein's Relativistic worldview represented a transition from Newtonian physics). According to Kuhn (1962/1996), scientific revolutions such as these are often problematic for dominant paradigms because new assumptions require the reconstruction of prior assumptions and a reassessment of concepts formerly held as 'facts'. Scientific revolutions and the paradigmatic shifts that follow can be quite overwhelming for conventionalists in this sense because they protract from practice and lay the groundwork for questioning the fabric of some hegemonic community, including its ideology. The stalwart communities ascribing to the dominant paradigm ultimately resist these shifts, resulting in competition between paradigms and significant deliberation regarding their compatibility. With these concepts in mind, the following sections provide a brief account of some of the paradigms evident in social science research and the scientific revolutions that have occurred therein, where it is proposed that at least one paradigmatic shift has taken place with the enlargement of post-positivism, and then possibly a second with the ascension of interpretivist schools of thought.

Traditionalist Paradigms: "The Received Views"

Positivism

Positivism has long been regarded as the grandfather of paradigms in scientific inquiry. As an organized branch of philosophical thinking, true positivism developed during the *Enlightenment Period* of the 17th and 18th centuries, wherein scholars of the time (e.g., Descartes, 1637/1968; Locke 1689/1956) began to think beyond the despotic influence of authoritarian

decree and instead initiated focus on the centrality of the individual, the world as objectively knowing and the use of language, including numerical language, as the conveyor of "truth" (Gergen, 2001; Ponterreto, 2005). The term positivism was originally coined by Auguste Comte (1830/1988) and used to describe a belief in "positive knowledge" generated by verifiable claims based directly on experience (Patton, 2002). In his discourse, Comte (1830/1988) described three stages of development for the human mind: the theological, the metaphysical and the positive. Comte was especially concerned with distinguishing empirical-based knowledge from theology and metaphysics, "which depended on fallible human reason and belief" (Patton, 2002, p. 92). In essence, the positivist ideals promulgated by Comte sought to eradicate superstition, purify knowledge and create a society in which erudition was to be generated through science, logic and empirical observation.

Messick (1989) described a view of positivism where "logic and syntax [were] critical to structuring knowledge of the world and that meaning is factual verifiability" (p. 22). Guba and Lincoln (1994) expanded this view within their conception of a positivist inquiry paradigm, suggesting positivists adhered to a belief in an independent, yet accessible reality (i.e., a realist ontology), which researchers explore in search of truth. Analogous to research in the "harder" sciences, positivists are thought then to maintain a dualist epistemological positioning to control for the contamination of research conclusions (i.e., threats to validity). Guba and Lincoln (1994) suggested positivists relied heavily on the hypothetico-deductive method, where hypotheses are formulated to explain phenomena by showing those phenomena follow deductively from a hypothesis (Mulaik & Lawrence, 1995). Consequently, any generalizations, correlations or prediction ns that are induced by data are thought to be discovered, rather than created, because they are implicit in the data (Fitzgerald & Cunningham, 2002), as opposed to being subject to the senses, values and biases of an investigator. Given these characteristics, positivism is often linked exclusively to quantitative methodology (Berg, 2001; Guba & Lincoln, 1994; Merriam, 1998). However, as will be described shortly, not all scholars agree with the exactitude of that relationship.

The Post-Positivist Shift

In the latter part of the 20th century, the positivist belief in absolute truths fell on hard times in a "skeptical postmodern age when honoring multiple perspectives and diverse points of

view [had] gained ascendancy" (Patton, 2002, p. 91). Dissatisfied with an absolutist view of the certainty of knowledge and other aspects of the positivist stance (Ponterreto, 2005), challengers to positivism argued that human intellectual mechanisms were inherently flawed, life's phenomena were basically intractable, and therefore, one can never fully capture a true reality (Guba & Lincoln, 1994; Ponterreto, 2005). As such, researchers across multiple disciplines began to question positivist notions and the ability to confirm facts through direct experience and observation.

Popper (1959) helped to undermine the positivist view of verifiability in his observation that theoretical propositions can never be determined to be true on the basis of experience (Mulaik & Lawrence, 1995). Throughout his discussions regarding the tenability of positive knowledge, Popper (1959) sought to distinguish valid science from other endeavors through the promulgation of a falsification criterion, arguing that, "the mark of science is to formulate potentially falsifiable propositions and put them to the test" (Mulaik & Lawrence, 1995, p. 122). Popper's (1959) *principle of falsifiability* was well illustrated by the example provided by Guba and Lincoln (1994): "Whereas a million white swans can never establish with complete confidence the proposition that all swans are white, one black swan can completely falsify it" (p. 107). In other words, while an inquirer might believe that they are justified in some conclusion, "that justification can be defeated by acquiring further relevant information" (Pollack, 1986, p. 16). Consideration for these distinctions helped set the stage for the ascension of a post-positivist paradigm and a belief in an "independent" reality that is only partially accessible and modeled, at best, probabilistically (i.e., critical realist ontology).

Guba and Lincoln (1994) suggested that post-positivism represented an adaptation of positivism, where the philosophical views that comprise each paradigm "differ very little" (Guba, 1990, p. 23). They intimated that the post-positivist movement illustrated a type of "damage control" and an effort by positivists to "respond in a limited way to the most problematic criticisms" of its positions, while still "remaining within the same set of basic beliefs" (Guba & Lincoln, 1994, p. 109). In contrast, post-positivists would argue their beliefs represent a significant departure from positivism, especially with regard to "truth", a term not found in post-positivist research (Morris, 1999). Post-positivists accept the notion that research and theory are, in part, value-laden and context dependent. In fact, post-positivists would say that it is precisely because inquirers are biased that they adopt "objective methods that are open to

scrutiny and challenge" (Charney, 1996, p. 570) so that others can review the authenticity and replicability of their actions and assertions. Under the post-positivist paradigm then, both the quantitative and qualitative traditions are often admissible, so long as certain aspects of rigor are met (Johnson & Onwuegbuzie, 2004). The primary goal for a post-positivist using these methods is to reason from the results of a study back to the overall tenets of theory (Morris, 1999) in order to build an edifice of knowledge that converges on a probable reality by explaining phenomena common to subject areas. Knowledge gained through research is then built upon other knowledge in an effort to present research conclusions that are probably authentic (but always subject to falsification), rather than an absolute certainty (Guba & Lincoln, 1994, p. 110).

The Ascension of Alternative Paradigms: Interpretivist Views

During closing decades of the 20th century, another "quiet revolution" began within social science research where strong counter-pressures against the received view emerged and broadened the repertoire of possibilities for exploring issues and topics (Guba & Lincoln, 1994, 2005; Ponterreto, 2005; Schnelker, 2006). Guba and Lincoln (1994) outlined some of the major issues that critics of the received view used to challenge its legitimacy for inquiry, based on the questions that those positioned within alternative paradigms are likely to ask. Included among the issues cited by Guba and Lincoln (1994) were positivism's and post-positivism's perceived stripping of context, exclusion of meaning and purpose, disjunction of grand theories with local context, inapplicability of general ideas to individual cases, exclusion of a discovery dimension in inquiry, theory-ladenness of facts, under-determination of theory, value-ladenness of facts and a failure to account for the interactive nature of an inquirer-inquired based dyad. In conjunction with these points, researchers operating within alternative paradigms suggested that the criteria used to judge the quality of research in the received communities were irrelevant to their work because they silenced too many voices (Guba & Lincoln 2005). Consequently, these researchers sought different, and in their view, more apposite methods for conducting research and evaluating the efficacy of their work. The birth of these new perspectives (e.g., critical theory and constructivism) suggested that alternative paradigms had come of age and begun to demonstrate legitimacy that was "well established and at least equal to the legitimacy of received and conventional paradigms" (Guba & Lincoln, 2005, p. 191).

Critical Theory

Critical theory is an appellation reserved for research traditions that challenge the status quo in an effort to expose oppression and empower the oppressed to work toward egalitarian transformation (Ponterreto, 2005, p. 129). The emergence of the criticalist tradition is often linked to Max Horkheimer, Theodor Adorno and Herbert Marcuse, and the contributions they made while at the University of Frankfurt's Institute of Social Research (Creswell, 1998). Though they themselves never claimed to have "developed a unified theory" (Ponterreto, 2005, p. 130), their work demonstrated a critical approach to inquiry that sought to catalyze change in society, as opposed to simply understanding or explaining it.

"Variants of critical theory abound in all of the social science disciplines" (Creswell, 1998, p. 80), so its adherents are quick to note the injustice of implying a false unity or consensus across all of its conceptualizations when none exists (Kincheloe & McLaren, 2005). Considering there is not any single critical theory, the critical tradition is perpetually changing and broadening, and there is frequently disagreement among critical theorists themselves, Kincheloe and McLaren (2005) cautioned, "critical theory should not be treated as a universal grammar of revolutionary thought objectified and reduced to discrete formulaic pronouncements or strategies" (p. 304). However, in an effort to "focus on the underlying commonality among critical schools of thought", Kincheloe & Steinberg (1997, as cited by Kincheloe and McLaren, 2005) provided the following definition of a criticalist:

> … a researcher or theorist who attempts to use her or his work as a form of social or cultural criticism and who accepts certain basic assumptions; that all thought is fundamentally mediated by power relations that are social and historically constituted; that facts can never be isolated from the domain of values or removed from some form of ideological inscription; that the relationship between concept and object and between signifier and signified is never stable or fixed and is often mediated by the social relations of capitalist productions and consumption; that language is central to the formation of subjectivity (conscious and unconscious awareness); that certain groups in any society and particular societies are privileged over others and, although the reasons for this privileging may vary widely, the oppression that characterizes contemporary societies is most forcefully reproduced when subordinates accept their social status as natural, necessary or inevitable; that oppression has many faces and that focusing on only one at the expense of others (e.g., class oppression versus racism) often elides the interconnections among them; and, finally, that mainstream research practices are generally, although most often unwittingly, implicated in the reproduction of systems of class, race and gender oppression. (p. 304).

Given these tenants of a criticalist ideology, critical race theory, queer theory and more recently post-structural, post-modern, critical pedagogy, feminism and cultural studies theory are often cited as examples of belief systems that are indicative of an adherence, either purposive or unintended, to the criticalist tradition (Denzin & Lincoln, 2000, 2005; Ponterreto, 2005). According to Guba and Lincoln (1994), a critical theorist's ontological positioning is one of "historical realism", where reality is considered only within the social, political, economic, ethnic or cultural climate in which an issue or phenomenon resides. Reality for the criticalist is shaped then not only by the context in which phenomena occur, but also throughout the course of time. Whereas traditionalists seek to describe or explain phenomena, critical researchers "view their work as the first step toward forms of political action that can redress the injustices found in the field site or constructed in the very act of research itself" (Kincheloe and McLaren, 2005, p. 305). Never satisfied with "merely increasing knowledge" (Horkheimer, 1972 as cited by Kincheloe & McLaren, 2005), the emancipatory principle that drives critical research "is committed to engaging oppressed groups in collective, democratic theorizing" (Denzin, 1994, p. 509) about their joint and diverse perceptions of repression and privilege (Ponterreto, 2005).

Critical theorists not only acknowledge the presence of values in the components of their research (purpose, methods, etc.), they are somewhat driven by it. Critical researchers "announce their partisanship in the struggle for a better world" (Kincheloe and McLaren, 2005, p. 305) by adopting an ontological and epistemological positioning that essentially blurs the separation between the two. Critical theorists are therefore often seen as activists or change agents with interests that are considered to be interwoven, if not inseparable, to that which is being investigated. As noted by Kincheloe and McLaren (2005), "critical researchers enter into an investigation with their assumptions on the table, so no one is confused concerning the epistemological and political baggage they bring with them to the research site" (p. 305-306). Because the "investigator and the object of investigation are assumed to be interactively linked so that 'findings' are *literally created* as the investigation proceeds" (Guba & Lincoln, 1994, p. 111), critical theorists are generally seen to use dialogic and dialectical methods of study within the qualitative tradition (Guba & Lincoln, 1994).

Constructivism

Constructivism represents a set of philosophical beliefs, which are based on the premise that persons construct an understanding of the world in which they live by using "fundamentally built-in cognitive principles" to organize and interpret experience (Airasian & Walsh, 1997; Heylighen, 1997). Constructivists claim their views represent a noticeable departure from positivism, especially in how those positioned within the paradigm interpret reality and existence (Guba, 1990). Given the belief that "human claims about nature cannot be independent of inside-the-head processes of the knowing subject" (Kant, 1781/1966 as cited by Hamilton, 1994, p. 63), constructivists suggest that knowledge of reality is constructed within the mind of an individual, rather than apprehended through the study of some external, singular and independent entity (Hansen, 2004). This belief demonstrates constructivists' adherence to a relativist ontology, where reality is apprehended, "in the form of multiple, intangible mental constructions, socially and experientially based, local and specific in nature (although elements are often shared among many individuals and even across cultures), and dependent for their form and content on the individual persons or groups holding the construction" (Guba & Lincoln, 1994, p. 110).

In stark contrast to positivist and post-positivist attempts at eradicating personal perspective through perceived systems of control, constructivists see themselves as part of the construction and/or reconstruction process, acting as a facilitator to consensual understanding between the researcher and some object of study (Guba & Lincoln, 1994, 2005). Although comparable phenomena may be observed across different populations, cultures, etc., knowledge gained within one context is not necessarily transferable to another. Even though patterns and themes across various settings may gravitate toward consensus or some notion of truth, allowances are always given to new interpretations as communication and understanding between an inquirer and the object of study proliferates (Guba & Lincoln, 1994). The goal of a constructivist then is to build an understanding of one's own mental constructions and foster the incorporation of others' in the "formation of ever more informed and sophisticated constructions" (Guba & Lincoln, 1994, p. 114). This is turn provides context for considering the implications those constructions and/or reconstructions have for their lives and interactions with others (Patton, 2002, p. 96).

Patton (2002) suggested that because constructivists view the human world differently than the physical world, they must study it differently (Patton, 2002). As such, constructivism is

sometimes seen as "the primary foundation and anchor for qualitative research methods" (Ponterreto, 2005, p. 129). These distinctions are best exemplified through the constructivist's use of hermeneutical and dialectical methods, as well as in their application of alternative criteria (e.g., trustworthiness and authenticity) for evaluating the efficacy of research.

Incompatibility Thesis?

Thus far, it has been proposed in this article that social science research has experienced at least one major scientific revolution, which resulted in a paradigmatic shift (i.e., the ascension of post-positivism), and then possibly a second resulting in rivals to the dominant paradigm (i.e., the ascension of interpretivism). As predicted by Kuhn, these revolutions were not effortless, or for that matter, resolved to everyone's satisfaction. In a not so distant past, issues about methods and how to interpret research results were a controversial topic for social science researchers. Branded the "paradigm wars" (Gage, 1989), these debates initiated around methods (e.g., quantitative vs. qualitative), but were expanded to include deliberations on conceived differences between epistemological paradigms (Howe, 1988). Some scholars (e.g., Shaeffer and Serlin, 2004) believe that social science research has moved beyond these divisive arguments, especially as it relates to the topic of methods, and is now beginning to recognize the value of multiple approaches (e.g., Fitzpatrick, Sanders & Worthen, 2004). Others have simply heeded Dewey's advice and abandoned the debate for the purpose of intellectual progress (Dewey, 1981). Nevertheless, theoretical purists have been less inclined to let go of their positions and continue to debate the appropriateness of combining epistemological paradigms.

Guba and Lincoln (1994) maintained a belief in an innate inseparability between researchers' ontological, epistemological and methodological orientations. Ercikan and Roth (2006) alluded to this notion of connectedness in their description of the popularly held view that the quantitative and qualitative traditions correlate with objectivist and subjectivist epistemologies respectively (p. 17). However, Ercikan and Roth (2006) questioned the value of making a "polarized" distinction between the quantitative and qualitative traditions and suggested the philosophical views of objectivism and subjectivism, which are believed to underpin these traditions, were "neither accurate nor useful" (Ercikan and Roth, 2006, p. 14). Because Ercikan and Roth (2006) believed that all phenomena are quantitative and qualitative at the same time; data construction processes follow similar interpretation processes for all

educational research; and, for most constructs that researchers are interested in, data construction processes are based on subjective, defensible judgments, they suggested research approaches should be considered across a continuum, where each tradition differed only by degree (p. 18) and in accordance with the research the questions being asked. In their view, an integrative framework that put research questions first would encourage investigators to join expertise and work together toward reaching solutions to problems and promote the conduct of quality research (Ercikan and Roth, 2006). Although their call to "transcend dichotomies" is promising and supported by this author, it still leaves open the matter of how philosophical assumptions (e.g., ontological views) impact the selection and structure of those research questions, if at all.

At one time, Guba & Lincoln (1994) suggested competing paradigms in the social sciences were seeking hegemony by jockeying for position as the paradigm of choice for guiding and informing inquiry (p. 105). The matter of competing paradigms they described is important because it brings to light issues about the commensurability between points of view, or a lack thereof, and the implication that divergence has for practice. Kuhn's exposition on scientific revolutions described them as "noncumulative" developmental episodes in which an elder paradigm is replaced by an emerging one that is "incompatible" with the former (Kuhn, 1962/1996, p. 92). According to Kuhn (1962/1996), rival paradigms have to be incompatible because "the assimilation of either a new sort of phenomenon or a new scientific theory must demand the rejection of an older paradigm" (p. 95). If this were not the case, then the revolutions described by Kuhn would be replaced by a type of recurring integration, resulting in an elusive meta-paradigm that "renders the older, accommodated paradigms not less true, but simply irrelevant" (Guba & Lincoln, 1994, p. 116).

In conjunction with the "incompatibility thesis*"* (Howe, 1988), Guba and Lincoln (2005) suggested making a one-to-one correspondence between certain paradigms (e.g., positivism and constructivism) at the philosophical level is impossible, rendering them incommensurable. However, in the application of methods, they concede mixing strategies "may make perfectly good sense" (Guba & Lincoln, 2005, p. 200). It appears then that questions about the commensurability between paradigms are best addressed at the specific level in which paradigms and/or their constituents are considered. For instance, at the worldview level, researchers are likely to demonstrate a tendency toward either a realist or a relativist ontology, while at the methodological level, "pragmatism" in the spirit of Charles Sanders Pierce, William James and

John Dewey (Johnson & Onwuegbuzie, 2004) may rule in the adoption of strategies that most appropriately address the research question at hand. This is not to suggest that ontology is not an important factor in determining how one goes about structuring research questions. Rather, this is to say that depending on which questions are being asked, either quantitative and/or qualitative strategies may be important and admissible, or possibly even "superior under different circumstances" (Johnson & Onwuegbuzie, 2004, p. 14).

As regards the tenability of methods and their compatibility with other paradigmatic dimensions, Patton (2002) believed that in real world practice methods could be separated from the epistemology out of which they emerged (p. 136). Johnson and Onwuegbuzie (2004) expounded upon Patton's position, stating that "the logic of justification (an important aspect of epistemology) does not dictate what specific data collection and data analytic methods researchers must use" (p. 15). In this, they emphasized epistemological beliefs should not preclude a researcher from using methods more closely associated with alternative perspectives or theories generated by those operating within alternative paradigms. This type of methodological eclecticism certainly appears to open the door to new opportunities for researchers, especially with regard to methods, but only if a researcher's core epistemological beliefs do not deny the acceptability of evidence produced by those methods.

Yanchar and Williams (2006) acknowledged that proponents of "methodological eclecticism" have helped initiate "a useful move beyond rigid adherence to traditional paradigms and a one-method-fits-all approach" (p. 3). However, they also claimed that advocates to this approach are only "half-correct", in that methodological eclecticism fails to take "seriously the inescapable assumptions and values that accompany the use of methods and the pursuit of practically usefully results" (p. 3). Consequently, they endorsed "a softer version of the incompatibility thesis" for both paradigmatic purists and advocates of methodological eclecticism in which contextual sensitivity, creativity, conceptual awareness, coherence and critical reflection should be emphasized in research and evaluation practices.

In later years, Guba and Lincoln (2005) appeared to soften a bit, as regards their beliefs about commensurability, in support of a view that "the boundaries between paradigms are shifting", resulting in a prophecy of "blurring genres" becoming rapidly fulfilled (Geertz, 1988, 1993 as cited by Guba and Lincoln, 2005). These acknowledgements apparently represent a manifestation of what their framework was originally intended to embody (i.e., a natural

connection between philosophical positions, as opposed to a system in which a particular positioning restricted the acceptance or use of alternative approaches, such as mixing research methods). The concept of blurring genres especially rang true in their discussion of values, where they saw the "expansion of basic issues to include axiology" as a way of achieving greater "confluence among various interpretivist inquiry models" (Guba & Lincoln, 2005, p. 200). However, the opinions they expressed were noticeably based on commonalities they believed to exist between interpretivist paradigms (e.g., critical theory and constructivism) and not necessarily between interpretivist and traditionalist perspectives. As such, Guba and Lincoln (2005) continued to support a clear distinction between paradigms at the comprehensive philosophical levels, such as an incongruity between the realist/objectivist and relativist/interpretivist positions.

Rather than considering paradigms as worldviews or epistemological stances, Morgan (2007) endorsed a re-connection with the version of paradigms he believed more closely mirrors what Kuhn had originally envisioned and preferred. According to Morgan (2007), social science research may be on the verge of another paradigmatic revolution, or a shift from what he calls a "metaphysical" account of paradigms (e.g., Guba and Lincoln, 2005) to an adaptation reflective of a community of researchers with shared beliefs about which questions are most meaningful and which procedures are most appropriate for answering those questions (Morgan, 2007, p. 53). To support his argument, Morgan (2007) cited three anomalies he believes exist within the metaphysical paradigm that a "pragmatic alternative" would address: (1) how to define paradigms; (2) whether those paradigms were incommensurate; and, (3) the extent to which metaphysical assumptions actually guide research. By endorsing a pragmatic approach in lieu of a metaphysical account of paradigms, Morgan (2007) promotes such concepts as shared "lines of action", "warranted assertions" and the "workability" of strategies (p. 66). In his words, "this means giving up on the assumption that there is some external system [i.e., a metaphysical paradigm] that will need to explain our beliefs to us" (Morgan, 2007, p.66). The consequence of this shift, he asserts, would be a greater focus on shared meanings and joint action across research communities. In his proposal, Morgan is clear to not endorse a complete abandonment of the tenets of a metaphysical paradigm, such as the importance of worldviews and epistemological issues. Rather, he is simply rejecting any privileging of ontological assumptions within a paradigmatic taxonomy and the restriction it argumentatively places on applied practice.

Because scientific revolutions are nearly invisible during their advance (Kuhn, 1962/1996, p. 140) and history is necessarily a retroactive analytic endeavor, it remains to be seen whether Morgan's appeal to the social sciences will actually come to fruition. Because of this, the balance of this article will continue to focus on a philosophical account of paradigms, which Morgan (2007) himself called "the currently dominant version" (p. 51). That said, I will still look with great optimism on his suggestion to deny any argument, paradigmatic or otherwise, that seeks to determine *a priori* limits for establishing meaningful communication between research communities.

Paradigm Legitimacy

Many scholars give credence to a philosophical account of paradigms and the propositions encompassed by it (Guba & Lincoln, 2005; Morris, 1999; Ponterreto, 2005; Toma, 1996; Yu, 2001), an acknowledgement presumably based on "adherence to what a paradigm provides" (Kuhn, 1970; Masterman, 1970). By way of example, Morris (1999) suggested theories derived from research reflect either a deliberate or tacit commitment to specific philosophical propositions, such as existence, knowledge, and values (Bahm, 1974). As described, these philosophical distinctions are formally compartmentalized within the referent framework as ontology, epistemology, and axiology, with methodology being the natural mechanism through which an inquirer goes about learning that which can be known (Guba & Lincoln, 1994, 2005).

Although acceptance may be given to the philosophical composition of research paradigms, there is still a great deal of controversy surrounding the legitimacy of certain paradigms themselves. As Morgan (2007) noted, this may be due to a lack of clarity or consensus about "what constitutes a paradigm within social science research" and "who gets to define and label the paradigms that are included in that list" (p. 60). For example, Morris (1999) questioned the authenticity of the classification scheme presented by Guba and Lincoln (1994) and its representation of positivism. Citing the work of Popper (1957, 1963, 1972) and Laudan (1977), Morris (1999) refuted the inclusion of positivism in any modern formulation of research paradigms and denigrated the manner in which Guba and Lincoln (1994) portrayed its relationship with post-positivism. Although Guba and Lincoln (2005) continued to include positivism in their representation of inquiry paradigms, Morris (1999) claimed new discoveries

and “changes in the nature of science’s understanding” challenged Comtian principles about the certainty of knowledge, most often associated with positivistic views (p. 19). The insurgency of these new ideals is well exemplified by the scientific revolutions in physics that began with Albert Einstein (Morris, 1999) and culminated in the Copenhagen conception of the uncertainty of quantum mechanics, where the relationship between the position and momentum of a subatomic particle, such as an electron, remains uncertain (Heisenberg, 1991).

The transformation of perspectives in the physical sciences was important in that it had profound implications for such fundamental notions as causality and the determination of future behaviors (Cassidy, 2006). Although these scientific revolutions seemed to echo through the realm of the physical sciences, Morris (1999) argued that some within the social sciences had been left behind, possibly ignoring John Stuart Mill’s (1843/1906) plea for the social sciences to mirror their older “harder” cousins. Again by citing the work of Popper (1957, 1959, 1963) and Laudan (1977), Morris (1999) claimed the scientific revolutions and paradigmatic shifts that occurred in the physical sciences had effectively destroyed “the polemics associated with notions of a positivistic social science” (p. 12). According to Morris (1999), the only place where positivism still existed was in the critiques of social scientists that operated under alternative research paradigms (p. 12), even though most philosophers of science had universally rejected its axioms for at least half a century (Feldman, 1998). Ironically, even Guba and Lincoln (1994) appeared to abandon an absolutist stance, stating that, “the naïve positivist position of the sixteenth through nineteenth centuries is no longer held by anyone even casually acquainted with these problems” (p. 116).

A review of Yu’s (2001) exposition on the relationship between positivism and quantitative methods helps to further the demise of positivist beliefs in any absolute sense. Past literature (Berg, 2001; Lincoln & Guba, 1994; Merriam, 1998) has often linked positivism exclusively to quantitative methodology. However, Yu (2001) cautioned researchers against making this assumption, noting that the terms are not synonymous and to presuppose as much would be an oversimplification of their respective traditions. Yu’s (2001) argument especially holds true in the context of null hypothesis testing, where concepts are never truly verified, as well as in measurement theory where theoretical constructs are implied, but not verified, to be causally responsible for observed behaviors across an instrument’s collective measures (Yu, 2001, p. 14). Given then the implausibility of Comtian conceptions about the certainty of

knowledge, many now contend that positivism as a research tradition is "dead" and should not be included in any modern formulation of inquiry paradigms (Jackson & Jeffers, 1989; Messick, 1993; Morris, 1999; Toma, 1996; Yu, 2001). In many respects, and understandably so, the only place where positivism still exists is in the language of those who use its features as criteria against which their own views about research can be compared and distinguished (e.g., Guba and Lincoln, 2005).

In his review of the paradigmatic framework proposed by Guba and Lincoln (1994), Morris (1999) also questioned the necessity of the constructivist paradigm across certain positions, as well as the internal consistency of philosophical beliefs within the paradigm itself. Morris (1999) suggested that the points used to distinguish constructivism from the received views of positivism and post-positivism (i.e., the theory-ladenness of facts, theory cannot be proven, value-ladenness of facts) were fully subsumed by post-positivism and/or critical theory, thereby reducing constructivism's ability to distinguish itself from other paradigms. Morris (1999) doubted the strong relativist position of constructivism, noting, "there exists no set of criteria by which one can judge the veracity of any construction" (p. 21). He suggested that a researcher must accept either that phenomena exist outside of mental constructions and cultural lenses, or the notion of cross-cultural universals. In either case, he deemed this an insoluble problem for the constructivist notion of relativism and the belief that multiple realities hold equal merit (Morris, 1999, p. 22). In his view, if the only things that can be known are one's own constructions and "the findings of an inquiry [are] not a report of what is out there but the residue of a process that literally creates them" (Guba, 1990, p. 26), then any attempt to "construct someone else's construction seems to be the most futile pursuit of all" (Morris, 1999, p. 23).

Critiques of alternative paradigms have certainly not been limited to constructivism, where philosophical views for the critical theorist represent one of the more controversial topics for the structure of the paradigm, especially regarding how each philosophical dimension within the paradigm is interpreted. For instance, Ponterreto (2005) suggested critical theorists adhere to ontology similar to that of a constructivist, wherein reality is advocated only in relation to some social, historical or political context. In contrast, Morris (1999) argued that critical theorists advocate the realist ontology of a traditionalist (p. 17). Morris (1999) based his assertion on what he believed to be a criticalist assumption of the existence of social forces that exert a consistent influence over the population and may be seen as dominant predictors or principle motivators of

human behavior. In his view, criticalists must accept their "veiled" view of reality because it is only by uncovering "true" reality that social change will be made possible (Morris, 1999, p. 19). In other words, one must determine "what is" because it holds dramatic consequences for how one engages "what ought to be" (Gadamer, 1989 as cited by Kincheloe and McLaren, 2005).

When evaluating critiques such as constructivists' assessment of the received view (e.g., Guba and Lincoln, 1994) and post-positivists' assessment of alternative views (e.g., Morris, 1999), there are two important points to consider. First, Guba and Lincoln (1994) presented their framework with the understanding that the views they expressed were in fact their own and "not open to proof in any conventional sense" (p. 108). In their opinion, any given paradigm simply represents "the most informed and sophisticated view that its proponents [had] been able to devise" (Guba & Lincoln, 1994, p. 108). Second, these critiques were made by individuals who assumed an etic positioning that resided outside of the paradigm in question. Albeit argumentative, I suggest that many of these critiques reflected efforts to discount rival paradigms, where they may have been inadvertently misrepresented, as opposed to an attempt to understand or appreciate them. It goes without saying that critiques of alternative perspectives in the social sciences, theoretical or otherwise, have always been extensive and not always without merit. That stated, the likelihood of any paradigmatic community converting an alternative school of thought to their own will always be questionable, if not impossible, because when two scientific schools disagree they inevitably "talk through each other when debating the relative merits of their respective paradigms" (Kuhn, 1962/1996, p. 109). As noted by Kuhn (1962/1996), in the circular arguments that follow, each is "shown to satisfy more or less the criteria that it dictates for itself and to fall short of a few of those dictated by its opponents" (p. 110).

Pitfalls to Reconciliation

Seeing the need to create joint criteria for evaluating the quality of social science research, some scholars have endorsed the exploration of similarities between communities of researchers (Howe, 1992; Johnson & Onwuegbuzie, 2004) and the development of standards that apply to multiple approaches (Kamill, 1995; Munby, 2001). As an example, Howe (1992) suggested researchers adopt consistent principles that are "broad and abstract enough" to encompass multiple perspectives such as the identification of grounded research questions, applying practices that define a theoretical approach, grounding research in credible background

knowledge and ensuring the overall warrant of conclusions and interpretations (p. 251). Howe's suggestions seem commonsensical enough and possibly even realized through recent efforts to legislatively define scientific research (ESRA, 2002; NCLB, 2002). However, before those types of principles can be upheld beyond authoritative mandate or legislative initiative, and endorsed by all social science research communities, which is currently not the case (Denzin & Lincoln, 2005, p. 8-9), there must be some semblance of dialogue between communities of researchers beyond arguments about the supremacy of their own paradigm.

Perceived incommensurability between paradigms has a great deal to do with researchers exploring only those studies and practices that are close to their own (i.e., operating within their paradigm) (Green, 1992) and an even greater deal to do with language, where the vocabularies of "different genres of inquiry have grown and become increasingly complex" (Paul & Marfo, 2001, p. 532). The product of this esoteric complexity has been a failure of meaning and concepts from one perspective to accurately translate or be wholly understood in another (Phillips, 1978). Presented here for heuristic purposes, an immoderate case of a failure to communicate may be exemplified in how different research communities define a "powerful analysis" (Schaeffer & Serlin, 2004), where a post-positivist using quantitative methods might be concerned with their ability to generalize a statistic to the population from which it was theoretically drawn and a phenomenonologist using qualitative methods would be concerned with their ability to provide a thick, rich description of some occurrence from an informant's perspective. Granted, there are instances when analogs transverse different perspectives, but are technically dissimilar enough to require unique labels (e.g., the quantitative concepts of internal validity, external validity and reliability loosely correlate with the qualitative concepts of credibility, transferability and dependability respectively). However, if the decision to create a unique term or language is based on some artificial rationale such as, because "reliability and validity are rooted in positivist perspectives then they should be redefined for their use in a naturalistic approach" (Golafshani, 2003, p. 597), then the consequence is an unmistakable impediment to the productive exchange of ideas across various perspectives and different research approaches (Kamill, 1995).

Because all social science research attempts to make reasonable assertions about human beings and their behaviors in relation to their environments (Biesta & Burbules, 2003), it seems likely that communication between researchers from different paradigms is possible, even if

those respective paradigms are seen as incompatible. In response to philosophers who claim Kuhn's remarks on incommensurability mean "incommensurable theories cannot communicate with each other at all" (Kuhn, 1962/1996, p. 198-199), Kuhn rejoins, "what the participants in a communication breakdown can do is recognize each other as members of different language communities and then become translators…taking the differences between their own intra- and intergroup discourse as itself a subject for study" (p. 202). Considering his proposal, it is suggested that an exposure to and continued focus on research paradigms from either the philosophical or practical viewpoint may provide a platform for constructive communication between research communities, as well as an understanding, if not acceptance, of views deemed disparate to one's own. With this in mind, the following paragraphs provide a few examples for where paradigms have been considered in research practice, as well as suggestions for how those conceptualizations might be used for the betterment of educational and/or social science research altogether.

Exemplars for Paradigm Fit and Application

During a qualitative interviewing process, Toma (1996) found that a sample of 22 legal scholars expressed assumptions that conformed to the social science typology outlined by Guba and Lincoln (1994). Because his sample was comprised of individuals from a discipline outside of traditional social science, Toma (1996) believed discovering these similarities may provide support, at a general level, for a common framework (i.e., inquiry paradigms) that avails comparisons among individuals from different disciplines, while also fostering "the interdisciplinary and substantive borrowing that reshapes and expands knowledge" (p. 35). Of note was Toma's inability to confirm the existence of a positivist paradigm during the course of his analyses, which resulted in collapsing positivism and post-positivism into a single category he entitled "legal realism." Given that the legal scholars he interviewed appeared to reject the notions of unequivocal discovery and absolute truths, his observation appears to again provide support for the assertion that a shift away from positivism into post-positivism has occurred, where there may not be any positivists among social scientists at all, only post-positivists (Toma, 1996, p. 34).

Several factors have been posited to be influential in the development of the philosophical views that structure an individual's paradigm such as the conventions promulgated

by respective disciplines (Kamill, 1995; J. Green, 1992; Green, Breitzin, Leininger, & Stauffer, 2001), the traditions of any given academic department, the reflective predisposition of a researcher and the level of familiarity a researcher has about epistemological issues altogether (Paul & Marfo, 2001). Because each of these factors appears to be grounded in an educational process, including a discussion on paradigms during the training of researchers might prove to be a worthwhile endeavor. By way of example, Schnelker (2006) found that doctoral students who were exposed to paradigmatic concepts were able to readily acquire a basic understanding of the major philosophical concepts associated with those distinctions. Consequently, she suggested using paradigmatic distinctions as a framework for qualitative research courses and then "discussing the philosophical assumptions that underlie the paradigms currently evident in social science research may contribute to clarifying many of the distinctions still debated in the literature" (Schnelker, 2006, p. 55). Even more promising, Schnelker (2006) suggested that an appreciation of paradigmatic distinctions would provide "a framework for understanding *why* researchers do what they do, resulting in more critical researchers and consumers of research" (p. 43).

Not unlike Schnelker, Tsoi-Hoshman (1989) believed that including discussions about paradigms in the education of students would promote epistemic development and facilitate the integration of theory, research and practice. Paul and Marfo (2001) developed this vision further by making the case for expanding research curriculum to include emphases on the nature of inquiry, including the epistemological, moral and aesthetic foundations of knowledge and knowing (p. 543). Paul and Marfo (2001) believed that "the continued socialization of students in the view that good science is necessarily quantitative would leave future researchers ill prepared to participate in informed critique or collaborative inquiry with other researchers who work within different paradigms of knowledge" (p. 543). As such, they argued for requiring doctoral students to take coursework on philosophical issues in preparation for, or as a supplement to, technical coursework such as statistics and research design. Such requirements, they suggested, would help to create an "atmosphere of interdisciplinary and multi-paradigmatic collaborative research that provides an informed context for students to experience and practice the values fostered by such coursework" (Paul & Marfo, 2001, p. 525).

On the matter of collaboration, Creamer (2003) found that researchers who worked together often shared many of the same attitudes about the nature of knowledge and how it can

be accumulated. Creamer's work was interesting in that it was only after she began a qualitative interviewing process she imagined "differences in epistemological assumptions might explain differences in the processes used by collaborators" (Creamer, 2003, p. 452). Creamer (2003) also found that some researchers who appeared to share the same research paradigm often employed strategies that were considerably more idiosyncratic than what could be explained exclusively by the philosophical views they held. This is not to say that the framework proposed by Guba and Lincoln (1994) was entirely inapplicable in this instance. To the contrary, it more than likely helped to illustrate again that there are worlds within worlds, each with its own paradigm and that inquiry paradigms can be considered "apt [only] as broad strokes" (Guba & Lincoln, 1994, p. 117). Further, it may suggest that the hypothesized structure of paradigms at the ontological and epistemological levels is justifiable, but then possibly unreasonable at another (e.g., methodology).

Closing Remarks and Suggestions for Future Research

This article provides an overview of the many ways in which research paradigms are conceptualized in the social sciences, describes a particular version of paradigms that focuses on philosophical propositions, presents multiple paradigms within that typology, discusses whether or not different paradigms are compatible and argues for sustained communication between communities using paradigmatic concepts as a platform. In light of what has been discussed, coupled with recent efforts to legislatively define scientific research (ESRA, 2002; NCLB, 2002) and the various degrees to which different communities' belief systems fit those molds, paradigms appear to have profound implications for inquirers. Although many scholars claim that the paradigm wars of the past are experiencing a cease-fire, the abundance of literature arguing different points of view suggests something to the contrary, albeit in a less strident format. From a philosophical standpoint, questions persist about whether alternative genres to inquiry have indeed reached a "paradigm" status, or if they remain "pre-paradigms" seeking hegemony and/or even acceptance in communities already dominated by an existing paradigm. Because it is difficult to argue against the verity that certain researchers subscribe to views and practices that fail to correspond with conventionalist typologies, these questions may not be as important as questions about whether various communities with divergent perspectives are

willing to enter into a dialogue with one another and/or how that dialogue might help to enhance the efficacy of educational and/or social science research in general.

In this article, I set out to provide a brief overview of paradigmatic concepts so that researchers can begin to recognize their own positions, as well as the philosophical forces guiding the work of others. The presentation of these concepts was not intended to suggest that any given paradigm was incontestably appropriate for all research questions demanding our attention. However, given what was explored, the following are loosely proposed:

- A positivist research paradigm is no longer relevant in contemporary social science inquiry. It has instead been replaced by post-positivism (Phillips & Burbles, 2000 as cited by Johnson & Onwuegbuzie, 2004).
- The paradigmatic structures described by Guba and Lincoln (1994; 2005) are compelling at the comprehensive philosophical levels (i.e. ontology, epistemology and axiology), but perhaps unreasonable in the application of methods where pragmatism may rule in accordance with the research question being asked.
- Although most inquirers agree that research questions should drive the selection and deployment of methods, ontological, epistemological and axiological assumptions may still be important components for determining which research questions are important and how those questions are structured.
- Given the increasing interdisciplinary nature of research, a continued focus on paradigms might help to create an understanding across different research communities and provide a conceptual framework that fosters communication and collaboration between researchers with similar and/or diverse points of view.

As has become commonplace, propositions like these tend to lead to other research questions, such as *What are the philosophical views of researchers?; To what degree are researchers familiar with their own philosophical orientations?;* and, *To what extent do researchers' belief systems and behaviors adhere to hypothesized paradigmatic structures?* As a necessary first step to answering questions such as these, it is suggested that there is a need for the creation of instrumentation to measure the dimensions that comprise research paradigms in the appropriate context. Beyond confirming hypothesized paradigmatic structures, such instrumentation could serve a more pragmatic role in the edification of graduate students who will serve as future researchers in the social sciences. For example, an instrument that delivered

valid and reliable measures of the dimensions that comprise research paradigms could be used to identify philosophical predilections of graduate students and help guide them toward coursework and/or research projects in which they might flourish. Such an instrument could also be used as a research tool for examining the views of current researchers and graduate students, possibly by measuring change(s) across the life of an educational program or career, and/or in the exploration of variables that contribute to the development of one's research approach altogether.

Like the supremacy of any given paradigm, the decision to focus on a philosophical account of paradigms in this article does not imply that its content unerringly mirrors the beliefs, values and activities of every researcher. The adaptation presented just happens to be the most popular conceptualization of paradigms at this point in time. Further, because "it is unlikely that a practitioner of any paradigm would agree that [Guba & Lincoln's] summaries closely describe what he or she does" (1994, p. 117), there is no expectation that the beliefs and practices of all researchers will fit neatly into discrete categories. However, if the dimensions contained within the referent framework are considered as broad strokes, it is strongly suggested that inquirers will likely demonstrate proclivity toward one ideology or another. Inasmuch as we are able to reflect on the referent framework, or a facsimile thereof, paradigmatic concepts may in time prove to be powerful tools for interpreting language and other concepts from different research genres. Further, by garnering an understanding of the various paradigms that such structures are likely to encompass, it is hoped that inquirers may eventually be judged solely on the relative merits of their research, the quality and dependability of their research assertions, and not necessarily by their comportment or the particular paradigmatic genre in which they practice.

CHAPTER 3
INSTRUMENT DEVELOPMENT PROCESS

Chapter Three of this document provides a detailed overview of the process associated with the development, review, trial and pre-testing of items that appear on the *Research Paradigm Inventory* (RPI). The amount of detail necessary to adequately explain the process for creating, refining, pilot testing and field testing items for the instrument would exceed the limited amount of space available in a journal article. However, because it is important to document all relevant activities through which an instrument and its items are created, Chapter Three is presented for that purpose. Much of what is presented here will be replicated in an article that describes this study (see Chapter Four), albeit in a more condensed format. In cases where the amount of information required to explain specific concepts exceeds the scope of such an article, readers will be referred to Chapter Three of this document for a more detailed description.

Framework for Instrument Development

In order to build a strong validity argument for an appropriate interpretation and use of instrument measures, aspects of the development process for the RPI were examined in accordance with the validity framework outlined by Messick (1995). Messick (1995) defined validity as, "an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment", wherein the validity of measure interpretation and resulting action can be assessed in terms of *content, substantive, structural, generalizability, external* and *consequential* forms of evidence (p. 741). Taken collectively, these six aspects serve as "general validity criteria or standards for all educational and psychological measurement" (Messick, 1995, p. 747). During this study, several of the aspects contained in Messick's framework were addressed through a series of activities designed for instrument development and measure validation (Wolfe & Smith, 2007a, 2007b). Table 3.1 provides an introduction to the activities that were performed in this regard, as well as a conceptual link to the specific aspect of validity each activity intended to address. Each step taken to provide evidence of validity for interpreting RPI scores is documented following the table.

Table 3.1
Instrument Development and Validation Activities Linked to Validity Aspects

	Validity Aspect			
	Content	**Structural**	**Substantive**	**Generalizability**
Validity Evidence	Instrument Purpose Instrument Specifications Item Development Expert Reviews Item Technical Quality	Instrument Dimensionality	Rating Scale Function Person Fit Item Difficulty Hierarchy	Item Calibration Invariance Reliability Precision

* Adapted from Wolfe & Smith, 2007b, p. 244

Instrument Purpose

As noted in Chapter One of this document, the purpose of this study was to develop an instrument to measure the views of educational researchers across the dimensions that comprise research paradigms. The development of the instrument was conducted in the hope of providing a mechanism for future research that will enable the examination of prospective links between the philosophical and methodological orientations of current researchers and graduate students in the field of education.

Instrument Specifications

Content specifications for the instrument were conceptualized following a thorough review of literature and then encapsulated within an instrument blueprint (see Appendix B). The blueprint for the instrument originally specified eight content domains, which were comprised of four inquiry dimensions anchored by extreme views within those respective dimensions (i.e., a 2x4 table consisting of):

- *Ontology* anchored by (1) *Absolute Realism* and (2) *Absolute Relativism*;
- *Epistemology* anchored by (3) *Objectivism* and (4) *Subjectivism/Interpretivism*;
- *Axiology* anchored by (5) *Values Excluded* and (6) *Values Included*; and,
- *Methodology* anchored by (7*) Hypothesis testing/Chiefly Quantitative Methods* and 8) *Emerging data/Chiefly Qualitative Methods*.

Once constructed, a small group of individuals who were familiar with concepts related to paradigms was consulted about the instrument blueprint and asked to provide input regarding its accuracy and relevancy to the construct domains. These individuals confirmed the viability of the instrument blueprint for the intended purpose and judged its specifications to "make sense" and be "uncontroversial" in terms of content and structure, although one individual strongly encouraged the inclusion of a "middle ground" for each of the philosophical dimensions. Because of this individual's expertise in this area, this suggestion was seriously considered. However, after much deliberation, it was decided that creating an additional set of cells within the blueprint to represent a "middle ground" would be problematic, especially as it related to drafting and interpreting content for such items. Ultimately, it was reasoned that intermediate views would essentially be represented by items reflecting less extreme views within the content domains that had already been conceptualized.

Item Development

Each of the eight content domains specified by the instrument blueprint was operationalized by a series of items to optimally represent each respective inquiry dimension. The item response format was constructed using a subject-centered Likert scale consisting of four ordinal categories, which were coded and labeled as: (0) *Strongly Disagree*, (1) *Disagree*, (2) *Agree*, and (3) *Strongly Agree*. Stems for newly crafted items were generated in accordance with content models prevalent in the literature using guidelines for writing rating scale items (Wolfe & Smith, 2007a, p. 230). In addition, a select number of item stems that had appeared on previously developed instruments were also included, if those item stems exhibited content relevant to the instrument blueprint. In total, 160 items were created and/or adapted. However, because some items illustrated considerable replicability in terms of content, the item pool was trimmed to 106 before further reviews on them were conducted.

Expert Reviews

Content Matching. A group of experts (n = 8) was invited to review the draft set of 106 items and provide links to the instrument blueprint. Participants for this exercise were recruited from a pool of doctoral students who had participated in coursework that included discussions related to research paradigms, instructors of research methods and individuals who were

conducting research in various fields of education. To facilitate the linking of items to the content domains of the instrument blueprint, an item-content linking instrument was created (see Appendix C). The item-content linking instrument included a table, which outlined and defined the eight domains of the instrument blueprint, as well as the draft set of 106 items, which had been randomized using a table of random numbers. Directions on the instrument asked respondents to review each of the 106 items carefully and then identify the content domain from the table they believed an item represented.

Two criteria were adopted to support the selection of items that would be retained for pilot testing, these being: (1) at least 6 of the 8 individuals who participated provided a response for an item, and (2) there was at least 70% agreement among those responding about an item's content. A total of 54 items met these criteria, with 25 items receiving 100% participant agreement, 8 items receiving 88% agreement, 7 items receiving 86% agreement, 11 items receiving 75% agreement and 3 items receiving 71% agreement. An additional 6 items were also selected for pilot testing because there was a need to ensure adequate representation of each content domain, there was agreement among a majority of respondents around item content and/or an item's associated concepts were clearly supported in the literature.

In total, 60 items were selected and included into a truncated version of the item-content linking instrument, which was subsequently administered to an additional three content experts. Based on their responses, 46 items (76.7%) received unanimous agreement. For those items in which there was not unanimity, at least 2 of 3 respondents were in agreement.

Item Edits. A group of individuals (n = 4) was asked to review the draft sets of 60 items and provide comments regarding their clarity, succinctness and potential for bias. Participants for this exercise were recruited from a pool of experts who had experience with item writing and editing such as instructors of research methods, measurement professionals and graduate students studying measurement. Following a review of their comments, 15 items received minor revisions, which did not impact item content, and one item was eliminated, yielding a final draft set of 59 items designated for pilot testing.

Item Pilot Tests

The draft 59-item set was included in an online survey created with *Survey Monkey*©, wherein potential respondents (educational research faculty and graduate students at research

institute located in Southwest Virginia) were sent an invitation to participate via e-mail, which contained a URL link to the survey. The online instrument was comprised of the draft 59-item set, which had been randomized using a table of random numbers; demographic variables; and, an open-ended question with which participants could provide additional comments regarding their impressions of the items, the ease with which they were able to respond, the efficacy of the response scale and so forth. In total, 48 individuals responded to the pilot test survey.

A preliminary scan of response category frequencies showed that, although the response scale appeared to be functioning with moderate success for most items, six items failed to receive a single response in at least one endpoint category. Therefore, for the purpose of assessing item quality and dimensionality, data were scaled to the Rasch Partial Credit Model (Rasch, 1960; Wright & Masters, 1982) given by:

$$\ln(P_{nix} / P_{ni\,x-1}) = B_n - D_i + F_{ix}$$

where:

P_{nix} = the probability that participant n will respond in category x on item i
$P_{ni\,x-1}$ = the probability that participant n will respond in category $x-1$ on item i
B_n = the ability for participant n
D_i = the difficulty for item i
F_{ix} = the step that corresponds to an equal probability of selecting category x and $x-1$

By allowing category thresholds to vary across individual items, the Partial Credit Model allows each item to employ a unique rating scale structure. Considering what had been observed during the initial review of response category frequencies, it was reasonable to assume this strategy would be appropriate for assessing dimensionality and item fit at this early stage of instrument development.

Parameters for both persons and items were estimated using the joint maximum likelihood estimation (JMLE) procedures contained in *Winsteps 3.62* (Linacre, 2006), wherein estimates for both item difficulty and person ability were calibrated along a common logit metric. Due to the relatively small sample size (n = 48 persons), item fit and dimensionality were evaluated in general terms, as opposed to strictly adhering to the criteria specified for the measures validation phase of this study. Criteria for the initial review of items included using a point-measure correlation criterion ($r_{pm} \geq .50$) for identifying items that were substantively contributing to measures, as well as the item unweighted mean square fit statistic (MS_{outfit} values

≥ 1.5), then the item weighted mean square fit statistic (MS_{infit} values ≥ 1.5) and finally the standardized version of each ($Z_{unweighted}$ and $Z_{weighted}$ values ≥ 2.0) for identifying potentially misfitting items. At this early stage of development, greater import was given to $Z_{unweighted}$ and $Z_{weighted}$ because of their greater insensitivity to false identifications of misfit when using smaller sample sizes. The dimensionality of item sub-sets was evaluated using a principal components analysis (PCA) of standardized residuals after scaling the data to the Partial Credit Model.

A series of eight initial parameter estimations were calibrated for items within each of the eight content domains specified by the instrument blueprint to ensure items were working cooperatively to represent the content domains respectively. After evaluating each item according to the aforementioned criteria, four items were deleted from the draft item pool, yielding an updated set of 55 items. Table 3.2 identifies the items that failed to meet the selection criteria, the items that were deleted and the content domain to which flagged items were linked.

Prior to deciding whether to exclude items that appeared as misfitting from further analyses, the standardized residual ($Z_{residual}$) for each item/person combination was inspected to determine if a single aberrant response might be contributing to the observed misfit. During this process, items that contained one or two large standardized residual values (i.e., $Z_{residual} > 3.0$), but not more than three values greater than 2.0 were considered. If it appeared as if a single response for a person might be contributing to the observed misfit, and it was felt that the item would potentially contribute substantive information about important instrument content, then the response associated with the large residual was eliminated prior to re-calibrating item parameter estimates for a content domain. For instance, after removing a single response for person 24 ($Z_{residual} = 3.973$), Item 53 demonstrated adequate fit to the model ($Z_{unweighted} = 1.4$, $Z_{weighted} = 1.0$, $r_{pm} = .61$). Item 29 likewise demonstrated adequate fit to the model ($Z_{unweighted} = -.6$, $Z_{weighted} = -.3$, $r_{pm} = .75$) following the removal of a single response for person 37 ($Z_{residual} = -5.56$). As such, Items 53 and 29 were retained for additional analyses.

Table 3.2
Item Misfit from Pilot Test Content Domain Calibrations

Item Text	Content Domain	MS_{infit}	$Z_{weighted}$	MS_{outfit}	$Z_{unweighted}$	r_{pm}	*Decision*
3. Subjective human thinking inherently limits the possibility for objectivity during research.	2. Absolute Relativism	1.74	3.1	1.92	3.7	.43	Delete
49. A researcher should limit his or her contact with participants to prevent biased conclusions.	3. Objectivism	1.32	1.4	1.26	1.0	.48	Delete
51. The quality of a research study is determined primarily by the action it prompts.	6. Values Included	1.87	2.9	1.82	2.7	-.10	Delete
57. Researchers should have a strong background in quantitative research methods.	7. Hypothesis Test	1.39	1.7	1.54	2.1	.43	Delete
2. Good research proceeds by drawing generalizable conclusions from empirical data.	7. Hypotheses Test	1.00	.1	1.00	.1	.45	Revise
1. A good researcher should have a strong background in qualitative research methods.	8. Emergent Data	1.27	1.3	1.31	1.4	.49	Revise
53. Reality is a stable entity – it is only people's perception of reality that varies.	1. Absolute Realism	1.40	1.7	1.66	2.77	.55	Retain
29. Reality is formed through individual cognition.	2. Absolute Relativism	1.17	.8	1.35	1.4	.54	Retain
28. Researchers are responsible for making sure participants benefit from their research.*	6. Values Included	.55	-2.3	.55	-2.2	.84	Retain
30. Larger, randomly selected samples provide the most useful data for research.*	7. Hypotheses Test	.60	-2.1	.60	-2.1	.80	Retain

* Items 28 and 30 were retained for future analyses because they appeared to be exhibiting overfit (i.e., potential redundancy with other items), rather than inconsistency with their respective modeled dimensions. Closer attention was given to potentially over-fitting items during the measures validation phase of this study.

Although Items 1 and 2 failed to meet the r_{pm} criterion, each exhibited reasonable fit to their modeled dimension in terms of their standardized fit. Because of their placement on the pilot test survey and the level of difficulty estimated for each, it was decided that these items should not be abandoned entirely. Further, given the degree to which their concepts were deemed important for constructing measures associated with instrument content, these items were designated for revision, followed by additional analyses.

In total, 55 items were retained following preliminary analyses of item fit and contribution. Although most of the retained items exhibited adequate fit and appeared to be contributing to the construction of measures, additional schema for item reduction were employed in an effort to maximize the response rate for item field tests, as well as the eventual usability of an operational form of the instrument. As an exploratory step to determining what those strategies might entail, the correlations for each pair of person measures from the content domains were estimated and reviewed (see Table 3.3).

Table 3.3
Content Domain Person Measure Correlations

	Traditionalist Content Domains				Interpretivist Content Domains			
	Domain 1. Absolute Realist	Domain 3. Objectivist	Domain 5. Values Excluded	Domain 7. Hypothesis Testing	Domain 2. Absolute Relativist	Domain 4. Subjectivist	Domain 6. Values Included	Domain 8. Emerging Data
Domain 1.	1.00							
Domain 3.	.513**	1.00						
Domain 5.	.423**	.496**	1.00					
Domain 7.	.535**	.709**	.363*	1.00				
Domain 2.	-.136	-.108	-.023	-.316*	1.00			
Domain 4.	-.100	-.463**	-.359*	-.559**	.439**	1.00		
Domain 6.	-.134	-.339*	-.264	-.498**	.307*	.501**	1.00	
Domain 8.	-.250	-.468**	-.304*	-.742**	.365*	.588**	.505**	1.00

** Correlation was statistically significant with $p < .01$ (2-tailed).
* Correlation was statistically significant with $p < .05$ (2-tailed).

After reviewing the patterns of correlation between pairs of measures from the content domains, it was believed that the truncated 55-item set might be comprised of at least a two-factor structure, which corresponded, in part, to the theoretical propositions that guided the creation of the instrument blueprint (i.e., paradigms could be modeled according to relationships between dimensions). Based on the associations that appeared to exist between certain content domains (i.e., domains 1, 3, 5, and 7, and domains 2, 4, 6, and 8), these potential factors were interpreted to represent a set of *Traditionalist* (e.g., positivism/post-positivism) and *Interpretivist* (e.g., constructivism) beliefs respectively. Therefore, subsequent parameter estimations were calibrated for several other item combinations, which included the truncated 55-item set; the same 55-item set after reverse-coding items from domains 2, 4, 6 and 8; and, additional sub-sets

of items using a Traditionalist/Interpretivist distinction, as well as a philosophy/methods distinction within both *Traditionalism* and *Interpretivism* respectively.

First Item Parameter Calibration (55-Item Set). An initial calibration of item parameters for the truncated 55-item set yielded unsatisfactory results, with the modeled dimension able to account for only 24.3% (eigenvalue = 17.6 / 72.6 total) of the variability in person measures. In total, just seven items exhibited adequate fit statistics and $r_{pm} \geq .50$. After the initial evaluation of item fit/contribution was conducted, the dimensionality of truncated 55-item set was explored. In the present study, if data were unidimensional, which is a primary specification for Rasch models, then the Rasch dimension should account for the predominant amount of systematic variation in responses, leaving only randomly distributed residuals (Linacre, 2002). Because absolute unidimensionality is implausible, the unidimensional specification was considered to have been met to a satisfactory extent by the instrument response data if a dominant component was seen to underlie performance on the draft set of items (Hambleton, 1993).

Winsteps 3.62 contains a feature in which a principal components analysis (PCA) of standardized residuals is performed after extracting the Rasch dimension. *Winsteps 3.62* constructs a series of "contrasts", where each contrast reflects opposing response patterns across items by persons. An analysis of the strength of those contrast components, expressed in eigenvalues units, is often useful in determining the degree to which opposing factors are providing structure in standardized residuals after data have been scaled to a model. As regards the interpretation of the magnitude of a contrast, an eigenvalue of 3, for example, may be interpreted as holding as much strength as 3 items. In terms of the relative importance of contrasts versus the modeled dimension, Linacre (2006) suggested using the following "Rules of Thumb": (a) percentage of variance explained by measures > 60% is good, (b) unexplained variance explained by the 1st contrast – eigenvalue < 3.0 is good, and (c) unexplained variance explained by 1st contrast < 5% is good.

The PCA of item standardized residuals for the truncated 55-item set revealed substantial systematicity beyond the modeled dimension. Using an eigenvalues criterion > 1.4 for determining importance (Smith & Miao, 1994), several important contrasts were identified, the first five of which demonstrated eigenvalues equal to 14.0, 4.1, 3.8, 3.5 and 3.4 respectively. To support the claim that residual components were reliably measured, only those items with absolute loadings ≥ .40 were used for interpretation (see Stevens, 2002, p. 394). On deciding

which contrasts to interpret, Stevens' (2002) guidelines again were employed, which recommend interpreting components with: (a) at least three absolute loadings above .80; (b) four or more absolute loadings above .60; or, (c) ten or more absolute loadings above .40 when the sample is greater than about 150 (p. 395).

According to Stevens' (2002) guidelines, only the first residual component was reliably represented and warranting interpretation, given the sample size available ($n = 48 < 150$). Examination of the first contrast revealed 20 meaningful (i.e., loadings $\geq |.40|$) negatively loading items and 22 meaningful positively loading items. All 20 of the meaningful negatively loading items were originally crafted and sustained by content reviewers to reflect *Traditionalist* beliefs from domains 1, 3, 5 and 7 of the instrument blueprint, whereas all 22 of the meaningful positively loading items exhibited content that was created to reflect *Interpretivist* beliefs from domains 2, 4, 6, and 8. This result was believed to be consistent with what had been previously observed, wherein it appeared as if the 55- item set might be representative of a two factor structure.

Second Calibration (Interpretivist Items Recoded). Because Rasch models require all items to be 'oriented' in the same direction, items that were created to represent *Interpretivist* views were reverse-coded (i.e. 0123 → 3210) prior to conducting a second calibration of item parameter estimates. It was predicted that reverse coding these items would orient all items toward a *Traditionalist* viewpoint. Therefore, and if the theoretical structure of the instrument specifications held, the reverse coding of *Interpretivist* items from domains 2, 4, 6 and 8 would account for a larger amount of systematic variation in person measures and a stronger fit of items to the modeled dimension.

The second calibration of item parameters for the 55-item set was again unacceptable, yet showed noticeable improvement, accounting for 46.2% (eigenvalue = 47.3 / 102.3 total) of the variability in person measures and 26 items with $r_{pm} \geq .50$. A PCA of item standardized residuals again revealed multiple contrasts with eigenvalues greater than 1.4, the first five of which demonstrated eigenvalues equal to 7.2, 4.7, 4.4, 4.1 and 3.6 respectively. However, considering the sample size, only the first contrast component, which was comprised of 9 meaningful (i.e., loadings $\geq |.40|$) positively loading items and 7 meaningful negatively loadings, met Stevens' (2002) criteria for interpretation. A review of the content for each of these items again illustrated a distinct contrast between *Traditionalist* and *Interpretivist* views.

Third Calibration (Traditionalism). Because there appeared to be a clear distinction between *Traditionalist* and *Interpretivist* views, as well as consistencies across certain philosophical dimensions within those categorizations, a third calibration of item parameters was conducted, which featured only those items that had been originally created to reflect *Traditionalist* views. The third calibration of item parameters was able to account for an acceptable 62.3% (eigenvalue = 47.8 / 76.8 total) of the variability in person measures. A PCA of item standardized residuals yielded 5 contrast components with eigenvalues greater than 1.4 (4.2, 3.3, 2.6, 2.6 and 2.3 respectively), one of which warranted interpretation according to the Stevens (2002) recommendations. The first residual contrast contained 7 meaningful (i.e., loadings $\geq |.40|$) positively loading items that, in general, reflected methodological views (i.e., items from domain 7) and 5 meaningful negatively loading items that reflected philosophical views (i.e., items from domains 1, 3 and 5). This was believed to represent a group of respondents that subscribe to realist/objectivist views, but do not necessarily endorse the use of hypothesis testing and/or quantitatively methods exclusively.

Sub-Set Calibrations. Because of the potential for delineations beyond a *Traditionalist* and *Interpretivist* distinction (e.g., a philosophy/methods distinction within *Traditionalism* and/or *Interpretivism*), each of the 55 items that were retained following the initial phase of item analyses was placed into one of four categories:

1. Traditionalist Philosophy – items from domains 1, 3 and 5;
2. Quantitative Methods – items from domain 7;
3. Interpretivist Philosophy – items from domains 2, 4 and 6; and,
4. Qualitative Methods – items from domain 8.

Separate parameter estimations were then calibrated for the items within each of the four categories, wherein items were deleted using the fit/contribution criteria employed during the first phase of item analysis (i.e., the initial 8 content domain item parameter calibrations). Table 3.4 highlights the performance of each sub-set of items in its final form, as well as the performance of the 55-item set and the 55-item set after reverse-coding *Interpretivist* items. Table 3.5 provides similar indices for the four sub-sets of items after data were scaled to the Rasch Rating Scale Model (Andrich, 1978).

Table 3.4
Performance Indices for Sub-Sets of Items (Partial Credit Model)

	All Items	All Items (Reverse Coded)	Sub-Set 1: Traditionalist Philosophy	Sub-Set 2: Quantitative Methods	Sub-Set 3: Interpretivist Philosophy	Sub-Set 4: Qualitative Methods	Traditionalist All Items[a]	Interpretivist All Items[b]
Number of Items	55	55	15	9	15	8	24	23
Rel_θ	.65	.85	.82	.77	.83	.64	.89	.85
Cronbach's α	.72	.94	.87	.85	.89	.81	.91	.91
$r_{pm} \geq .50$	7 items	26 items	14 items	8 items	15 items	6 items	22 items	19 items
$r_{pm} \geq .40$	12 items	42 items	15 items	9 items	15 items	7 items	24 items	23 items
$MS_{infit} \geq 1.5$	0 items	0 items	0 items	0 items	0 items	0 items	0 items	1 items
$Z_{unweighted} \geq 2.0$	0 items	0 items	0 items	0 items	0 items	0 items	1 item	0 items
Total Var_θ explained by Rasch Dimension	24.3%	43.8%	71.7%	53.6%	82.4%	50.7%	77.3%	84.7%
Eigenvalue for 1st Contrast	14.0	7.2	2.7	2.5	2.8	2.0	3.7	3.2

a. *Traditionalist* sub-set represents a combination of the *Traditionalist Philosophy* and *Quantitative Methods* sub-sets.
b. *Interpretivist* sub-set represents a combination of the *Interpretivist Philosophy* and *Qualitative Methods* sub-sets.

Table 3.5
Performance Indices for Sub-Sets of Items (Rating Scale Model)

	Sub-Set 1: Traditionalist Philosophy	Sub-Set 2: Quantitative Methods	Sub-Set 3: Interpretivist Philosophy	Sub-Set 4: Qualitative Methods	Traditionalist All Items	Interpretivist All Items
Number of Items	15	9	15	8	24	23
Rel_θ	.82	.77	.78	.73	.89	.82
Cronbach's α	.87	.85	.89	.81	.91	.91
$r_{pm} \geq .50$	14 items	9 items	15 items	7 items	22 items	18 items
$r_{pm} \geq .40$	15 items	9 items	15 items	8 items	24 items	23 items
$MS_{infit} \geq 1.5$	0 items	0 items	0 items	0 items	0 items	1 items
$Z_{unweighted} \geq 2.0$	1 items	0 items	0 items	0 items	1 item	0 items
Total Var_θ explained by Rasch Dimension	67.5%	53.7%	69.6%	71.3%	74.1%	65.8%
Eigenvalue for 1st Contrast	2.7	2.5	2.9	2.0	3.7	3.2

Following the estimation of parameters for items within each of the respective 4 sub-sets, a total of 47 items were selected for the field test/measure validation phase of this study. In order to enhance the reliability of potential sub-scales, 3 additional items were drafted, which yielded a revised set of 50 items. Prior to conducting field tests of items, the correlations between all pairs

of sub-set measures were estimated to examine relationships that might be used for future exploration, specification and confirmation of instrument dimensionality (see Table 3.6).

Table 3.6
Correlations between Person Measures from Item Subsets

	1. Traditionalist Philosophy	2. Quantitative Methods	3. Interpretivist Philosophy	4. Qualitative Methods
1. Traditionalist Philosophy	1.00			
2. Quantitative Methods	.741**	1.00		
3. Interpretivist Philosophy	-.382*	-.614**	1.00	
4. Qualitative Methods	-.522**	-.742**	.670**	1.00

** Correlation was statistically significant with p < .01 (2-tailed).
* Correlation was statistically significant with p < .05 (2-tailed).

Because of the correlations between measures from the sub-sets of items and the multidimensionality that appeared to exist within the full 55-item set, it was determined that both a total score for *Traditionalism*, a total score for *Interpretivism*, as well as sub-scale scores within each of those would be explored during future iterations of instrument development. Further, it was determined that a multidimensional Rasch model and/or confirmatory factor analytic strategies would be considered during the measure validation phase of this study to confirm a hypothesized structure of the instrument before a final version was reported.

Item Tryouts (Think Aloud Interviews)

In order to assess the degree to which future respondents might process items that appear on the RPI, as well as identify potential sources of construct-irrelevant variability in response data sets, four individuals were administered the draft form of the RPI in an interview format. Three of these individuals were doctoral students in adult learning, and the fourth was an individual with a doctorate in education and other advanced degrees in philosophy. Through the course of this process, participants were presented with the draft form of the RPI, asked to 'think aloud' as they proceeded through the instrument's items and verbally provide context and rationale for the answers they gave. By and large, participants appeared to process the items as intended, where most of the comments given related to disagreements with item statements, as opposed to issues with the structure of an item and/or the terminology used. Following a review

of participant comments, none of the items that were included on the draft form were deleted, and only a few received minor updates in order to provide clarity in terms of item meaning.

Item Field Tests

A draft of the RPI was formatted, included in a survey created with *Survey Monkey*© and made available for response in order to field test items and document the psychometric properties of the instrument. The survey was comprised of the revised 50-item set, demographic variables and open-ended questions with which participants could provide additional comments regarding their impressions of items and the survey in whole. Directions were provided within each section of the survey that informed participants about the means by which they were to provide responses, and submit those responses by a specified date.

Participants for the field test and measures validation phases of this study were petitioned from various online list serves, which included the American Educational Research Association's (AERA) Division D: Measurement and Research Methodology, the list serve of the Graduate Student Forum of the AERA and the Qualitative Research for the Human Sciences List (QUALRS-L). Potential participants were sent an invitation via e-mail that contained a secure, encrypted URL to which participants could link and respond anonymously. In total, 210 individuals began the survey, with 189 submitting responses that were considered adequate for data analyses (i.e., provided a response to at least 50% of the scale items).

Preliminary analyses into the dimensionality of the pilot test items suggested the full item set was tapping at least four dimensions. A replication of those analyses using field test data suggested the RPI was more than likely tapping six dimensions, where a further delineation within *Traditionalist Philosophy* and *Interpretivist Philosophy* could be made using an ontological and epistemological/axiological distinction. This observation was particularly noteworthy, as it seemed to directly correspond with a popular depiction of paradigms that has appeared in the literature, at least in terms of identifying the important elements that serve to structure such (see Table 8.1 in Guba & Lincoln, 2005, p. 193).

Because the goal for creating the RPI was to measure views within the dimensions that comprise research paradigms, as opposed to identifying an individual's belief system for the purpose assigning a label to them, the decision was made to model various item groupings as

individual sub-scales, wherein the instrument's dimensionality in its entirety would be evaluated through confirmatory analyses. As such, six potential scales were identified, these being:

1) *Realism in Research*;

2) *Research Objectivism*, which included 'Values Excluded' items;

3) *Quantitative Methodology*;

4) *Relativism in Research*;

5) *Research Interpretivism*, which included 'Values Included' items; and,

6) *Qualitative Methodology*.

Even through the revised categorizations were not markedly different than what had been identified following item pilot tests (i.e., four categories), aside from separating ontological concepts from epistemology/axiology, the re-apportioning of content did require a re-assessment of item fit. Therefore, items were re-scaled and their fit re-evaluated according to the revised six scale depiction detailed above. Table 3.7 identifies those items that were subsequently deleted and their associated indices of fit and contribution.

Table 3.7
Item Misfit from Field Test Scale Calibrations

Item Text	Scale	MS_{infit}	$Z_{weighted}$	MS_{outfit}	$Z_{unweighted}$	r_{pm}	*Decision*
13. Knowledge is morally neutral until it is applied.	Research Objectivism	1.55	4.5	1.71	4.5	.53	Delete
31. A researcher must set aside his or her personal beliefs in order to reach accurate conclusion.	Research Objectivism	.66	-3.6	.66	-3.6	.81	Delete
45. The research topics that I study are independent of me.	Research Objectivism	1.08	.7	1.18	1.4	.59	Delete
46. Researchers who study the same topic and use the same tools should arrive at similar conclusions.	Quantitative Methodology	1.16	1.5	1.17	1.5	.56	Delete
48. Every step of a research design should be structured before the process of data collection begins.	Quantitative Methodology	1.31	2.8	1.37	3.0	.58	Delete
12. Researchers are responsible for making sure participants benefit from the research he or she conducts.	Research Interpretivism	1.4	3.5	1.4	3.6	.34	Delete
20. Researchers are responsible for how their conclusions are used.	Research Interpretivism	1.64	5.2	1.67	5.3	.34	Delete
24. Knowledge gained through research is based on subjective choices made by a researcher.	Research Interpretivism	.98	-.1	.98	-.1	.69	Delete
40. One of my goals for research is to establish equality where none existed before.	Research Interpretivism	1.73	5.9	1.73	5.9	.29	Delete
27. It is unreasonable to generalize conclusions beyond the immediate setting of a study.	Qualitative Methodology	1.35	3.1	1.34	3.0	.32	Delete
43. The hypothesis that I formulate emerge following a research study.	Qualitative Methodology	1.37	3.3	1.4	3.5	.15	Delete

a. Item 31 was deleted because it appeared to exhibit overfit and its difficulty estimate was closely matched to another item that was retained.
b. Item 46 was deleted after a re-evaluation of item content determined it was probably more representative of objectivist views.
c. Item 45 was deleted after several respondents indicated its content was confusing to them.
d. Item 24 was deleted because it closely matched Item 23 in terms of content and difficultly estimate, which was interpreted to represent an idiosyncratic response pattern.
e. Note: Item 1 was deleted, as it failed to load onto any of the sub-scales, presumably due to a lack of variability in responses. A review of response category frequencies demonstrated 94% of respondents either Agreed or Strongly Agreed with this item's content.

After confirming the contribution and fit of items to the six potential sub-scales, a principal components analysis (PCA) of standardized residuals was performed to evaluate the dimensionality of each. Table 3.8 summarizes the results of those analyses by providing information about the amount of variance explained by the modeled Rasch dimension, the amount of variance explained by the first standardized residual component, the number of residual components with loadings greater than .60 in absolute value, and whether residual components met an abbreviated version of Steven's (2002) criteria for interpretation.

Table 3.8
Principal Components Analysis of Item Standardized Residuals

Sub-Scale	# of Items	Component	Eigenvalue	Total % of Var_θ Explained	Number of $\lambda > \|.60\|$	Met criteria* for interpretation?
1. Realism in Research	6	Rasch	12.3	67.3%	--	☒
		1st residual	2.1	11.4%	3	
2. Research Objectivism	6	Rasch	10.7	61.8%	--	☒
		1st residual	1.7	11.0%	2	
3. Quantitative Methodology	7	Rasch	17.8	71.8%	--	☒
		1st residual	1.4	5.8%	2	
4. Relativism in Research	6	Rasch	6.0	50.2%	--	☒
		1st residual	1.8	14.5%	2	
5. Research Interpretivism	6	Rasch	6.8	53.0%	--	☒
		1st residual	1.5	11.7%	2	
6. Qualitative Methodology	7	Rasch	11.1	61.3%	--	☒
		1st residual	1.6	8.8%	2	

* Because each sub-scale contained less than 10 items, interpretation criteria were: (a) at least three items have loadings greater than |.80|; or, (b) at least four items have loadings greater than |.60| (See Stevens, 2002). Components that failed to meet the criteria for interpretation are identified by ☒.

Principal components analyses of standardized residuals yielded zero components that met the abbreviated criteria for reliable interpretation. Considering the ratio of modeled variance to variance explained by the residual components, wherein it appeared as if a dominant component was underlying each set of scale responses, it was concluded that each scale gravitated toward unidimensionality. However, prior to drawing any conclusions regarding the overall dimensionality of the RPI, confirmatory exercises were conducted, the results of which are detailed in Chapter Four of this document.

In total, 38 items were selected to represent the six scales included on the updated version of the RPI. The psychometric properties of the RPI were then explored through a series of analyses designed to document the dimensional structure of the RPI, the efficacy of the item response scale, the reliability of the six scales and the invariance of measures obtained from the instrument. Each of these analyses are reported and discussed in Chapter Four of this document.

CHAPTER 4

MANSCRIPT 2

Development and Validation of an Instrument to Measure Educational Research Paradigms

Abstract

The purpose of this study was to develop an instrument, entitled the *Research Paradigm Inventory* (RPI), to measure views that comprise research paradigms in education. In accordance with recommendations for instrument development and measure validation using Rasch Models, a series of activities was performed to address various aspects of the validity framework outlined by Messick (1989) during the course of this project. This article describes these activities and documents content, substantive, structural and generalizability forms of validity evidence in support of an appropriate interpretation and use of the instrument's measures.

Keywords: Educational Research, Paradigms, Rasch Measurement

Theoretical Background

In recent years, a great deal of attention has been paid to how scientific research in education is defined and conducted (ESRA, 2002; NCLB, 2002), as well as to what constitutes quality in that regard (Hostetler, 2005; Lather & Moss, 2005). In the literature, quality often refers to an inquirer's ability to adhere to specific research traditions, but then also to opinions about what constitutes credible evidence for inquiry-based assertions and the philosophical foundation undergirding various research approaches. The attention given to issues about quality has helped catalyze the emergence of alternative perspectives and exploratory archetypes for examining topics in education, where certain research communities have shifted from a "traditionalist" position (e.g., positivism, post-positivism), sometimes referred to as the "received view" (Guba & Lincoln, 1994), to a more milieu-dependent positioning of "interpretivism" (Howe, 1998; Nisbet, 2005; Rabinow & Sullivan, 1979; Tsai & Liu, 2005). Investigators with diverse philosophical orientations often view the purpose of their research differently, incorporate the use of different methods and apply an assortment of standards to evaluate the quality of a research study.

Consequential to Thomas Kuhn's (1962/1996) popularization of the term paradigm, a number of scholars have offered up the concept of paradigmatic distinctions between research communities as a possible source for this diversity (Creswell, 1998; Guba & Lincoln, 1994; Schwandt, 1989). Discussions about research paradigms typically embrace, but then extend beyond the topic of methods (i.e., research design, data collection and analysis) to include deliberations on specific philosophical propositions, most notably existence, knowledge and values (Bahm, 1974; Johnson, Germer, Overton & Overton, 1988). For example, Guba and Lincoln (1994, 2005) suggested researchers explore issues in accordance with an "inquiry paradigm" that reflects their underlying philosophical perspectives. They defined inquiry paradigms as, "the basic belief systems or worldviews that guide the investigator, not only in choices of method, but in ontologically and epistemologically fundamental ways" (Guba & Lincoln, 1994, p. 105). Guba and Lincoln (1994, 2005) did not believe that inquiry paradigms are restricted to methodology, where nonconforming perspectives might represent an incompatibility between the quantitative and qualitative traditions. Rather, they contended inquiry paradigms demonstrate an interdependency between *ontology* (the form and nature of reality), *epistemology* (the relationship between the researcher and what can be known) and

methodology (the manner in which one learns what is to be known) (Guba & Lincoln, 1994, p. 108). Interlaced within each of these is *axiology* (the philosophical discipline concerned with values), where moral forces, either intrinsically or extrinsically based, feed into the inquiry process by helping to define the types of questions that are asked, the paradigmatic choices that guide an investigation, the types of methods that are deployed and the format in which research-based assertions are made accessible (Guba & Lincoln, 2005, p. 197).

Guba and Lincoln (2005) identified the inquiry paradigms that they believed help to structure contemporary social science inquiry: *positivism*, *post-positivism*, *constructivism, critical theory* and *participatory research* (p. 195). An appraisal of the belief systems within those averred paradigms demonstrates how each may be placed along continua representing the dimensions to inquiry they described. By expanding upon conceptualizations originally depicted by Morris (1999), Table 4.1 juxtaposes the five paradigms submitted by Guba and Lincoln (2005) and illustrates the point along those continua where each is positioned with respect to its adherents' views.

Table 4.1
Positioning of Paradigms Across Philosophical Dimensions

Ontology				
Positivism Critical Theory[a]		Post-Positivism	Participatory	Constructivism Critical Theory[b]
Absolute Realism				Absolute Relativism
Epistemology				
Positivism		Post-Positivism		Constructivism Critical Theory Participatory
Objectivism				Interpretivism
Methodology				
Positivism		Post-Positivism		Constructivism Critical Theory Participatory
A priori hypotheses / Focus on hypothesis testing				*A posteriori* hypotheses / Focus on emerging data
Axiology				
Positivism		Post-Positivism		Constructivism Critical Theory Participatory
Ethics extrinsic / Values Excluded				Ethics intrinsic / Values Included

a. See Morris (1999)
b. See Ponterreto (2005)

Instrument Purpose

The purpose of this study then was to develop an instrument to measure views across the dimensions that comprise educational research paradigms. The development of the instrument, entitled the *Research Paradigm Inventory* (RPI), was conducted in the hope of providing a mechanism for future research that will enable the examination of prospective links between the philosophical and methodological orientations of current researchers and graduate students in the field of education. Beyond exploring hypothesized paradigmatic structures and relationships among various philosophical belief systems, the RPI could be useful as a research tool for measuring change across the life of an educational program and career, or in the exploration of variables that contribute to the development of one's research approach altogether. In addition to addressing ideological questions about research paradigms and potentially catalyzing researcher self-reflection, the RPI likewise can serve a more pragmatic role in the edification of graduate students who will serve as future researchers in the field of education. For example, an instrument that delivered valid and reliable measures of the referent concepts could be used in conjunction with other activities to identify the philosophical predilections of graduate students and help guide them toward coursework and research projects in which they might flourish.

Instrument Development

In order to build a strong validity argument for an appropriate interpretation and use of RPI measures, aspects of its development were examined in accordance with the validity framework outlined by Messick (1995). Messick (1995) defined validity as "an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment", wherein the validity of measure interpretation can be evaluated in terms of *content, substantive, structural, generalizability, external* and *consequential* forms of evidence (p. 741). Taken collectively, these six aspects serve as "general validity criteria or standards for all educational and psychological measurement" (Messick, 1995, p. 747). During this study, several of the aspects described by Messick (1995) were addressed through a series of activities designed for instrument development and measure validation using Rasch models (Wolfe & Smith, 2007a, 2007b). Table 4.2 provides an introduction to the activities that were

performed during the course of this study and a link to the specific aspect of validity each was designed to address.

Table 4.2
Instrument Development and Validation Activities Linked to Validity Aspects[a]

	Validity Aspect			
	Content	Structural	Substantive	Generalizability
Validity Evidence	Instrument Purpose	Instrument Dimensionality	Rating Scale Function	Item Calibration Invariance
	Instrument Specifications		Person Fit	Reliability
	Item Development		Item Difficulty Hierarchy	Precision
	Expert Reviews			
	Item Technical Quality			

a. Adapted from Wolfe & Smith, 2007b, p. 244

Instrument Specifications

Content specifications for the instrument were conceptualized following a thorough review of literature and then encapsulated within an instrument blueprint. The blueprint for the instrument originally specified eight content domains, which were comprised of the four inquiry dimensions anchored by extreme views within those respective dimensions (see Appendix B). Once constructed, a small group of individuals who were familiar with concepts related to research paradigms was consulted about the blueprint and asked to provide input regarding its accuracy and relevancy to the construct domains. These individuals confirmed the viability of the blueprint for the intended purpose and judged its specifications to "make sense" and be "uncontroversial" in terms of content and structure, although one individual strongly encouraged the inclusion of a "middle ground" for each of the philosophical dimensions. After much deliberation, I decided that intermediate views would be represented by items reflecting less extreme views within the content domains that had already been conceptualized and modeled accordingly.

Item Development

Each content domain specified by the instrument blueprint was operationalized by a series of items to optimally represent each respective inquiry dimension. Items employed a subject-centered Likert response scale consisting of four ordinal categories: (0) *Strongly Disagree*, (1) *Disagree*, (2) *Agree*, and (3) *Strongly Agree*. Stems for newly crafted items were generated in accordance with content models prevalent in the literature using guidelines for writing rating scale items (Wolfe & Smith, 2007a). A select number of items that had appeared on previously developed instruments were also included, if they exhibited content relevant to the instrument blueprint. In total, 160 items were created or adapted. However, because some items illustrated considerable replicability in terms of content, the item pool was trimmed to 106 items before further reviews were conducted.

Expert Reviews

Content Matching. A group of individuals (n = 8) comprised of practicing researchers, graduate students studying research paradigms and instructors of research methods was invited to review the draft set of 106 items and provide links to the instrument blueprint. To facilitate the linking of items to the content domains of the instrument blueprint, an item-content linking instrument was created. The item-content linking instrument included a table, which outlined and defined the domains of the instrument blueprint, as well as the draft set of 106 items, which had been randomized using a table of random numbers. Directions on the instrument asked respondents to review each of the 106 items carefully and then identify the content domain from the table they believed an item represented.

Two criteria were adopted to support the selection of items that would be retained for pilot testing: (1) at least 6 of the 8 individuals who participated provided a response for an item; and, (2) there was at least 70% agreement among those responding about an item's content. In total, 54 items met these criteria. Because there was a need to ensure adequate representation of each content domain, an additional 6 items were drafted, reviewed and added to the 54 original items that met the selection criteria. This resulted in a 60-item truncated version of the item-content linking instrument, which was then administered to an additional three content experts. Based on their responses, 46 items (76.7%) received unanimous agreement. For those items in

which there was not unanimity, at least 2 of the 3 respondents were in agreement regarding content links for the item.

Item Edits. A group of individuals (n = 4) comprised of measurement professionals, instructors of research methods and graduate students studying measurement was asked to review the draft sets of 60 items and provide comments regarding clarity, succinctness and potential for bias. Following a review of their comments, 15 items were revised and one item was eliminated, yielding a final draft set of 59 items designated for pilot testing.

Item Pilot Tests

The draft 59-item set was included in an online survey created with *Survey Monkey*© and made available for response for the purpose assessing potential item performance. The survey was comprised of the draft 59-item set, which had been randomized using a table of random numbers; demographic variables; and, an open-ended question with which participants could provide additional comments regarding their impressions of the items, the ease with which they were able to respond, the efficacy of the response scale and so forth. Potential respondents (education faculty and graduate students at a research institute located in Southwest Virginia) were sent an invitation to participate via e-mail, which contained a URL link to the survey. In total, 48 individuals responded to the pilot test survey, wherein analyses of item fit, substantive contribution and dimensionality were conducted using parameters estimated from the Rasch Partial Credit Model (Rasch, 1960; Wright & Masters, 1982).

Preliminary analyses of pilot test data entailed several iterations of calibrating parameters for item groupings, evaluating item quality and then performing principal components analyses of standardized residuals to detect systematicity beyond a respective modeled dimension (see Chapter Three). These analyses suggested that response data supported: (a) a clear distinction between *Traditionalist* (e.g., post-positivist) and *Interpretivist* (e.g., constructivist) views; (b) a likely distinction between philosophical and methodological views within both *Traditionalist* and *Interpretivist* item groupings respectively; and, to a lesser extent, (c) a distinction between ontology, epistemology/axiology and methodology. At the completion of these analyses, a total of 47 items were selected for future development activities. An additional 3 items were also drafted to enhance the reliability of potential subscales, yielding a final set of 50 items designated for field tests.

Item Try Outs (Think Aloud Interviews)

To assess the degree to which future respondents might process items that appear on the RPI, as well as identify potential sources of construct-irrelevant variability in response data sets, four individuals completed the draft form of the instrument during a one-on-one interview. Through the course of this process, participants were presented with the draft set of items, asked to 'think aloud' and provide verbal context and rationale for the answers they gave. By and large, participants appeared to process the items as intended, where most of the comments given related to disagreements with item statements, as opposed to issues with the structure of an item or the terminology used. Following a review of participant comments, none of the items that were included on the pilot test survey were deleted, and only a few received minor updates in order to provide clarity in terms of item meaning.

Item Field Tests

A draft of the RPI was formatted and included in a survey created with *Survey Monkey*© in order to field test items and document the psychometric properties of the instrument. Once data were collected, the analyses performed following item pilot tests were replicated to confirm items were adequately representing content domains. During these analyses, a total of 12 items were deleted, yielding a final set of 38 items measuring six dimensions. The six-dimensional structure was noteworthy, in that it appeared to match the instrument specifications to a large degree, as well as a popular depiction of paradigm structure that has appeared in the literature (see Guba & Lincoln, 2005, p. 193). For a complete review of the item and instrument development process for the RPI, see Chapter Three.

Methods for Instrument Measure Validation

Instrumentation

The revised form of the RPI contained 38 items designed to measure six distinct, yet in some cases, related dimensions. The original blueprint for the instrument specified eight content domains designed to measure at least two dimensions (e.g., Traditionalist vs. Interpretivist views). However, analyses of pilot test and field test data suggested the RPI was more accurately gauging six dimensions, which were subsequently modeled as individual subscales labeled:

(1) *Realism in Research,* (2) *Research Objectivism,* (3) *Quantitative Methodology,* (4) *Relativism in Research,* (5) *Research Interpretivism,* and (6) *Qualitative Methodology*. The constituent items for each of the scales are presented in Appendix A.

Participants

Participants for the field test and measure validation phase of this study were petitioned from various online list serves, which included the American Educational Research Association's (AERA) Division D: Measurement and Research Methodology, the list serve of the Graduate Student Forum of the AERA, and the Qualitative Research for the Human Sciences List (QUALRS-L). Potential participants were sent an e-mail invitation that contained a secure, encrypted URL address to which participants could link and respond anonymously. In total, 210 individuals began the survey, with 189 individuals submitting responses that were considered adequate for data analyses (i.e., provided a response for at least 50% of the scale items). Table 4.3 details the demographic characteristics of the individuals who responded.

Table 4.3
Demographic Characteristics of Field Test Respondents

Variable	Category	*n*	% of those responding
Gender	Female	138	77.1
	Male	41	22.9
	Missing	31	--
Role in Education	Graduate Student	122	67.4
	Tenure Track Faculty	29	16.0
	Non-tenure Track Faculty	8	4.4
	Other	22	12.2
	Missing	29	--
Race	White	131	74.4
	Black or African American	13	7.4
	American Indian or Alaska Native	2	1.1
	Asian	10	5.7
	Native Hawaiian/ Other Pacific Islander	1	.6
	Bi/Multi-Racial	6	3.4
	Other	13	6.2
	Missing	34	--
Education Level	Bachelor's Degree	19	10.5
	Master's Degree	99	54.7
	Doctoral Degree	57	31.5
	Other	6	3.3
	Missing	29	--

Analyses

Measurement Model. Response data for each phase of this study were scaled according to variations of the Rasch Polytomous Model (Luo, 2007; Rasch, 1960). The decision to use Rasch analyses during the development of the RPI was based on the presumed importance of providing construct-related evidence of validity for interpreting measures obtained from instrumentation (Messick 1989, 1995). Rasch models may be viewed as instruments of construct validity (Fisher, 1994) by cooperatively working with instrument developers to determine the extent to which data actually measure a hypothesized trait, thereby permitting the strong inference that measured behaviors are indeed expressions of that trait (Bond & Fox, 2007). In other words, by asking empirical data to adhere to a Rasch model's specifications, rather than creating structures to model response data, stronger confidence can be placed in the supposition that an instrument's measures are indeed expressions of some trait(s) of interest.

Because each item on the field test form of the RPI employed an identical response format, responses were scaled to the Rasch Rating Scale Model (Andrich, 1978) given by:

$$\ln(P_{nix} / P_{ni\,x-1}) = B_n - (D_i + F_x)$$

where:

P_{nix} = the probability that participant n will respond in category x on item i
$P_{ni\,x-1}$ = the probability that participant n will respond in category $x-1$ on item i
B_n = the ability for participant n
D_i = the difficulty for item i
F_x = the step that corresponds to an equal probability of selecting category x and $x-1$

To facilitate analyses, parameters for both persons and items were estimated using the joint maximum likelihood estimation (JMLE) procedures contained in *Winsteps 3.62* (Linacre, 2006), where estimates for both item difficulty and person ability (i.e., the degree to which persons exhibited a measured trait) were calibrated along a common metric delineated by logit units (θ), which is the log odds transformation of the probability of a given response option.

Instrument Dimensionality. Evidence relating to the structural aspect of validity for interpreting measures was accumulated through an exploration and confirmation of the instrument's dimensionality. A primary specification for the use of Rasch models is the unidimensionality of measures, which is considered to have been met to a satisfactory extent by a set of instrument data when a dominant component underlies performance on an instrument's scale(s) (Hambleton, 1993). During the instrument development process for the RPI, the procedures used to explore the unidimensional specification for scales were: (a) correct for systematic contradictions to the Rasch dimension using point-measure correlations (r_{pm}), (b) diagnose idiosyncratic response patterns using fit statistics, and (c) examine potential systematic patterns among residuals using principal components analyses (Linacre, 1992).

Preliminary analyses into the dimensionality of the instrument, using the criteria described, suggested a dominant component was underlying responses for each of the six subscales respectively (see Chapter Three). To confirm an optimal depiction of the overall structure for the instrument, a series of model comparisons were made using confirmatory factor analyses (CFA). During these analyses, model fit was assessed using the combination rule of Comparative Fit Index (CFI values ≥ .95) and the Standardized Root Mean Square Residual (SRMR values < .09) as recommended by Hu and Bentler (1999) when $n \leq 250$ (p. 28). In

addition, the Root Mean Square Error of Approximation (RMSEA) was reviewed and is reported for conventional purposes. However, because a recent study (Chen, Curran, Bollen, Kirby & Paxton, 2008) found RMSEA cutoff values ≥ .05 too conservative and RMSEA cutoff values ≤ .10 too liberal when using smaller samples (e.g., ≤ 200), a RMSEA criterion value of .08 was used to identify reasonably acceptable model fit (Browne & Cudeck, 1993). As regards improvement across competing models, statistical significance was determined using a chi-square difference test (Δ_{χ}^{2}) (Steiger, Shapiro & Browne, 1985).

Item Technical Quality. Aside from mapping items to the construct domains of an instrument blueprint, strong content-related evidence of validity requires the adoption of items that demonstrate high technical quality. The technical quality of items that appear on the RPI was explored using point-measure correlations (r_{pm}) and mean square fit statistics. Items that exhibited negative r_{pm} or values ≤ .40 were flagged and reviewed to determine the source of their failure to contribute to a measured dimension. Estimates of item weighted (*infit*) and unweighted (*outfit*) mean squares (*MS*) and the standardized unweighted mean squares ($Z_{unweighted}$) were used to identify potential item misfit. Values for both MS_{infit} and MS_{outfit} were inspected due to the fact that MS_{infit} values are sensitive to non-extreme unexpected responses and MS_{outfit} values are sensitive to extreme unexpected responses (Karabatsos, 2000). Linacre's (2002) guidelines for the interpretation of *MS* values were used to assess potential item misfit, where values > 2.0 distort the measurement system; values = 1.5 - 2.0 are unproductive for the construction of measurement, but not degrading; values = 0.5 - 1.5 are productive for measurement; and, values < .05 are less productive for measurement, but not degrading, and may produce misleading reliabilities and separations.

Unweighted mean squares can be transformed using a cubic root transformation (Smith, Schumacker & Bush, 1998) to produce a standardized mean square fit index (i.e., $Z_{unweighted}$). When data conform to the model, $Z_{unweighted}$ have an expected value of 0.00, a standard deviation near 1.00 and a distribution that is approximate normal (Tejada, Gomez, Garcia & Melendez, 2002). Smith et al. (1998) provided the following suggestions regarding threshold $Z_{unweighted}$ values for flagging items that may exhibit misfit in relation to the sample size being used: N < 1,000, z = 2.00; N > 3,000, z = 3; N = 5,000, z = 4. It should be noted that these values were based on simulations performed by Smith et al. (1998) on dichotomous data. Authors of a recent study concluded that standardized statistics such as $Z_{unweighted}$ were sample size dependent when

used with polytomous data, whereas invariance appears to exist for mean square fit statistics (Smith, Rush, Fallowfield, Velikova & Sharpe, 2008). As such, values for MS_{infit}, MS_{outfit} and $Z_{unweighted}$ were reviewed collectively to assess item fit, as opposed to relying on a single standardized test statistic for rejecting or retaining an individual item.

Rating Scale Function. The substantive aspect of validity refers to the degree to which theoretical rationales relating to both item content and processing models adequately explain the observed consistencies among item responses (Wolfe & Smith, 2007a, p. 205). In this study, the substantive aspect of validity was addressed, in part, through an exploration of item rating scale functioning using guidelines presented by Linacre (2002). As presented, these guidelines were: (1) at least ten observations are present for each response category; (2) each response scale exhibits a unimodal and smooth distribution; (3) average measures advance monotonically across each response category within an item's response scale; (4) unweighted mean squares are less than 2.0; (5) step calibrations advance; (6) ratings imply measures and measures imply ratings (i.e., coherence); (7) step difficulties advance by at least 1.4 logits; and, (8) step difficulties advance by less than 5.0 logits.

Person Fit. The substantive aspect of validity for interpreting measures was also explored through an examination of person fit. Person misfit is seen to occur when respondents with more 'ability' fail to endorse less 'difficult' items and/or respondents with less 'ability' tend to endorse more 'difficult' items. Wolfe and Smith (2007b) suggested that the degree to which person responses conform to the model's expectations can be examined using the standardized mean square unweighted *person* fit index ($Z_{unweighted}$ for *persons*). A $Z_{unweighted}$ for *persons* may assume values - ∞ to ∞, and have an expected value of 0.00 and a standard deviation near 1.00 when data conform to the model. Wolfe and Smith (2007b) suggested using a value of 2.00 as a criterion for flagging potential person misfit. Using this criterion, persons with $Z_{unweighted}$ values $\geq$ 2.00 were flagged, wherein attempts were made to provide plausible explanations for the observed misfit. If appropriate, persons exhibiting extreme misfit (i.e., $Z_{unweighted} > 2.5$) were deleted from subsequent parameter calibrations before additional reviews of item fit and reliability were conducted.

Item Difficulty Hierarchy. An additional means through which substantive evidence of validity can be provided is by demonstrating empirically derived item difficulties correspond to theory-predicted item difficulties (Wolfe & Smith, 2007b, p. 255). Prior to calibrating item

parameters for RPI scales, there was an expectation that certain items would be more or less difficult to endorse by certain respondents depending upon the magnitude of their individual views. By reviewing hierarchical plots of empirically derived item difficulty estimates, an assessment was made regarding the degree to which observed difficulties corresponded to those predicted, as well as the breadth of item coverage across the dimension they were tapping.

Reliability. The generalizability aspect of validity relates to the degree to which score properties and interpretations generalize to and across population groups, settings, and tasks, as well as the generalization of criterion relationships (Wolfe & Smith, 2007a, p. 206). The generalizability aspect was explored, in part, through an analysis of person reliability. Within the context of Rasch measurement, Person Separation Reliability (R_P) may be seen as an analog to coefficient alpha (α), where the degree of internally consistent measures is expressed as one minus the ratio of the average squared errors (MSE_θ) to the variance of measures (V_θ):

$$R_P = 1 - [MSE_\theta / V_\theta].$$

Although R_P and α estimate the same theoretical relationship, R_P may be viewed as a lower bound estimate of true person reliability. As such, and for the benefit of those who are unfamiliar with concepts associated with Rasch measurement, values for both R_P and α are presented, as well as R_P estimates for scales after the removal of persons exhibiting extreme misfit. In terms of assessing the adequacy of reliability estimates for future use in research settings, values of at least .60 were considered adequate, with a goal of producing estimates of .70 or higher.

Precision. The precision of parameter estimates for both persons and items was explored using each estimated parameter's standard error (*SE*), which is the standard deviation of hypothetical parameter estimates around a true parametric value. This index serves to depict the stability of parameter estimates for respondents and items respectively, with more stable estimates depicted by smaller standard errors (Wolfe & Smith, 2007b). Considering this, confidence intervals were constructed around each parameter estimated by the model. Further, information functions (i.e., the inverse of the squared standard error of an ability estimate at a specified level along the latent trait) were examined to explore the amount of information available for a respondent, given his or her level of the measured trait(s).

Item Calibration Invariance. Another important component to building a strong validity argument in terms of generalizability concerns the degree to which item calibrations maintain

their meaning and interpretability across various contexts (Wolfe & Smith, 2007b, p. 261). When items fail to exhibit invariant properties, they are said be functioning differentially (i.e., exhibiting DIF). Although various tests of statistical significance exist for identifying DIF, these were not considered during the course of this study because of the rather modest sample sizes available for variable sub-groups (e.g., males, tenure track faculty). However, in an effort to provide some information about the invariance of item parameters for the RPI, as well as context that might serve future exploration in that regard, a criterion value for Signed Area Index (*SAI*) (Raju, 1988, 1990) was used to identify potentially problematic items. To facilitate analyses, responses were scaled separately for each variable sub-group of interest (i.e., a reference group and focal group) and then standardized by constraining the means and variances of the sets of parameters to be equal. *SAI* was computed to express the distance between pairs of item parameter estimates along the logit scale. Draba's criterion (*SAI* > .50 logits) (Draba, 1977) was then applied to identify items that might potentially be exhibiting DIF. A *SAI* value of .50 logits (i.e., half a logit value) was selected because it corresponds to a shift in probability along the logit scale of about 12%. Flagged items were reviewed, first to determine whether they maintained their rank order across subgroups and then to provide plausible explanations for the observation.

Results

Instrument Dimensionality

In order to determine an optimal representation of structure for the instrument, confirmatory factor analyses (CFA) were performed. In preparation for those analyses and as a step toward providing evidence of substantive validity for interpreting instrument measures, the correlations between all pairs of scale measures were reviewed. Tables 4.4 and 4.5 provide information about the correlation between pairs of measures from each of the instrument's six subscales.

Table 4.4
Correlations Between Pairs of Scale Measures

	Traditionalist Scales			Interpretivist Scales		
Scale	*Realism in Research*	*Research Objectivism*	*Quantitative Methodology*	*Relativism in Research*	*Research Interpretivism*	*Qualitative Methodology*
1. Realism in Research	1.000					
2. Research Objectivism	.581	1.000				
3. Quantitative Methodology	.523	.642	1.000			
4. Relativism in Research	-.286	-.264	-.356	1.000		
5. Research Interpretivism	-.317	-.530	-.391	.369	1.000	
6. Qualitative Methodology	-.288	-.424	-.648	.474	.444	1.000

NOTE: All correlations were statistically significant with $p < .01$ (2-tailed).

Table 4.5
Correlations Between Pairs of Scale Measures - Corrected[a]

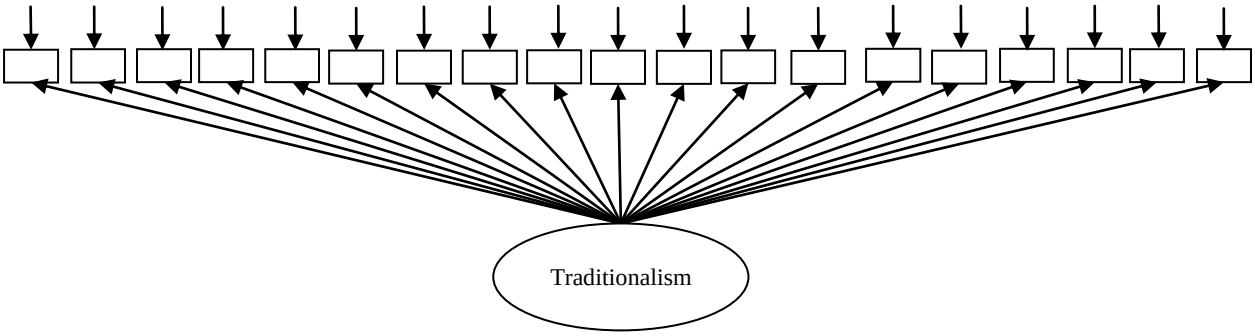
	Traditionalist Scales			Interpretivist Scales		
Scale	*Realism in Research*	*Research Objectivism*	*Quantitative Methodology*	*Relativism in Research*	*Research Interpretivism*	*Qualitative Methodology*
1. Realism in Research	1.000					
2. Research Objectivism	.745	1.000				
3. Quantitative Methodology	.655	.783	1.000			
4. Relativism in Research	-.401	-.341	-.475	1.000		
5. Research Interpretivism	-.417	-.680	-.489	.517	1.000	
6. Qualitative Methodology	-.374	-.538	-.801	.656	.577	1.000

a. Coefficients were corrected for attenuation according to $r_{xy} /(r_{xx} * r_{yy})^{½}$, where r_{xy} represents the correlation between pairs of measures from scales X and Y, r_{xx} is the R_P estimate for scale X and r_{yy} is the R_P estimate for scale Y.

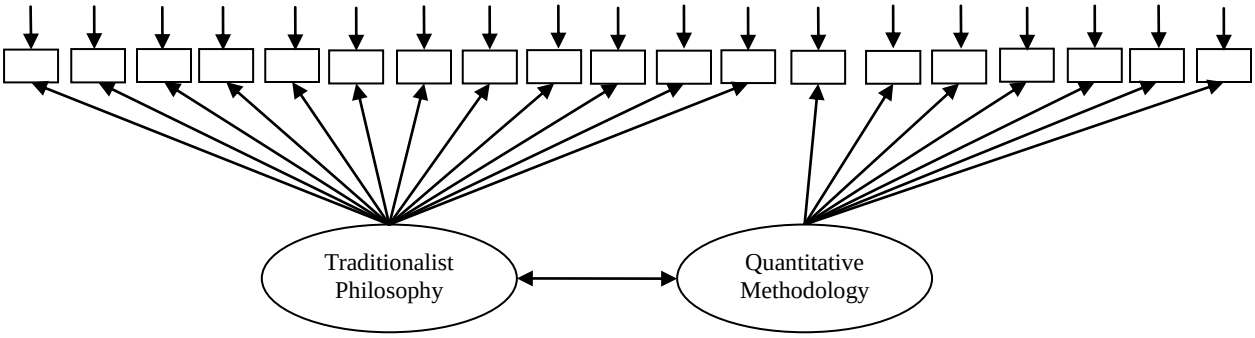
Considering the associations that appeared to exist between pairs of scale measures, coupled with the theoretical propositions that supported the creation of the instrument blueprint, it was determined that any accurate representation of structure for the instrument would necessarily consist of an oblique factors model. Therefore, a series of comparative analyses across oblique factor models was conducted for the three Traditionalist and three Interpretivist scales respectively. The first set of comparative analyses was conducted for the Traditionalist scales and included a one factor model (*Traditionalism*), a two correlated factors model (*Traditionalist Philosophy, Quantitative Methodology*) and a three correlated factors model (*Realism in Research, Research Objectivism, Quantitative Methodology*). The second set of

comparative analyses was conducted for the Interpretivist scales and included a one factor model (*Interpretivism*), a two correlated factors model (*Interpretivist Philosophy, Qualitative Methodology*) and a three correlated factors model (*Relativism in Research, Research Interpretivism, Qualitative Methodology*). Figures 4.1 and 4.2 provide a conceptual depiction of each series of competing models, both for the Traditionalist and Interpretivist comparisons respectively.

Model 1 (One Factor)



Model 2 (Two Correlated Factors)



Model 3 (Three Correlated Factors)

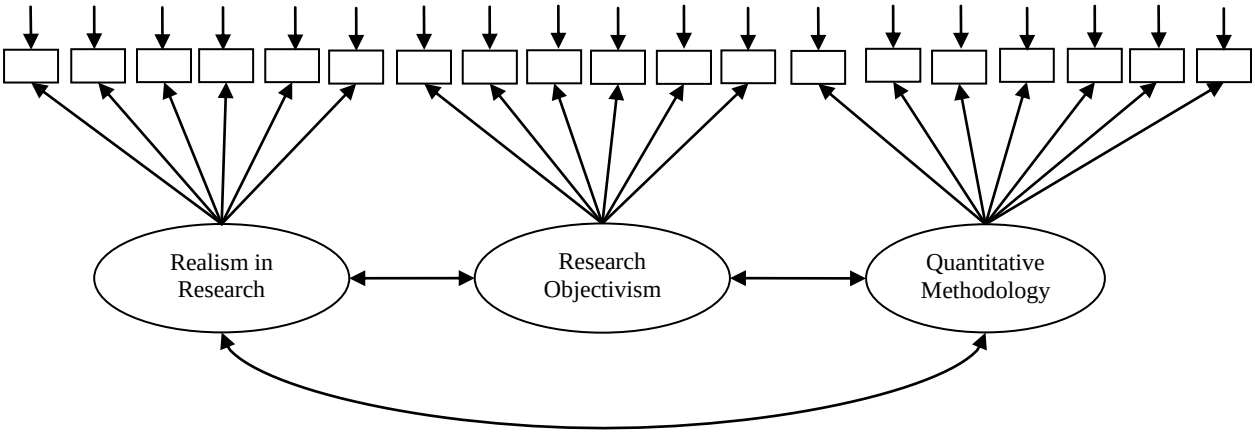
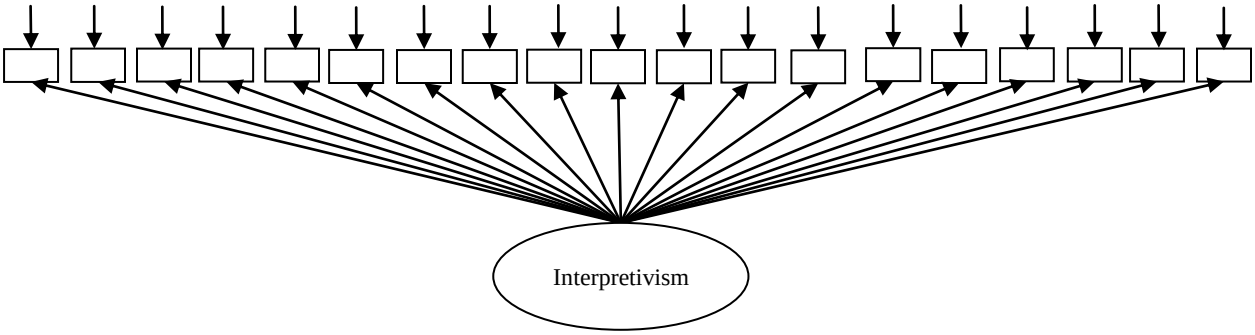
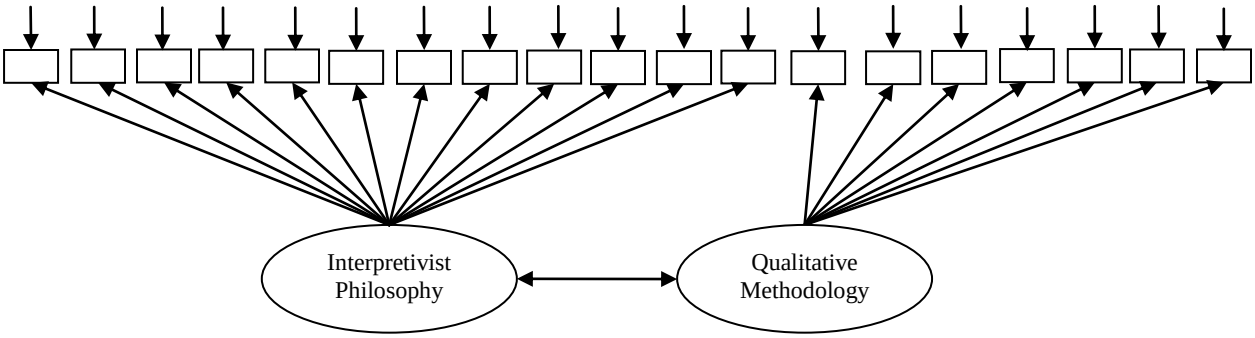


Figure 4.1
Conceptual Depiction of Traditionalist Model Comparisons

Model 4 (One Factor)



Model 5 (Two Correlated Factors)



Model 6 (Three Correlated Factors)

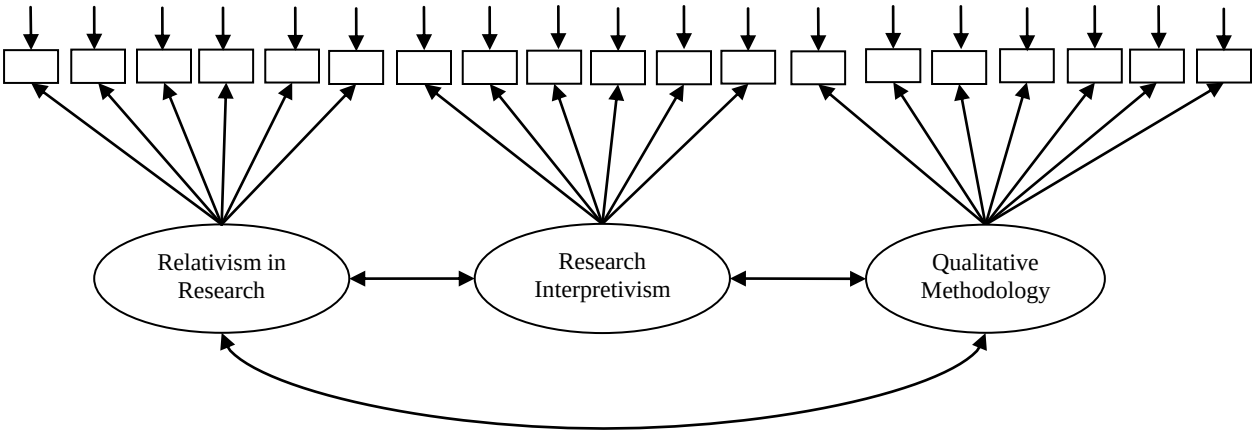


Figure 4.2
Conceptual Depiction of Interpretivist Model Comparisons

To facilitate analyses, model parameters were estimated from the inter-item correlation matrix using the Maximum Likelihood estimation procedures in *LISREL 8.72* (Jöreskog & Sörbom, 2005). Maximum Likelihood estimation was chosen because of its fairly robust performance in less than optimal analytic conditions such as small sample size and excessive kurtosis (Hoyle & Panter, 1995, p. 164), which were issues for this particular data set. With n = 175, indices of model fit produced from each of the respective CFA analyses are detailed in Tables 4.6 and 4.7.

Table 4.6
CFA of Traditionalist Scales - Indices of Model Fit

Model	$\chi^2_{(df)}$	*CFI*	*SRMR*	*RMSEA*
1. One Factor Model	$528_{(152)}$	.92	.084	.12
2. Two Oblique Factors Model	$395_{(151)}$	.95	.074	.10
3. Three Oblique Factors Model*	$287_{(149)}$	.97	.065	.07

NOTE: 90% Confidence Interval for $RMSEA_{Model\ 3}$ = .060 ; .086

Table 4.7
CFA of Interpretivist Scales - Indices of Model Fit

Model	$\chi^2_{(df)}$	*CFI*	*SRMR*	*RMSEA*
4. One Factor Model	$519_{(152)}$	.86	.093	.12
5. Two Oblique Factors Model	$412_{(151)}$	.90	.084	.10
6. Three Oblique Factors Model*	$239_{(149)}$	.96	.065	.06

NOTE: 90% Confidence Interval for $RMSEA_{Model\ 6}$ = .045 ; .072

As depicted by Tables 4.6 and 4.7, the one factor Models 1 and 4 produced unsatisfactory results and were associated with relatively poor levels of fit. The two factor Models 2 and 5 were statistically significantly better fitting than each of the one factor models ($\Delta\chi^2_{Model\ 1\ vs.\ Model\ 2}$ = $133_{(1)}$ and $\Delta\chi^2_{Model\ 4\ vs.\ Model\ 5}$ = $107_{(1),}$ respectively). However, these improvements were marginal in terms of producing a well-fitting model. The three factor Models 3 and 6 were statistically significantly better fitting than the two factor models ($\Delta\chi^2_{Model\ 2\ vs.\ Model\ 3}$ = $108_{(2)}$ and $\Delta\chi^2_{Model\ 5\ vs.\ Model\ 6}$ = $173_{(2),}$ respectively) and associated with relatively better levels of model fit. Indices of fit for the three factor Models 3 and 6 fell within the cutoff values for CFI, SRMR and RMSEA. As

a cautionary note, although the upper band of a 90% confidence interval (CI) for $RMSEA_{Model\ 6}$ fell below the .08 criterion, this was not the case for $RMSEA_{Model\ 3}$, where the upper band was observed to be .086. Again though, the efficacy of using CI for RMSEA in absolute terms was seen as questionable in this instance, given the relatively small sample size available for these analyses (see Chen et al., 2008).

A review of the modification indices provided in *LISREL 8.72* indicated the greatest improvement to model fit could be accommodated by including a correlated errors term for a few pairs of items within the certain scales, as opposed to allowing items to load across scales. For example, including a correlated errors term for a single pair of same scale items in Model 3 improved overall model fit with $\chi^2 = 253_{(148)}$, CFI = .97, SRMR = .063 and RMSEA = .06. Likewise, including a correlated errors term for a single pair of same scale items in Model 6 improved overall model fit with $\chi^2 = 219_{(148)}$, CFI = .97, SRMR = .063 and RMSEA = .05. After re-scaling each of these pairs of items to the Rating Scale Model and then plotting their resulting measures against the sets of measures originally obtained from their respective scales, it was determined that the impact of these correlated items was not substantial in terms of confounding the measurement system. Rather, it was believed that these item pairs simply represented highly correlated items that were matched, but not overmatched, in terms of content and difficulty estimates. Further, because the exercises described were confirmatory, as opposed to exploratory, additional modification to the models was not indicated, as such modification would essentially represent parameterization simply for the sake of producing a better fitting model. Rather, given the sample size available, the goal of confirming instrument dimensionality and the inability to disconfirm Models 3 and 6 based on their satisfactory fit, it was determined the most reasonable depiction of structure for the instrument at this point in time entailed six dimensions, which were believed to be well-represented by the six modeled scales.

Item Technical Quality

Table 4.8 provides information about average difficulty estimate for items, their unweighted standardized mean square fit statistics ($Z_{unweighted}$) and their correlation (r_{pm}) with measures produced by the scales of which they are a part. In general, most items fit within the criteria designated for determining adequate fit and useful contribution. Across the six scales, all items exhibited $r_{pm} > .40$ and mean square fit statistics that fell within recommended .5 and 1.5

values. With the exception of Item 30, where MS_{infit} = .69 and MS_{outfit} = .69, all items exhibited MS_{infit} and MS_{outfit} values > .70 and < 1.3. In total, only six items were flagged based on standardized outfit values (i.e., MS_{outfit} > 2.0). However, in most cases, these items exhibited adequate fit and contribution according to the other indices employed. Where potential item misfit was seen to occur, the rationale for why an item was retained is described.

Table 4.8
Summary of Item Quality Indicators

Scale	*Statistic*	*Logit*	$Z_{unweighted}$	r_{pm}	*Scale*	*Statistic*	*Logit*	$Z_{unweighted}$	r_{pm}
1. Realism in Research	Mean	.00	-.2	--	*4. Relativism in Research*	Mean	.00	.0	--
	SD	.96	1.5	--		SD	.76	1.1	--
	Minimum	-1.02	-2.2	.62		Minimum	-.95	-1.6	.46
	Maximum	1.70	2.3	.76		Maximum	1.28	2.1	.73
	N Flagged	--	1[a]	--		N Flagged	--	1[c]	--
2. Research Objectivism	Mean	.00	-.1	--	*5. Research Interpretivism*	Mean	.00	-.1	--
	SD	.94	1.3	--		SD	.66	1.7	--
	Minimum	-1.14	-2.3	.68		Minimum	-.76	-2.7	.59
	Maximum	1.93	1.6	.82		Maximum	1.23	2.5	.76
	N Flagged	--	--	--		N Flagged	--	1[d]	--
3. Quantitative Methodology	Mean	.00	-.2	--	*6. Qualitative Methodology*	Mean	.00	-.2	--
	SD	1.09	1.5	--		SD	.73	2.1	--
	Minimum	-2.17	-2.5	.67		Minimum	-.90	-3.3	.52
	Maximum	1.38	2.1	.83		Maximum	1.39	2.4	.79
	N Flagged	--	1[b]	--		N Flagged	--	2[e]	--

a. Item 8 was retained because it contained essential instrument content and exhibited adequate mean square fit statistics (MS_{infit} = 1.25, MS_{outfit} = 1.25), strong correlation with the modeled dimension (r_{pm} = .66) and acceptable standardized fit ($Z_{unweighted}$ = 1.9) after deleting persons exhibiting extreme misfit.
b. Item 19 was retained because it exhibited important instrument content, adequate mean square fit statistics (MS_{infit} = 1.17, MS_{outfit} = 1.25), strong correlation with the modeled dimension (r_{pm} = .67) and acceptable standardized infit ($Z_{weighted}$ = 1.7). In addition, after deleting persons exhibiting extreme misfit from subsequent analyses, items 19 exhibited adequate fit across all indices (MS_{infit} = 1.03, MS_{outfit} = 1.06, $Z_{weighted}$ = .5 and r_{pm} = .73).
c. Item 29 was retained because it exhibited important instrument content, adequate mean square fit statistics (MS_{infit} = 1.21, MS_{outfit} = 1.29) and strong correlation with the modeled dimension (r_{pm} = .68).
d. Item 4 was retained after deleting persons exhibiting extreme misfit resulted in adequate fit across all indices (MS_{infit} = 1.12, MS_{outfit} = 1.13, $Z_{unweighted}$ = 1.2 and r_{pm} = .66).
e. Item 14 was retained because it exhibited important instrument content and adequate mean square fit statistics (MS_{infit} = 1.22, MS_{outfit} = 1.27). Item 30 was retained because it exhibited important instrument content, adequate mean square fit statistics (MS_{infit} = 1.27, MS_{outfit} = 1.21) and strong correlation with the modeled dimension (r_{pm} = .63).

Rating Scale Function

An initial review of response category frequencies demonstrated that some items failed to receive the recommended number of ten responses in all end-point categorical options (i.e., *Strongly Disagree* and *Strongly Agree*). As such, the interpretation of rating scale efficacy, in

addition to the accuracy and stability of less-centralized parameter estimates, would have to be met with a certain degree of caution. With that reproach in mind, the adequacy of the item response scale for each of the instrument's six scales was examined in accordance with Linacre's (1998) criteria, the results of which are presented in Table 4.9.

Table 4.9
RPI Rating Scale Analyses

	Realism in Research	*Research Objectivism*	*Quantitative Methodology*	*Relativism in Research*	*Research Interpretivism*	*Qualitative Methodology*
$\boldsymbol{MS_{outfit}}$						
Person	.97	.98	.97	.98	1.00	.98
Item	.98	.98	.98	1.00	.99	.99
Linacre's Criteria						
$N_g > 10$	☒	☒	☒	☒	☒	☒
Distributions	☑	☑	☑	☑	☑	☑
$M(\theta)$	☑	☑	☑	☑	☑	☑
$MS_{unweighted}$	☑	☑	☑	☑	☑	☑
τ *increase* > 1.4	☑	☑	☑	☑	☑	☑
τ *distance* < 5.0	☑	☑	☑	☑	☑	☑
$Coherence_{Measures}$	☑	☑	☑	☑	☒	☑
$Coherence_{Categories}$	☒	☒	☒	☒	☒	☒

NOTE: ☑ - Indicates criterion was met.
NOTE: ☒ - Indicates criterion was not met.

A closer examination of response category frequencies revealed that a few Traditionalist scale items failed to receive the recommended number of responses within the *Strongly Agree* category and a few Interpretivist scale items failed to receive the recommended number of responses in the *Strongly Disagree* category. This pattern likely contributed to the lack of sufficient coherence statistics for categories (see Table 4.9), which represents the percentage of observations in a category that matched model-based expected ratings. Wolfe and Smith (2007b) recommend that coherence statistics be greater than 40% for each category within an item response scale. In the present study, each category met this criterion, with the exceptions of *Strongly Agree* for items within the Traditionalist scales and *Strongly Disagree* for items within the Interpretivist scales. The implication for the lack of category coherence in this instance was

that inferences from these data might be stronger if categories 0 and 1 were combined for the Traditionalist scales, and categories 2 and 3 combined for the Interpretivist scales.

In order to explore potential improvements to the item response scale and subsequently, item model-data fit statistics, alternative rating scale schema were explored by collapsing responses from adjacent categories. Given the original structure of the item response scale in terms of label connotation, coupled with the need to remain substantively consistent, only schema that combined *Strongly Disagree* and *Disagree* or *Agree* and *Strongly Agree* were considered. Combining these adjacent categories did provide an improvement in terms of category to measures coherence. However, these improvements were made at the cost of producing an unacceptable reduction in reliability estimates for scales. Although reliability was not used as a primary criterion for constructing scales, it was still a preeminent consideration for producing acceptable separation of persons across the measured traits. Because of this, the original four-point response scale was retained and is suggested for future use until such time as additional analyses of rating scale functioning are conducted, where persons with more extreme views are targeted during data sampling.

Person Fit

Table 4.10 provides information about the degree to which cases of person misfit were observed across each of the instrument's scales. An examination of the response strings for these individuals demonstrated that the observed misfit was likely due to individuals with low ability measures endorsing one or two difficult items, or individuals with high ability estimates failing to endorse one or two items that were closely matched to them along the latent trait measured by a respective scale. In order to assess the degree to which person misfit might have impacted the measurement system, persons exhibiting extreme misfit (i.e., $Z_{unweighted} \geq 2.5$) were removed from the data set and model parameters re-calibrated for each of the six scales. The removal of these persons did improve scale reliabilities to a small extent (see Table 4.12), and for two items, bring fit statistics for flagged items into acceptable ranges. However, these improvements were marginal, as the inclusion of persons exhibiting extreme misfit did not appear to substantially impact the measurement system, where the difficulty estimates for items and their hierarchy remained consistent across all sets of parameter calibrations.

Table 4.10
Summary of Person Misfit

Scale	Mean $Z_{unweighted}$	Minimum $Z_{unweighted}$	Maximum $Z_{unweighted}$	Number Flagged > 2.0	%	Number Flagged > 2.5
1. Realism in Research	-.1	-2.4	3.1	10	5.3 %	5
2. Research Objectivism	-.2	-2.4	4.9	8	4.2 %	6
3. Quantitative Methodology	-.1	-2.5	5.4	10	5.3 %	7
4. Relativism in Research	-.2	-2.1	4.9	12	6.3 %	5
5. Research Interpretivism	-.2	-2.2	5.3	9	4.8 %	4
6. Qualitative Methodology	-.1	-2.9	5.0	8	5.3 %	5

Item Difficulty Hierarchy

An examination of the hierarchy of item difficulty estimates suggested that the parameters estimated by the model tended to correspond with those envisioned during instrument construction, especially in those cases where items were created to reflect extreme views. In general, items representing more extreme views within a respective scale were endorsed by individuals who were closely matched to that item, or in cases where a person's ability estimate exceeded the difficulty estimate for an item. For example, within the *Quantitative Methodology* scale, Item 19 (*The strength of theory is best determined by using data-based quantitative methods*) received a higher difficulty estimate than Item 41 (*Replicable findings lead to valid research conclusions*). From a conceptual standpoint, this seemed reasonable, as a person who endorses the use of mixed methods would likely respond affirmatively with *Agree* or *Strongly Agree* to Item 41. However, when asked about the preeminence of quantitative methods, such as in the case of Item 19, these individuals were likely to respond with *Disagree* or *Strongly Disagree*. Appendix A provides a list of items that were included in each scale, as well as information about each item's difficulty estimate along the latent trait measured by respective scales.

As regards the breadth of item coverage along the latent trait continua, Table 4.11 reports the average person measure observed within each response category of a given scale. A review of these averages suggests that persons are given a wide range of possibilities for representing their views, in terms of magnitude, along each of the six scales included on the RPI.

Table 4.11
Average Person Measure Observed in Rating Scale Categories

Scale	Strongly Disagree	Disagree	Agree	Strongly Agree
1. Realism in Research	- 4.75	- 2.13	.96	3.65
2. Research Objectivism	- 4.61	- 1.76	1.75	4.62
3. Quantitative Methodology	-4.74	-1.95	1.81	5.01
4. Relativism in Research	- 3.79	- 1.50	1.33	4.11
5. Research Interpretivism	- 4.81	- 1.90	1.85	4.90
6. Qualitative Methodology	- 3.63	- 1.18	1.27	3.47

Reliability

Table 4.12 provides information about the internal consistency of Rasch measures, as well as raw scores, by reporting indices of Rasch Person Separation Reliability (R_P) , R_P after deleting persons exhibiting extreme misfit and coefficient alpha (α) for the benefit of those who are unfamiliar with concepts related to Rasch measurement.

Table 4.12
Internal Consistency of Measures Produced by Scales (Reliability)

Scale	R_P	R_P [a]	α
1. Realism in Research	.74	.76	.79
2. Research Objectivism	.75	.80	.84
3. Quantitative Methodology	.82	.84	.85
4. Relativism in Research	.63	.67	.73
5. Research Interpretivism	.74	.76	.77
6. Qualitative Methodology	.75	.78	.81

a. R_P after removal of persons exhibiting extreme misfit (i.e., $Z_{unweighted} \geq 2.5$).

With the exception of *Relativism in Research,* where R_P = .63, all scales produced reliability estimates that were greater than .70 in magnitude. On average, the instrument's scales appear to be able to delineate persons across at least two statistically different performance strata the instrument identified in the sample. Considering the item response format asked participants to agree or disagree with statements with varying degrees of magnitude therein, it is suggested that the reliability of separation for persons provided by instrument scales is sufficient for future research, which is the purpose for which the instrument was created.

Precision

The precision of parameters estimated by the model was explored through a review of their respective standard errors. Standard errors for item difficulty estimates ranged from .12 to .16 logits across all six scales. The size of a 95% confidence interval (95% CI) for the item with the largest standard error was ± .31. As regards the precision of person estimates, ten individuals were selected to help represent the full range of abilities across the traits measured by the six scales. Individual measures estimated for the persons selected are depicted in Tables 4.13 – 4.18, as well as the standard error (*SE*) and a 95% confidence interval (95% CI) for those point estimates. In addition, the maximum and minimum estimated measures from the scales are highlighted.

Table 4.13
Indices of Precision for Persons Measured by *Realism in Research* Scale

		← -6.18 logits (*SE* = 1.88)									→ 6.51 logits (*SE* = 1.91)
	Mean θ	Person 164	Person 3	Person 93	Person 1	Person 53	Person 12	Person 146	Person 181	Person 24	Person 113
Logit Measure	-.38	-4.82	-3.20	-2.09	-1.11	-.01	.93	2.05	3.09	4.03	5.07
SE	.77	1.11	.78	.71	.69	.88	.74	.76	.87	.92	1.16
95% CI Upper	1.13	-2.64	-1.67	-.69	.24	1.71	2.38	3.54	4.80	5.83	7.34
95% CI Lower	-1.89	-6.99	-4.73	-3.40	-2.46	-1.73	-.52	.56	1.38	2.23	2.80

Table 4.14
Indices of Precision for Persons Measured by *Research Objectivism* Scale

		← -6.73 logits (*SE* = 1.88)									→ 6.93 logits (*SE* = 1.94)
	Mean θ	Person 153	Person 28	Person 37	Person 3	Person 1	Person 124	Person 172	Person 177	Person 166	Person 107
Logit Measure	-.06	-5.36	-4.39	-3.02	-2.38	-1.16	-.03	1.16	2.38	3.61	5.43
SE	.84	1.18	.82	.77	.78	.75	.76	.80	.80	.89	1.12
95% CI Upper	1.59	-3.04	-2.78	-1.51	-.85	.31	1.46	2.73	3.95	5.35	7.62
95% CI Lower	-1.61	-7.67	-5.99	-4.53	-3.91	-2.63	-1.52	-.41	.81	1.87	3.23

Table 4.15
Indices of Precision for Persons Measured by *Quantitative Methods* Scale

		← -5.42 logits (*SE* = 1.16)								7.02 logits (*SE* = 1.88) →
	Mean θ	Person 37	Person 16	Person 32	Person 188	Person 110	Person 142	Person 169	Person 171	Person 91	Person 113
Logit Measure	.37	-4.40	-3.15	-2.20	-.86	.05	1.04	2.02	3.11	4.69	5.65
SE	.75	.90	.72	.67	.67	.69	.72	.84	.85	.88	1.11
95% CI Upper	1.84	-2.64	-1.74	-.89	.45	1.40	2.45	3.67	4.78	6.41	7.83
95% CI Lower	-1.10	-6.16	-4.56	-3.51	-2.17	-1.30	-.37	.37	1.44	2.97	3.47

Table 4.16
Indices of Precision for Persons Measured by *Relativism in Research* Scale

		← -5.56 logits (*SE* = 1.89)								6.20 logits (*SE* = 1.89) →
	Mean θ	Person 90	Person 24	Person 66	Person 48	Person 222	Person 134	Person 1	Person 150	Person 181	Person 23
Logit Measure	.95	-3.56	-2.94	-1.96	-1.09	-.05	1.08	1.91	3.11	3.97	4.81
SE	.78	.84	.75	.67	.66	.77	.83	.76	.81	1.29	1.13
95% CI Upper	2.48	-1.91	-1.47	-.65	.20	1.46	2.71	3.40	4.70	6.50	7.02
95% CI Lower	-.58	-5.21	-4.41	-3.27	-2.38	-1.56	-.55	.42	1.52	1.44	2.60

Table 4.17
Indices of Precision for Persons Measured by *Research Interpretivism* Scale

		← -3.48 logits (*SE* = .88)								6.94 logits (*SE* = 1.89) →
	Mean θ	Person 125	Person 14	Person 8	Person 29	Person 101	Person 107	Person 186	Person 18	Person 17	Person 60
Logit Measure	1.48	-3.48	-2.53	-1.26	-.08	1.06	1.98	3.20	3.85	4.57	5.55
SE	.84	.88	.80	.78	.77	.89	.92	.80	.82	.89	1.13
95% CI Upper	3.13	-1.76	-.96	.27	1.43	2.80	3.78	4.77	5.46	6.31	7.76
95% CI Lower	-.17	-5.20	-4.10	-2.79	-1.59	-.68	.18	1.63	2.24	2.83	3.34

Table 4.18
Indices of Precision for Persons Measured by *Qualitative Methods* Scale

		← -5.42 logits (*SE* = 1.16)									→ 5.67 logits (*SE* = 1.87)
	Mean θ	Person 172	Person 75	Person 187	Person 174	Person 76	Person 8	Person 1	Person 140	Person 67	Person 13
Logit Measure	.87	-2.98	-2.02	-.82	.02	.59	.93	2.00	2.94	3.46	4.33
SE	.65	1.09	.82	.82	.62	.58	.58	.64	.61	.66	.74
95% CI Upper	2.14	-.84	-.41	.78	1.24	1.73	2.07	3.25	4.14	4.75	5.78
95% CI Lower	-.40	-5.11	-3.63	-2.43	-1.20	-.55	-.21	.75	1.74	2.17	2.88

Clearly, each of the respective scales tended to provide estimates for this sample that were most precise along the center of the logit scale. In other words, as person estimates deviated from centrality, their associated standard errors grew in magnitude. This was not unexpected, in that most researchers are likely to hold views that are less extreme in relative terms. However, confidence bands around point estimates of person ability appeared much larger than what was anticipated, which is an issue that will be addressed through future validation activities. This observation is perhaps due to the relatively small sample, as well as the ambiguity normally associated with many of the philosophical concepts measured by the RPI scales.

Item Calibration Invariance

As an exploratory exercise, items exhibiting potential DIF were identified using Draba's Criterion ($SAI \geq |.50|$ logits). These analyses included comparisons between parameter estimates from sub-groups within gender (males and females) and role in education (graduate students and faculty). In total, 12 items were flagged across the six scales for exhibiting $SAI \geq |.50|$ logits. The maximum number of flagged items within a given scale was two. In all cases, *SAI* for flagged items were < .80 logits, and all but two were < .70 (*SAI* for 5 items was < .60). Table 4.19 provides information about the number of flagged items based on *SAI* across each sub-group comparison set.

Table 4.19
Exploratory DIF Analyses of RPI Items

		GROUP COMPARISON	
		Graduate Students vs. Faculty	*Females vs. Males*
Scale	*Number of items*	*Number of SAI* ≥ .50 logits	*Number of SAI* ≥ .50 logits
1. Realism in Research	6	0	1
2. Research Objectivism	6	0	2
3. Quantitative Methodology	7	2	0
4. Relativism in Research	6	2	0
5. Research Interpretivism	6	1	1
6. Qualitative Methodology	7	1	2

NOTE: As a point of information, all standardized *SAI* were non-statistically significant with $p < .05$. However, as described, the limited sample size for sub-groups of variables draws into question the fidelity of parametric testing within the context of this study.

Among those items that were flagged, only items 15 and 17 produced difficulty estimates that were discrepant in terms of their rank ordering across sub-groups. Item 15 (*One of my objectives for research is to understand the individual realities that exist in the world*) produced a *SAI* = .70 between graduate students and faculty. In terms of substantive differences, this item appeared to be more difficult for faculty to endorse, perhaps due to their presumed exposure to philosophical concepts that graduate students might not otherwise receive during their normal coursework. Item 17 (*I need to have a strong background in qualitative research*) produced a *SAI* = .62 and appeared to be more difficult for males to endorse. This may be due to a residual effect of females filtering out of upper level mathematics and science at a much higher rate than males, wherein males would presumably be more comfortable with quantitative methods and less comfortable with qualitative methods (Nosek, Mahzarin & Greenwald, 1998). In either case, the observed shift for males spanned just one item in terms of rank order. Further, estimates for both pairs of groups fell well below the average difficulty estimate for items in the associated scale, where the $D_{females}$ was = -1.03 and D_{males} was = -.41 respectively.

As noted, the statistical significance of *SAI* was not considered during the course of this study due to the small sample sizes available for variable sub-groups. As such, any conclusions about or justification for potential differences between parameter calibrations is purely speculative at this point in time. However, considered what was observed, it does appear as if item parameter estimates are relatively invariant for this sample, at least in terms of their ability

to maintain a rank order. Where potential areas of concern were observed, results were noted and will be considered during future explorations into the invariance of RPI items.

Relative Performance of Combined Scales

Early analyses into the dimensionality of the RPI revealed its most viable structure to be across six dimensions. However, the strong associations between pairs of measures from certain scales suggests also that it would not be inappropriate to assign a total score if one were attempting to represent a *Traditionalist* or *Interpretivist* set of beliefs respectively. As such, the three *Traditionalist* scales and the three *Interpretivist* scales were combined to produce a single score for *Traditionalism* and a single score for *Interpretivism*. Indices of performance for these full scales are highlighted in Table 4.20.

Table 4.20
Performance Indices for Full Scales

	Traditionalist Items	Interpretivist Items
Number of Items	19	19
R_P	.89	.83
α	.91	.87
$r_{pm} \geq .50$	17 items	13 items
$r_{pm} \geq .40$	19 items	18 items
$MS_{infit} \geq 1.5$	0 items	0 items
$Z_{unweighted} \geq 2.0$	3 items	2 items
Total Var_θ explained by Rasch Dimension	62.1%	40.4%
Total Var_θ explained by 1st Residual Component	5.6%	8.5%
Minimum Item Difficulty Estimate and *SE*	-1.90 (.12)	-.70 (.12)
Maximum Item Difficulty Estimate and *SE*	1.69 (.13)	1.31 (.11)
Mean Item Difficulty Estimate and *SE*	.00 (.12)	.00 (.12)
Minimum Person Ability Estimate and *SE*	-5.57 (.36)	-2.17 (.34)
Maximum Person Ability Estimate and *SE*	4.24 (1.04)	4.01 (.63)
Mean Person Ability Estimate and *SE*	-.03 (.39)	.77 (.37)

Conclusions

The goal for this study was to develop an instrument, which provides measures of the various dimensions that comprise research paradigms in education. Given the evidence that has been presented, it is suggested that RPI scales provide reasonably reliable and defensible estimates of individuals' attitudes toward various research practices and beliefs. Empirical evidence for the dimensionality of the instrument, coupled with observed relationships between certain perspectives, appear to support a popular typology for paradigms that has emerged in the literature (e.g., Guba & Lincoln, 1994, 2005), at least in terms of describing the components that structure such, if not content therein. Because of the likelihood for this structure to hold in other contexts, it is anticipated these scales will serve as useful tools for future research studies that explore paradigms in greater detail. Such studies may include measuring the impact that a graduate program of study has on the development of a student's approach to inquiry, or even how one's theoretical predilections change over time. In addition, continued focus will be given to refining the RPI through the use of larger cross-validation samples in which persons with more extreme views are targeted to reduce standard errors for parameter estimates, as well as studies that explore the external and consequential aspects of validity for interpreting and using RPI measures.

CHAPTER 5
DISCUSSION

The goal for this study was to develop an instrument, which provides measures of the various dimensions that comprise research paradigms in education. Given the evidence presented, the *Research Paradigm Inventory* (RPI) appears to provide reasonably reliable and defensible estimates of individuals' attitudes toward various research practices and beliefs. In addition, the observed relationships between measures from RPI scales help to support a popular typology for paradigms that has emerged in the literature (Guba & Lincoln, 1994, 2005), at least in terms of structure, if not content therein. Because of the likelihood for this structure to hold in future contexts, these scales should serve as useful tools for research studies that explore paradigms in greater detail. Such studies may include measuring the impact a graduate program of study has on the development of a student's approach to inquiry or even how one's theoretical predilections change over time.

In addition to using the instrument in research settings, further focus should be given to refining RPI scales as appropriate, most likely through the use of cross-validation samples in which persons with more extreme views are targeted to reduce error in parameter estimates. Supplementary activities should also include a focus on the external aspect of validity for interpreting RPI measures through Multitrait-Multimethod analyses or perhaps even qualitative analyses (e.g., interviews, document analyses of research articles). Qualitative efforts could include examining the degree to which instrument measures correspond to how researchers and graduate students views themselves, which would certainly contribute to building a stronger case of validity for RPI measure use and provide further context for identifying the various belief systems that researchers hold in this regard.

As noted, the reason for developing the RPI was to explore a hypothesized structure for research paradigms and then measure corresponding views therein. These reasons did not include arguing for the supremacy of any particular paradigm or authentication of all paradigms that might exist at this point in time. Given the controversy surrounding the legitimacy of some of the paradigms arguably evident in contemporary educational research, coupled with the concept of blurring genres across paradigms (Geertz, 1988), such claims would be premature until the viability of individual paradigms can be determined. Because their legitimacy is often based in

argumentative discourse, paradigms sometimes appear as idiosyncratic, abstruse and a sensitive topic for some. Evidence for this appeared to manifest during the course of this study through responses given to the open-ended questions on the field test survey, where one individual commented the instrument was an attempt to "resurrect the paradigm wars of the past", while another stated the survey "appeared to be highly biased toward qualitative research." These comments were particularly interesting, considering that an equal number of Traditionalist and Interpretivist items were selected in an effort to represent many different views, as opposed to promulgate one particular outlook over another. A few respondents also commented that they felt like they wanted to respond with "it depends" to some of the items, or that an attempt was being made to dichotomize their belief systems. In response, it must be stated that the purpose for developing the RPI was not to dichotomize views (e.g., label someone as quantitative or qualitative), but rather to model an assortment of views with varying degrees. A necessary condition for capturing views across the respective dimensions measured by the RPI scales required the adoption of items that reflected relatively extreme concepts, as well as those that might be considered more moderate in nature. The fact that individuals with moderate views would not endorse items that reflected extreme views was by design, where each individual was given the same opportunity to agree or disagree with a concept according to the magnitude of his or her own beliefs. Further, because these scales were modeled individually, participants were never forced to choose one position over another. Whether an individual chose to respond affirmatively to any item simply represented a manifestation of their own belief system (i.e., paradigm) in relation to how well those beliefs matched an item in terms of relative intensity.

Traditionalist beliefs were far less demanding than Interpretivist beliefs in terms of modeling scales during the course of this study. This is perhaps best exemplified by the ability of Traditionalist scales to capture a greater amount of variability in person measures relative to Interpretivist scales, with the exception of *Qualitative Methodology*, which performed comparably to all three Traditionalist scales (see Table 3.8). This was not entirely unexpected, given that Traditionalist views (e.g., positivism and then post-positivism) have dominated educational research until very recently. Traditionalist views are well-defined and deeply engrained in the minds of individuals, regardless of whether those views are part of the paradigm to which he or she adheres. A review of the literature on research paradigms demonstrates that Traditionalist perspectives are often used as an archetype from which alternative views are

distinguished. Interpretivist beliefs are, by definition, more individualized than the paradigms from which they sometimes try to disassociate (e.g., post-positivism). This is not to suggest that Traditionalist views are not also highly nuanced. Rather, this more than likely represents what Robert Stake (1993) had in mind when he said, "there are worlds within worlds, unending, each with its own paradigms" (*personal communication* as cited by Guba and Lincoln, 1994, p. 117). Certainly, this sentiment was never more accurate than when attempting to describe Interpretivist beliefs (e.g., constructivism).

As regards the labels that were used to identify scales and scale combinations during the course of this study, these labels were selected in consideration for how paradigms have developed and been discussed in recent history. The label "Traditionalist" was selected to characterize a certain combination of scales because the concepts measured by those scales represent components to a system that, until very recently, was the dominant paradigm in educational research. The label "Interpretivist" was selected to represent an alternative combination of scales because newer, alternative paradigms often focus on the researcher as being the primary instrument for data collection. Because Interpretivist paradigms have a long and distinguished history, it could be well-argued that their associated beliefs and practices are, in fact, traditional. As such, and for the purpose of assessing the results of this study, the reader is asked to review each of the RPI scales in terms of how they have been operationalized, and not necessarily by any label given to an individual scale or scale combination.

One of the more salient points to materialize during the course of this study was the instrument's apparent ability to catalyze self-reflection. This phenomenon first appeared during the think aloud exercises of the instrument development process, where all four of the individuals who participated commented that the instrument "really made them think" about themselves and their views on research. Interestingly, all four also stated that they might not otherwise think about many of the instrument's associated concepts during normal practice, perhaps echoing the belief that researchers "rarely have the time or inclination to assess what they do in philosophical terms" (Guba & Lincoln, 1994, p. 117) and proceed through an investigation without the necessity of long existential consideration (Patton, 2002). In the end, all four participants noted that they felt like the exercise was worthwhile, as it forced them to self-reflect and look at themselves critically in terms of what they were doing on a day to day basis. The administration of any instrument has some didactic effect, which may or may not be

desirous. In the future, if this effect represents researchers recognizing their own views, as well as the views of others for the purpose of understanding and engaging them in dialogue (Guba & Lincoln, 2005), then I am in agreement with the participants who considered the exercise meaningful.

Although this proposition has not been tested, it is possible that the RPI could be used to generate dialogue among research teach members regarding the similarities and differences in their paradigmatic assumptions. Given the increasingly interdisciplinary nature of research, an understanding of multiple perspectives might help to foster collaboration among individuals from various academic disciplines and research communities, wherein we are able to better understand the places in which our own paradigm is applicable, as well as the areas in which it might fall short. In such instances, garnering an understanding of and appreciation for alternative paradigms could help researchers to identify a more apposite approach to answering research questions, work with others toward building a body of evidence that effectively addresses those issues, and share our discoveries with a broader audience who might find them useful.

Some have endorsed the development of standards that apply to multiple approaches (Kamill, 1995; Munby, 2001). For instance, Howe (1992) suggested the adoption of a "critical educational research model," which consists of principles that are "broad and abstract enough" to encompass multiple perspectives such as the identification of questions grounded in education, applying practices that define a theoretical approach, grounding research in credible background knowledge and ensuring the overall warrant of conclusions and interpretations (p. 251). Any effort to develop quality criteria across paradigms or understand axioms associated with alternative approaches must be made in earnest, because those axioms are likely the reasons why a researcher chose the paradigm that he or she did, or why some might claim the "paradigm chose them". Further, any discussion about prospective criteria still has the potential to result in continued deliberation regarding the supremacy of one's own paradigm, where one is "shown to satisfy more or less the criteria that it dictates for itself and to fall short of a few of those dictated by its opponents" (Kuhn, 1962/1996, p. 110).

Many scholars claim the paradigm wars of the past are experiencing a cease-fire. The abundance of literature arguing different points of view suggests something to the contrary. When they originally presented their paradigm framework, Guba and Lincoln (1994) dutifully acknowledged that the views they presented were their own, asserting also that paradigms were

not open to proof in any conventional sense and must be accepted on faith (p. 107). Although we might concede that adherence to any particular paradigm remains an exercise in faith, the strong associations between philosophical views and methodological approaches that appeared in this study (see Tables 4.4 and 4.5) support the relevancy of Guba and Lincoln's framework (1994, 2005), at least in terms of describing some of the important components that serve to structure one's belief system regarding research. That stated, the majority of researchers are unlikely to "swear vows of allegiance to any single epistemological perspective" (Patton, 2002, p. 136) and will claim their selection of a methodological approach is based predominantly on the research questions they are asking. As evidence to this, 94% of those who participated in the field test portion of this study were in agreement with the statement: *The research methods I use depend on the research questions I ask.* This may be true, but as described in Chapter Two of this document, philosophical beliefs may still have an impact on deciding which research questions are important and how those questions are structured. This issue is itself one of the many questions that will be addressed through future research using the scales developed during the course of this study, in conjunction with other methodological approaches (e.g., qualitative analyses).

On the topic of methods, recent articles have promulgated different opinions about methodological eclecticism (Ercikan & Roth, 2006), endorsed a softer version of the incompatibility thesis (Yanchar & Williams, 2006), promoted a reconnection with pragmatism in lieu a metaphysical account of paradigms (Morgan, 2007) and endorsed the use of mixed research methods (Johnson & Onwuegbuzie, 2004). All of these authors should be congratulated for their efforts and willingness to discuss their views on these controversial topics. I would especially like to applaud proponents of the concept of methodological eclecticism (Ercikan & Roth, 2006), because I believe that many of today's researchers tend to operate in different methodological paradigms all the time. However, where Ercikan and Roth (2006) may have been slightly off the mark is in asking researchers to consider methods as existing across a continuum. With all due respect to Ercikan and Roth (2006), this conceptualization necessarily requires endorsing one view at the expense of another (i.e., as you move closer to one end of a continuum you move further away from the other). This may seem pedantic, but because all phenomena are both quantitative and qualitative at the same time (Ercikan & Roth, 2006), different methodological approaches are probably more accurately representative of a family of distinct

ingredients, based in different language, from which researchers select what they believe to be the most appropriate combination for studying and making reasonable assertions about a particular topic. For some researchers, that combination reflects his or her ontological and epistemological belief system, while others simply select a workable combination based on a more pragmatic approach.

On the matter of pragmatism in educational research, several arguments promoting the acceptance of mixed methods (i.e., combining quantitative and qualitative approaches) have emerged in recent years, perhaps signaling that we are in the midst of a paradigm shift. Even though debates about educational research paradigms were expanded during the closing decades of the 20th century to include dialogue related to specific philosophical propositions, much of today's discussion still focuses on methods and the appropriateness of their application. It may be that Morgan's (2007) appeal is coming to fruition, where pragmatism might eventually serve as the guiding paradigm in social science research, "both as a basis for supporting work that combines qualitative and quantitative methods and as a way to redirect attention to methodological rather than metaphysical concerns" (p. 48). Johnson, Onwuegbuzie and Turner (2007) apparently endorsed this view by suggesting we recognize mixed methods as a third, major "research paradigm" in addition to qualitative and quantitative approaches. Perhaps in response to those who support a metaphysical account of paradigms, Johnson et al. (2007) claimed that philosophical commitments are welcomed and embraced as part of a mixed methods research paradigm. At the same time, Johnson et al. (2007) endorsed continued discussions about various philosophical propositions so that the "field is self-reflexive and continues to grow (philosophically)" (p. 126). It is difficult for me to formulate any compelling argument against the use of mixed research methods, especially because the development of the RPI entailed the use of strategies that I consider both qualitative and quantitative in nature. I believe that the choices I have made and will continue to make in that regard reflect my own commitment to exploring and thinking seriously about many of the approaches that exist in contemporary educational research.

References

Airasian, P. W. & Walsh, M. E. (1997). Constructivist cautions. *Phi Delta Kappan, 78*(6), 444-449.

Andrich D. (1978). A rating scale formulation for ordered response categories. *Psychometrika, 43*, 561-573.

Bahm, A. J. (1974). *Metaphysics: An introduction.* New York: Barnes and Noble.

Berg, B. L. (2001). *Qualitative research methods for the social sciences.* Boston, MA: Allyn & Bacon.

Biesta, G. J. J., & Burbles, N. C. (2003). *Pragmatism and educational research*. Lanham, MD: Rowman and Littlefield.

Bond, T. G. & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Burrell, G. & Morgan, G. (1979). *Sociological paradigms and organizational analysis.* Brookfield, VT: Ashgate Publishing.

Cassidy, D. (2006). *Quantum mechanics (1925-1927): The Uncertainty Principle*. American Institute of Physics. Retrieved April 27, 2006 from http://www.aip.org/history/heisenberg/p08.htm.

Charney, D. (1996). Empiricism is not a four-letter word. *College Composition and Communication, 47*(4), 567-593.

Chen, F., Curran, P. J., Bollen, K. A., Kirby J. & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in Structural Equation Models. *Sociological Methods & Research, 36*(4), 462-494.

Comte, A. (1988). *Introduction to positive philosophy.* Indianapolis, IN: Hackett Publishing. (Originally published in 1830).

Creamer, E. G. (2003). Exploring the link between inquiry paradigms and the process of collaboration. *The Review of Higher Education, 20*(4), 447-465.

Creswell, J. W. (1998). *Qualitative inquiry and research design: Choosing among five traditions.* Thousand Oaks: Sage.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Denzin, N. K. (1994). The art and politics of interpretation. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 500-515). Thousand Oaks, CA: Sage.

Denzin, N. K. & Lincoln, Y. S. (2000). Paradigms and perspectives in transition. In N. K. Denzin and Y. S. Lincoln (Eds.), *Handbook of qualitative research* (2nd ed., pp. 155-162). Thousand Oaks, CA: Sage Publications.

Denzin, N. K. & Lincoln, Y. S. (2005). *The Sage handbook of qualitative research* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Descartes, H. (1968). Discourse on method and the meditations. *Harmondsworth: Penguin.*(Original work published 1637).

Dewey, J. (1981). The influence of Darwinism on philosophy. In J. J. McDermott (Ed.), *The philosophy of John Dewey.* Chicago: University of Chicago Press.

Draba, R. E. (1977). *The identification and interpretation of item bias* (Research Memorandum No. 26). Chicago, IL: University of Chicago.

Education Sciences Reform Act of 2002 (Pub. L. No. 107-279). Retrieved August 30, 2006 from http://www.ed.gov/policy/rschstat/leg

Eisenhart, M. & Towne, L. (2003). Contestation and change in national policy on "scientifically based" educational research. *Educational Researcher, 32*(7), 31-37.

Ercikan, K. & Roth, W. M. (2006). What good is polarizing research into qualitative and quantitative? *Educational Researcher, 35*(5), 14-23.

Erwin, T. A. (1981). *Scale of Intellectual Development* [Manual]. Harrisonburg, VA: Developmental Analytics.

Erwin, T. A. (1983). The Scale of Intellectual Development: Measuring Perry's scheme. *Journal of College Student Personnel*, *24*, 6-12.

Feldman M. (1998). Re: Rethinking quantitative social research. *Statistical Consulting Newsgroup.* [Online]. Available newsgroup: sci.stat.consult.

Filstead, W. J. (1979). Qualitative methods: A needed perspective in evaluation research: In T. D. Cook & C. S. Reichardt (Eds.), *Qualitative and quantitative methods in evaluation research* (pp. 33-48). Beverly Hills, CA: Sage.

Fisher, W. P., Jr. (1994). The Rasch debate: Validity and revolution in educational measurement. In M. Wilson (Ed.), *Objective measurement: Theory and practice* (Vol. 2, pp. 36-72). Norwood, NJ: Ablex.

Fitzgerald, J. & Cunningham, J. W. (2002). Mapping basic issues for identifying epistemological outlooks. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 209-228). Mahwah, NJ: Lawrence Erlbaum.

Fitzpatrick, J. L., Sanders, J. R. & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines.* Boston, MA: Pearson.

Gadamer, H. G. (1989). *Truth and method,* (2nd rev. ed.) (J. Weisheimer & D. G. Marshall, Eds. & Trans.). New York, NY: Crossroad.

Gage, N. L. (1989). The paradigm wars and their aftermath: A historical sketch of research on teaching since 1989. *Teachers College Record, 91*(2), 135-149.

Geertz, C. (1988). *Works and lives: The anthropologist as author.* Cambridge: Polity. Geertz, C. (1993). *Local knowledge: Further essays in interpretive anthropology*. London: Fontana.

Gergen, K. J. (2001). Psychological science in a postmodern context. *American Psychologist, 56,* 803-813.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8*(4), 597-606. Retrieved June 13, 2007 from http://www.nova.edu/ssss/QR/QR8-4/golafshani.pdf

Green, J. (1992). Multiple perspectives: Issues and directions. In R. Beach, J. Green, M. L. Kamil & T. Shanahan (Eds.), *Multidisciplinary perspectives on literary research* (pp. 19-34). Urbana, IL: National Conference on Research in English

Green, R. G., Bretzin, A., Leininger, C., & Stauffer, R. (2001). Research learning attributes of graduate students in social work, psychology and business. *Journal of Social Work Education, 37*(2), 333-341.

Guba, E. G. (1990). The alternative paradigm dialogue. In E. G. Guba (Ed.), *Paradigm dialogue* (pp. 17-27). Newbury Park, CA: Sage Publications.

Guba, E. G. & Lincoln, Y. S. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage.

Guba, E. G. & Lincoln, Y. S. (2005). Paradigmatic controversies, contradictions and emerging confluences. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research, 3rd ed.* (pp. 191-216). Thousand Oaks, CA: Sage.

Hambleton, R. K. (1993). Principles and selected applications of Item Response Theory. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 147-200). New York, NY: Macmillan Publishing.

Hamilton, D. (1994). Traditions, preferences and postures in applied qualitative research. In N. K. Denzin and Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 60-69). Thousand Oaks, CA: Sage.

Hansen, J. T. (2004). Thoughts on knowing: Epistemic implications of counseling practice. *Journal of Counseling and Development, 82,* 131-138.

Heisenberg, W. (1991). The Copenhagen interpretation of quantum theory. In T. Ferris (Ed.), *The world treasury of physics, astronomy and mathematics (pp. 86-97).* Boston, MA: Little, Brown and Company.

Heron, J. & Reason, P. (1997). A participatory inquiry paradigm. *Qualitative Inquiry, 3*, 274-294.

Heylighen, F. (1997). Epistemological constructivism. In F. Heylighen, C. Joslyn and V. Turchin (Eds.): *Principia cybernetica web.* Retrieved May 22, 2006 from http://pespmc1.vub.ac.be/CONSTRUC.html

Horkheimer, M. (1972). *Critical theory.* New York, NY: Seabury.

Hostetler, K. (2005). What is "good" education research? *Educational Researcher, 34*(6), 16-21.

Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis (or dogmas die hard). *Educational Researcher, 17*(8), 10-16.

Howe, K. R. (1992). Getting over the quantitative-qualitative debate. *American Journal of Education, 100*(2), 236-257.

Howe, K. R. (1998). The Interpretive turn and the new debate in education. *Educational Researcher, 27*(8), 13-20

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling* (pp. 158-176). Thousand Oaks, CA: Sage.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Jackson, L. A. & Jeffers, D. L. (1989). The Attitudes About Reality Scale: A new measure of personal epistemology. *Journal of Personality Assessment, 53*(2). 353-365.

Johnson, J. A., Germer, C. K., Overton, J. A. & Overton, W. F. (1988). Personality as the basis for theoretical predilections. *Journal of Personality and Social Psychology, 55*(5). 824-835.

Johnson, R. B. & Onwuegbuzie, A. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.

Johnson, R. B., Onwuegbuzie, A. & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112-133.

Jöreskog, K. G. & Sörbom, D. (2004). *LISREL 8.7* [Computer Software]. Lincolnwood, IL: Scientific Software International, Inc.

Kamill, M. L. (1995). Some alternatives to paradigm wars in literary research. *Journal of Reading Behavior, 27*(2). 243-260.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Westport, CT: Praeger Publishers.

Kant, I. (1996). In W. S. Pluhar (Trans.), *Critique of pure reason*. Indianapolis, IN: Hackett Publishing. (Original work published 1781).

Karabatsos, G. (2000). A critique of Rasch residual fit statistics. *Journal of Applied Measurement, 1*(2), 152-176.

Kim, S. (2003). Research paradigms in organizational learning and performance: Competing modes of inquiry. *Information Technology, Learning and Performance, 21*(1), 9-18.

Kincheloe, J. L. & McLaren, P. (2005). Rethinking critical theory and qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (3rd ed., pp. 303-342). Thousand Oaks, CA: Sage Publications.

Kincheloe, J. L. & Steinberg, S. R. (1997). *Changing multiculturalism: New times, new curriculum.* London: Open University Press.

Kuhn, T. (1962/1996). *The structure of scientific revolutions* (3rd ed.). Chicago, IL: University of Chicago Press.

Lather, P. & Moss, P. A. (2005). Introduction: Implications of the scientific research in education report for qualitative inquiry. *Teachers College Record, 107*(1). 1-3.

Laudan, L. (1977). *Progress and its problems.* Berkeley, CA: University of California.

Linacre, J. M. (1992). Prioritizing misfit indicators. *Rasch Measurement Transactions, 9*, 422-423.

Linacre, J. M. (1997) KR-20 / Cronbach alpha or Rasch reliability: Which tells the "truth"? *Rasch Measurement Transactions, 11*(3). 580-581

Linacre, J. M. (1998). Detecting multidimensionality: Which residual type works best? *Journal of Outcome Measurement, 2*(3), 266-283.

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1). 85-106.

Linacre, J. M. (2006). *A user's guide to Winsteps: Rasch model computer programs.* Chicago, IL: Retrieved August 1, 2007 from www.winsteps.com.

Locke, J. (1956). *An essay concerning human understanding.* (P. H. Nidditch, Ed.). Oxford: Clarendon Press. (Original work published 1689).

Luo, G. (2007). The relationship between the Rating Scale and Partial Credit Models and the implication of disordered thresholds of the Rasch models for polytomous responses. In

Smith, E. V. & Smith, R. M. (Eds.), *Rasch measurement: Advanced specialized applications* (pp. 181-201). Maple Grove, MN: Journal of Applied Measurement Press.

Masterman, M. (1970). The nature of a paradigm. In I. Lakatos & A. Musgraves (Eds.), *Criticism and the growth of knowledge* (pp. 59-89). Cambridge: Cambridge University Press.

Masters, G. N. (1982) A Rasch model for partial credit scoring. *Psychometrika*, *47,* 149-174.

Maxwell, S. E. & Delaney, H. D. (2004). *Designing experiments and analyzing data.* Mahwah, NJ: Lawrence Erlbaum and Associates.

Merriam, S. B. (1998). *Qualitative research and case study: Applications in education.* San Francisco, CA: Jossey-Bass.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: Macmillan Publishing.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*. 741-749.

Mill, J. S. (1906). *A system of logic.* London: Longman. (Original work published 1843).

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications for combining quantitative and qualitative methods. *Journal of Mixed Methods Research, 1*(1), 48-76.

Morris, B. J. (1999). *The poverty of constructivism: A comparison of philosophies of inquiry in the social sciences.* Pittsburgh, PA: Carnegie Mellon University. (ERIC Document Reproduction Service No. ED440903)

Mulaik, S. A. & Lawrence, J. R. (1995). Objectivity and reasoning in science and structural equations modeling. In R. H. Hoyle (Ed.). *Structural equations modeling: Concepts, issues and applications* (118-137). Thousand Oaks: Sage.

Munby, H. (2001, March). *Educational research as disciplined inquiry: Examining the facets of rigor in our work.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, St. Louis, MO.

Nisbet, J. (2005). What is educational research? Changing perspectives through the 20th century. *Research Papers in Education, 20*(1). 25-44.

No Child Left Behind Act of 2001, Pub. L. No. 107-110.

Nosek, B. A., Mahzarin, R. B. & Greenwald, A. G. (2002). Me = Female, Math = Male, therefore Math ≠ Me. *Journal of Personality and Social Psychology, 83*(1). 44-59.

Nott, M. & Wellington, J. (1993). Your nature of science profile: An activity for science teachers. *School Science Review, 75*(270). 109-112.

Papanastasiou, E. C. (2005). Factor structure of the "Attitudes toward Research" Scale. *Education Research Journal, 4*(1), 16-26.

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

Paul, J. L. & Marfo, K. (2001). Preparation of educational researchers in philosophical foundations of inquiry. *Review of Educational Research, 71*(4). 525-547.

Phillips, D. C. (1987). *Philosophy, science and social inquiry.* New York, NY: Pergamon Press.

Pollack, J. L. (1986). *Contemporary theories of knowledge*. Totowa, NJ: Rowan and Littlefield.

Ponteretto, J. G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology, 52*(2).126-136.

Popper, K. R. (1957). *The poverty of historicism*. London: Routledge.

Popper, K. R. (1959). *The logic of scientific discovery*. London: Routledge.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. London: Routledge.

Popper, K. R. (1972). *Objective knowledge*. Oxford: Clarenden Press.

Rabinow, P. & Sullivan, W. (1979). The interpretive turn: Emergence of an approach. In P. Rabinow & W. Sullivan (Eds.), *Interpretive social science* (pp. 1-21). Los Angeles, CA: University of California Press.

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika, 53*, 495–502.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement, 14*, 197-207.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen: Danish Institute for Educational Research.

Royce, J. R. & Mos, L. (1980). *Psycho-Epistemological Profile* [Manual]. Edmonton, Alberta, Canada: University of Alberta Printing Office.

Schaeffer, D. W. & Serlin, R. C. (2004). What good are statistics that don't generalize? *Educational Researcher, 33*(9), 3-13.

Schnelker, D. L. (2006). The student-as-bricoleur: Making sense of research paradigms. *Teaching and Teacher Education, 22*. p. 42-57.

Schommer, M. (1990). The effects of beliefs about the nature of knowledge in comprehension. *Journal of Educational Psychology, 82*, 498-504.

Schraw, G., Bendixen, L. D. & Dunkle, M. E. (2002). Development and validation of the Epistemic Beliefs Inventory (EBI). In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 261-275). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Schwandt, T. (1989). Solutions to the paradigmatic controversy: Coping with uncertainty. *Journal of Contemporary Ethnography, 17*(4), 379-407).

Sechrest, L. & Sidana, S. (1995). Quantitative and qualitative methods: Is there an alternative? *Evaluation and Program Planning, 18*, 77-87.

Smith, A. B., Rush, R., Fallowfield, L. J. Velikova, G. & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology, 8*(33). 1-11.

Smith, B. & Thomas, A. (1998). Axiology. In *Routledge encyclopedia of philosophy*, Version 1.0, London: Routledge.

Smith, B. (2003). Ontology. In L. Floridi (Ed.), *Blackwell guide to philosophy of computing and information* (pp. 155-166). Oxford: Blackwell.

Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using fit statistics and principal components analysis. *Journal of Applied Measurement, 3*(2), 205-231.

Smith, J. K. (1983). Quantitative versus interpretive: The problem of conducting social inquiry. In E. R. House (Ed.) , *Philosophy of evaluation* (pp. 27-51). San Francisco, Jossey-Bass.

Smith, R. M., Linacre, J. M., & Smith, Jr., E.V. (2003). Guidelines for manuscripts. *Journal of Applied Measurement, 4*, 198-204.

Smith, R. M and Miao, C. Y. (1994). Assessing unidimensionality for Rasch measurement. In M. Wilson (Ed.), *Objective Measurement: Theory into practice*. Greenwich: Ablex.

Smith, R. M., Schumacker, R. E. & Bush, M. J. (1998). Using item mean squares to evaluate fit to the Rasch model. *Journal of Outcome Measurement, 2*(1), 66-78.

Steiger, J. H., Shapiro, A. Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50*, 253-264.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Tejada, A J. R., Gomez, A. G. & Melendez, C. P. (2002). Two strategies for fitting real data to Rasch polytomous models. *Journal of Applied Measurement, 3*(2), 129-145.

Toma, J. D. (1996, April) *Scholars and their inquiry paradigms: Exploring a conceptual framework for classifying inquiry and inquirers based upon paradigmatic assumptions.* Paper presented at the annual meeting of the American Educational Research Association, New York, NY.

Tsai, C. & Liu, S. (2005). Developing a multi-dimensional instrument for assessing students' epistemological view toward science. *International Journal of Science Education, 27*(13). 1621-1638.

Tsoi-Hoshman, L. S. (1989). Alternate research paradigms: A review and teaching proposal. *The Counseling Psychologist, 17*(1), 3-79.

Unger, R. K., Draper, R. D. & Pendergrass, M. L. (1986). Personal epistemology and personal experience. *Journal of Social Issues, 42*(2). 67-79.

Wolfe, E.W., & Smith, E.V., Jr. (2007a). Instrument development tools and activities for measure validation using Rasch Models: Part I - Instrument development tools. In Smith, E. V. & Smith, R. M. (Eds.), *Rasch Measurement: Advanced specialized applications* (pp. 202-242). Maple Grove, MN: Journal of Applied Measurement Press.

Wolfe, E.W., & Smith, E.V., Jr. (2007b). Instrument development tools and activities for measure validation using Rasch Models: Part II – Validation activities. In Smith, E. V. & Smith, R. M. (Eds.), *Rasch Measurement: Advanced specialized applications* (pp. 243-290). Maple Grove, MN: Journal of Applied Measurement Press.

Wright, B. D. & Masters, G. N. (1982). *Rating scale analyses*. Chicago: MES Press.

Yanchar, S. C. & Williams, D. D. (2006). Reconsidering the compatibility thesis and eclecticism: Five proposed guidelines for method use. *Educational Researcher, 35*(9), 3-12.

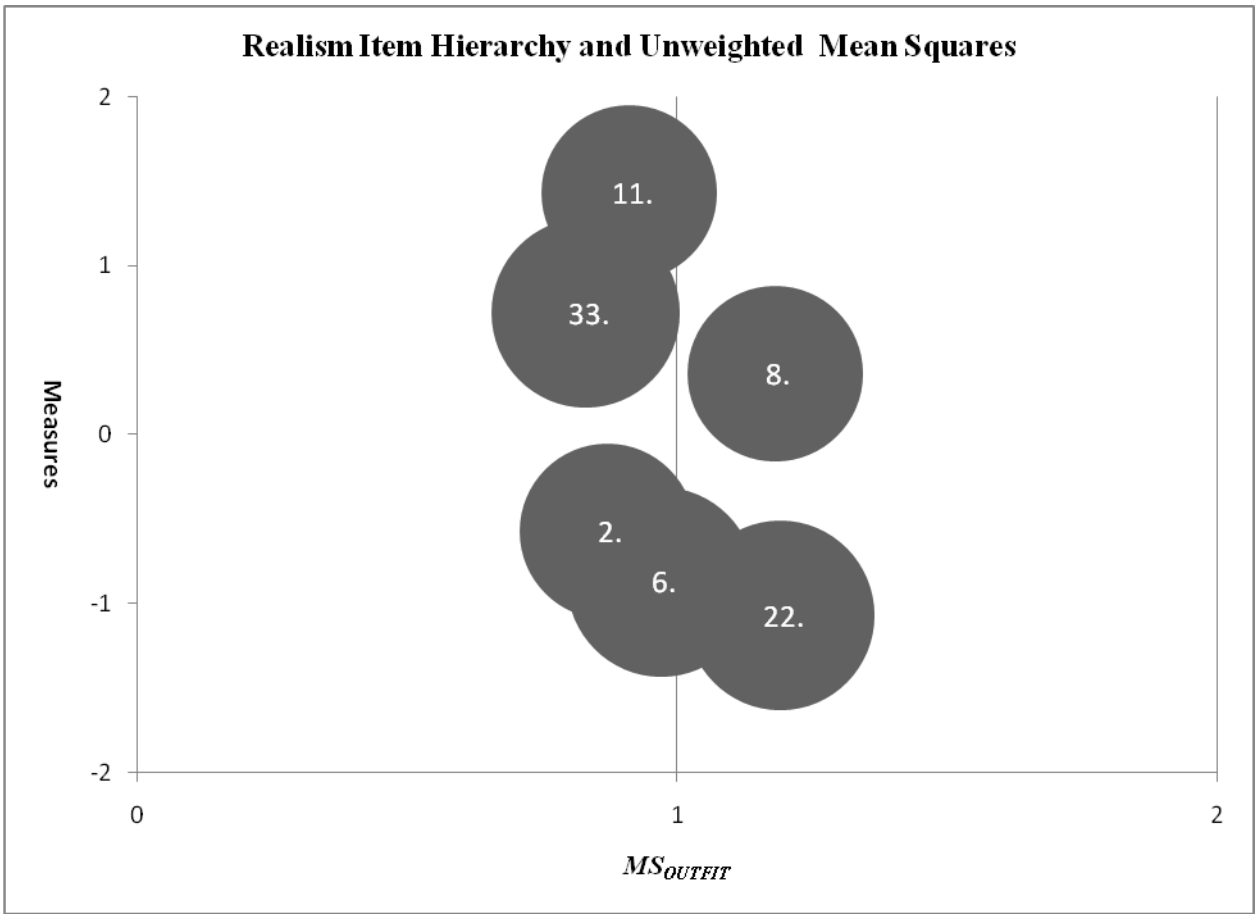
Yu, C. H. (2001, April). *Misconceived relationships between logical positivism and quantitative research.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Appendix A. RPI Scales and Item Difficulty Estimates for the Final 38 Item Instrument

1. Realism in Research Subscale

Field Test Item Number	*Item*	*Measure*	SE_δ
11.	Although there may be many opinions, there is only one real truth.	1.70	.14
33.	The truth is out there waiting to be discovered by researchers.	.52	.14
8.	Reality is a stable entity – it is only people's perception of reality that varies.	.33	.14
2.	Quality research uncovers the truth.	- .63	.14
6.	A primary objective of research is to reveal reality.*	-.96	.14
22.	Research helps to explain the world as it actually is.	- 1.02	.14

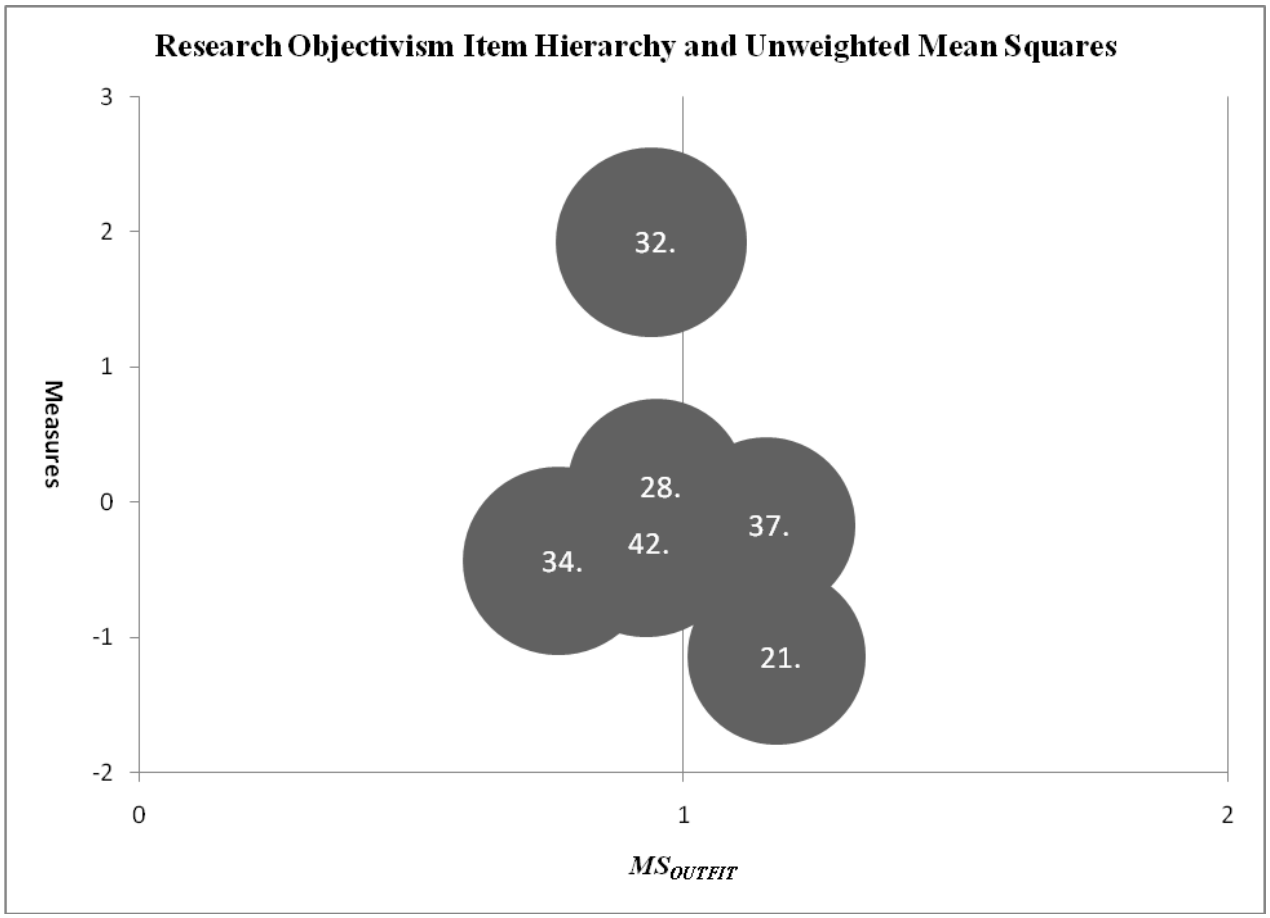
* Adapted from Nott and Wellington (1993)



2. Research Objectivism Subscale

Field Test Item Number	*Item*	*Measure*	SE_δ
32.	The research methods I use are value-free.	1.93	.16
28.	Once a topic is selected, the personal values of a researcher should be separated from the research process.	.12	.15
37.	Researchers are capable of making unbiased observations during the course of a study.*	-.17	.15
42.	Objectivity is a standard against which research should be compared.	-.30	.15
34.	Objectivism is a principle to which researchers should adhere.	-.43	.15
21.	In order to reach accurate conclusions, researchers should remain impartial during the course of a study.	-1.14	.15

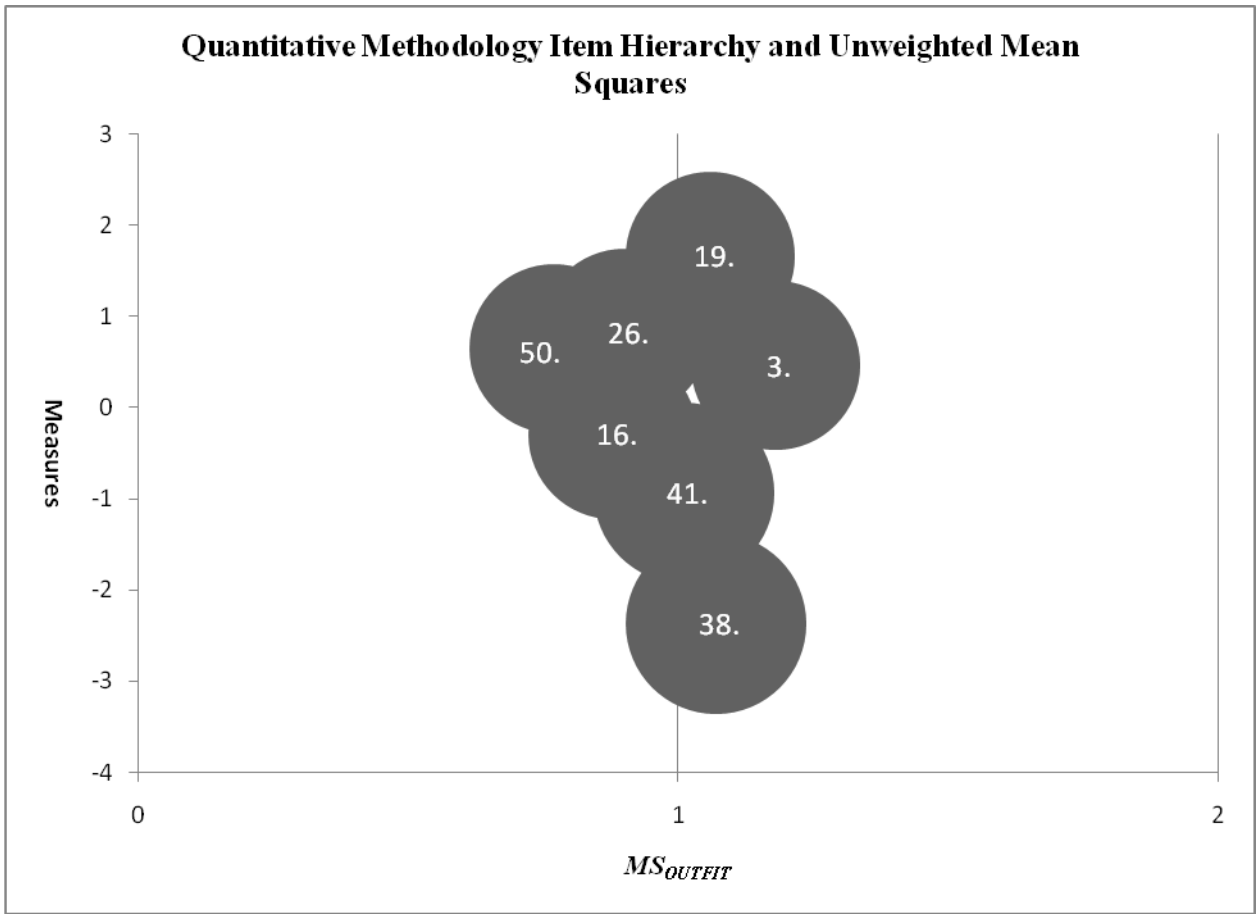
*Adapted from Tsai and Liu (2005)



3. Quantitative Methodology Subscale

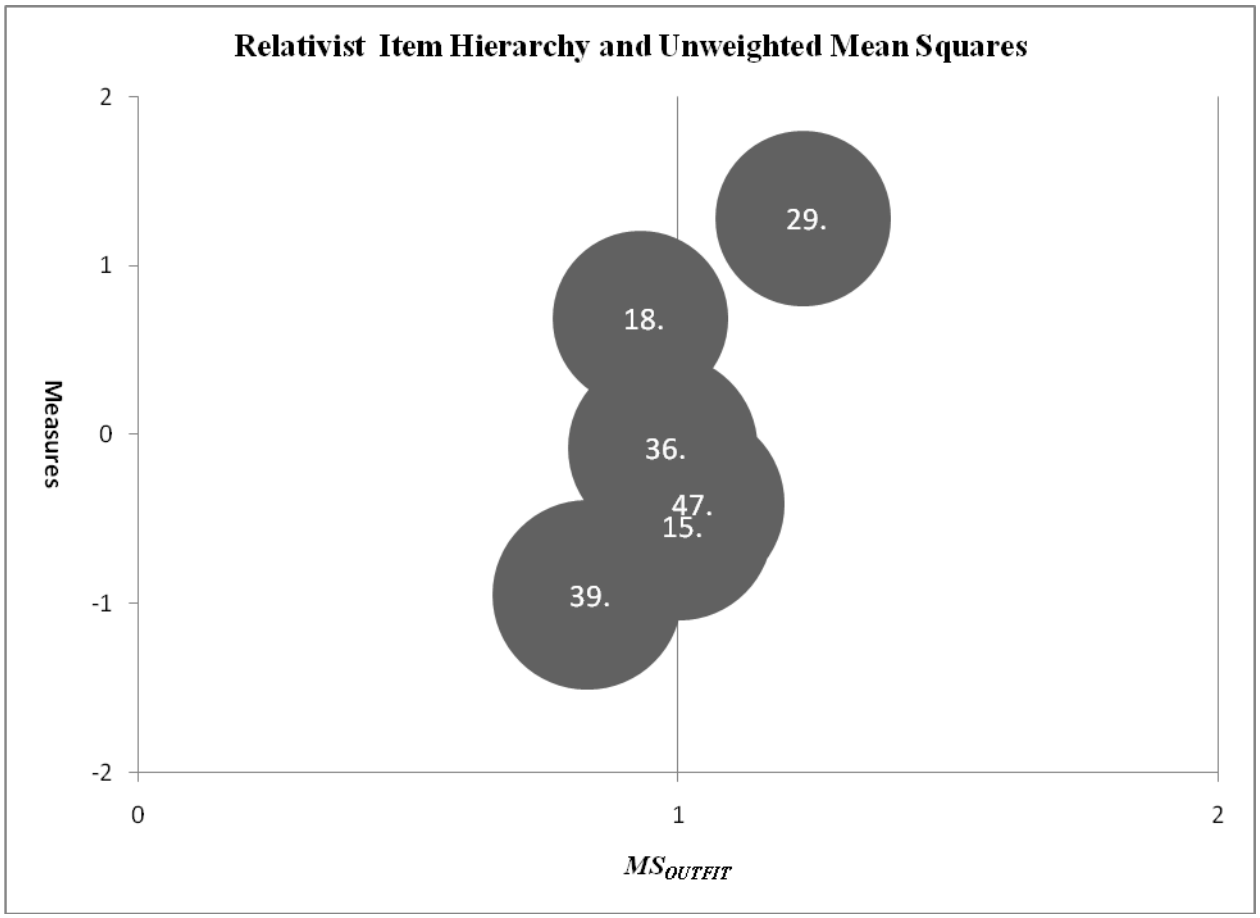
Field Test Item Number	*Item*	*Measure*	SE_δ
19.	The strength of theory is best determined by using data-based quantitative methods.*	1.38	.14
26.	The most productive way for me to explore research topics is to formulate a hypothesis and then test it using empirical data.*	.75	.14
50.	The most effective research designs feature systems for controlling variables.	.57	.14
3.	Large, randomly selected samples provide useful data for the research studies I am likely to conduct.	.44	.14
16.	The generalizability of conclusions is an important component to my research.	-.20	.14
41.	Replicable findings lead to valid research conclusions.	-.77	.15
38.	Carefully constructed instruments are effective tools for collecting quality data for my research.	-2.17	.15

* Adapted from Chang (1996)



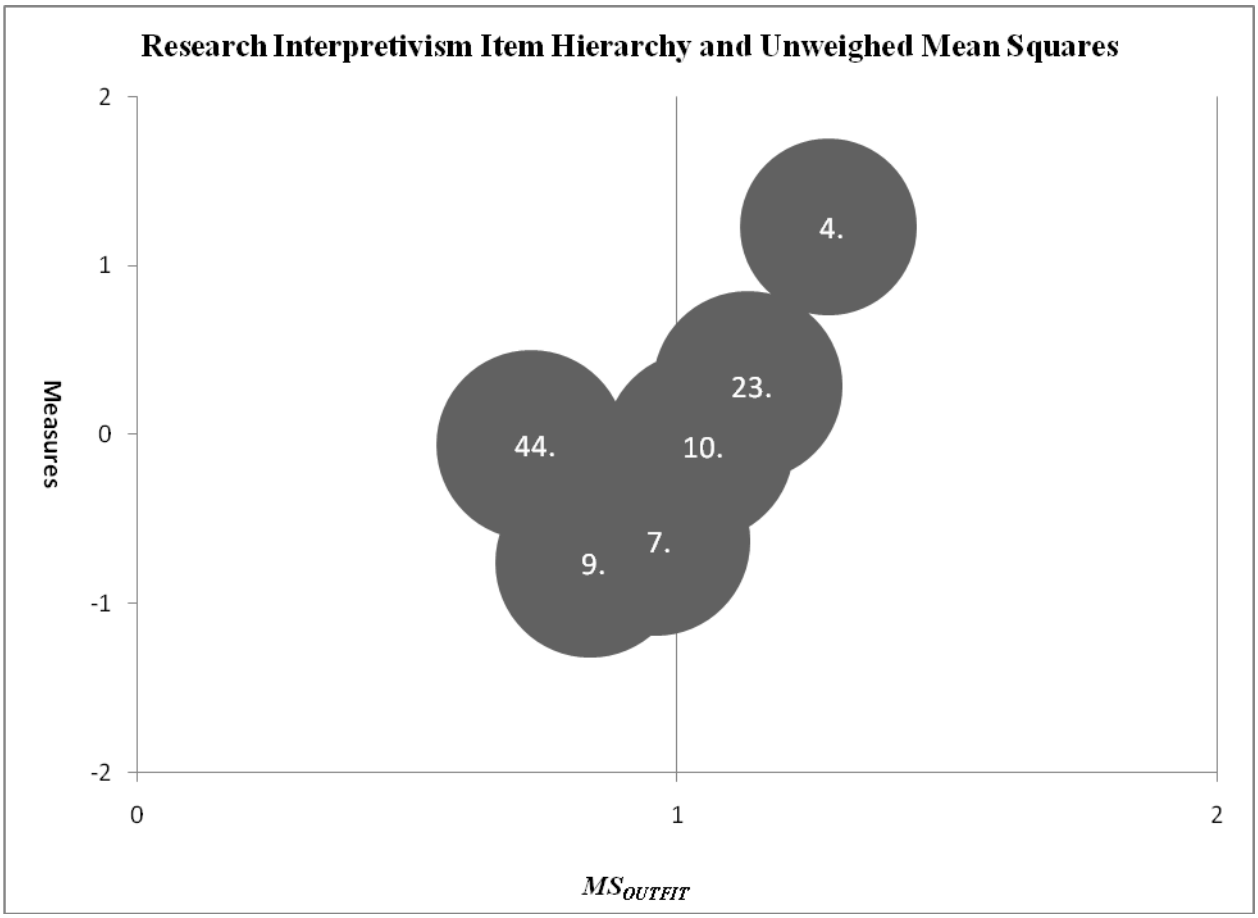
4. Relativism in Research Subscale

Field Test Item Number	*Item*	*Measure*	SE_δ
29.	Reality is determined by who holds power in society.	1.28	.13
18.	Reality is an entity that is constructed through interactions between researchers and research participants.	.69	.14
36.	Truth is determined by the cultural, economic and/or political climate of the time.	-.08	.14
47.	One of my objectives for research is to understand the individual realties that exist in the world.	-.41	.14
15.	Reality is formed in part through individual cognition.	-.54	.14
39.	Reality exists in different forms – each mediated by an individual's own experiences.	-.95	.15



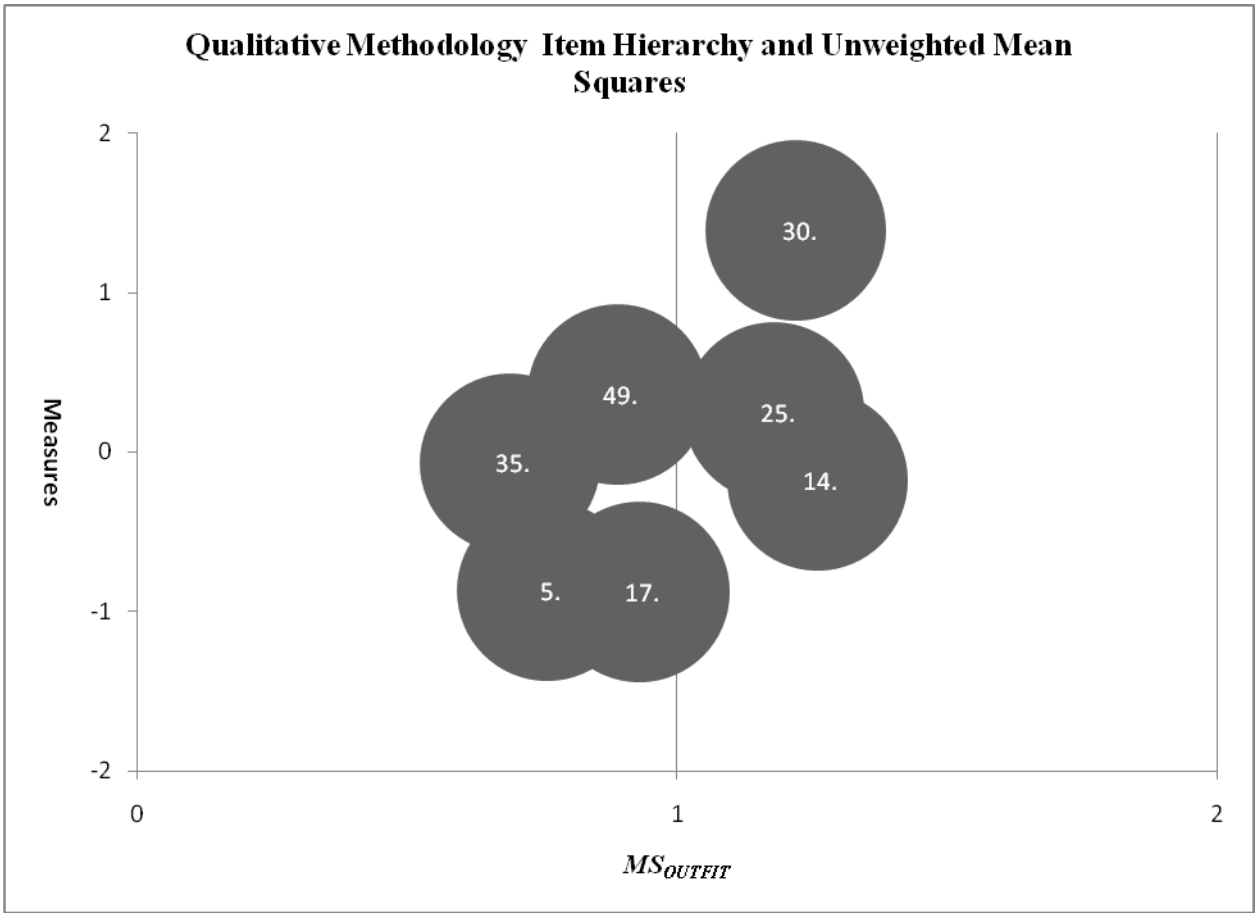
5. Research Interpretivism Subscale

Field Test Item Number	*Item*	*Measure*	SE_δ
4.	Whenever we study something we change it.	1.23	.15
23.	Researchers create knowledge by attaching their own meaning to things.	.29	.15
44.	Research is a subjective process.	-.06	.15
10.	A researcher's intuition plays a decisive role in the process of data analysis.	-.07	.15
7.	The data a researcher collects is inherently laden with his or her own values.	-.63	.15
9.	Researchers construct knowledge by making personal interpretations of the data he or she collects.	-.76	.15



6. Qualitative Methodology Subscale

Field Test Item Number	*Item*	*Measure*	SE_δ
30.	Quantitative methods are often inappropriate because they silence the voice of an individual.	1.39	.12
49.	Smaller, purposefully collected samples tend to provide the most meaningful data for my research.	.36	.12
25.	The steps of a research design should be allowed to emerge as a study progresses.	.25	.12
35.	Exploring research topics in their natural environment is more valuable to me than the statistical testing of hypotheses.	-.07	.12
14.	The research topics I am likely to explore require the use of qualitative methods.	-.18	.12
5.	Developing a rapport with research participants is optimal for collecting quality data.	-.87	.13
17.	I need to have a strong background in qualitative methods.	-.90	.13



Appendix B. Instrument Blueprint (Original)

	Instrument Content Dimensions				Total
	Ontology	***Epistemology***	***Axiology***	***Methodology***	
Traditionalist Paradigms	**Absolute Realism** (Minimum of Four Items)	**Objectivism** (Minimum of Four Items)	**Ethics Extrinsic** (Minimum of Four Items)	**Focus on hypothesis testing/chiefly quantitative methods** (Minimum of Four Items)	**Minimum of 16 Items**
Interpretivist Paradigms	**Absolute Relativism** (Minimum of Four Items)	**Subjectivism/Interpretivism** (Minimum of Four Items)	**Ethics Intrinsic** (Minimum of Four Items)	**Focus on emerging data/chiefly qualitative methods** (Minimum of Four Items)	**Minimum of 16 Items**
Total	**Minimum of 8 Items**	**Minimum of 8 Items**	**Minimum of 8 Items**	**Minimum of 8 Items**	**32 Items**

Appendix C. Item Content Linking Instrument

Introduction: The following exercise is designed to link a draft set of items to a specific content domain within an instrument blueprint. This instrument is being created to measure the views of educational researchers across the philosophical dimensions that comprise research paradigms (i.e. ontology, epistemology, axiology and methodology). Part I of this exercise asks you to indicate the domain from Table 1 that you believe the content of individual items represents. Part II asks you to provide any additional statements you feel would be necessary for measuring the domains contained in Table 1, but were not included in the list of items you were asked to review. Before you begin, please detach the first page of this document, which contains Table 1, to use as a reference as you provide responses.

Table 1: Instrument Blueprint Content Domains

Ontology	Epistemology	Axiology	Methodology
What is the form and nature of reality?	*What is the nature of the relationship between the knower and what can be known?*	*What role do an individual researcher's values play in the research process?*	*How can the researcher go about finding out whatever can be known?*
Domain 1: ***Absolute Realism*** There is a single, independent reality that exists outside of oneself.	**Domain 3:** **Objectivism** The researcher and research object are assumed to be independent entities, wherein the researcher is able to study an object without influencing it or being influenced by it.	**Domain 5:** **Ethics Extrinsic** The researcher's individual values and ethics are purposively excluded from the research process and thought to be non-influential.	**Domain 7:** **Focus on hypothesis testing** The researcher explores hypotheses *a priori*, often through the use of quantitative methods.
Domain 2: **Absolute Relativism** Multiple realities are local, socially constructed entities and always relative to some particular frame of reference	**Domain 4:** **Subjectivism/Interpretivism** The researcher and research object are assumed to be so interactively linked that "findings" are literally created as an investigation proceeds.	**Domain 6:** **Ethics Intrinsic** The researcher's individual values and ethics are accepted, and sometimes embraced, as part of the research process.	**Domain 8:** **Focus on emerging data** The researcher develops hypotheses *a posteriori*, often through the use of qualitative methods.

***Definitions have been adapted from Guba & Lincoln (1994)**

Name: _____

Part I:
Directions: Please indicate the content domain that each of the following philosophical statements represents by writing the corresponding number from Table 1 in the spaces provided.

2	⬅	**Example:** If you believe a statement reflects an ***absolute relativist*** view within the ***ontology*** dimension (i.e. Domain 2), you would respond by writing **2** in the space provided.

	Content Domain	Philosophical Statement
1.	_____	Research does not describe what is real – it describes people's perceptions of what is real.
2.	_____	Researchers are responsible for making sure participants benefit from their research.
3.	_____	Although complete objectivism in research is difficult to achieve, it is the standard to which all researchers should aspire.
4.	_____	In order to reach accurate conclusions, researchers must make dispassionate observations and decisions during the course of a study.
5.	_____	Theories derived from research are as much a result of imagination and intuition as inferences from data.
6.	_____	Researchers are not the ones responsible for participants benefiting from their research.
7.	_____	Developing a relationship with research participants is optimal for collecting quality data.
8.	_____	Researchers are not necessarily responsible for how their research will be used.
9.	_____	Research conclusions are a result of how the words spoken by research participants are interpreted.
10.	_____	The purpose of research is to establish equality where none existed before.
11.	_____	The more emotionally attached you become to something, the less likely you are to view it in its real form.
12.	_____	Research exists to correct society's wrongs.
13.	_____	The truth is out there and waiting to be discovered by researchers.
14.	_____	The truth is dependent upon with whom a researcher is speaking.
15.	_____	A researcher's emotions play no part in the creation of true knowledge.
16.	_____	Although an inquirer's personal views might impact the selection of a research question, those views have no place in the actual research process.
17.	_____	In practice, choices between competing theories are made predominantly on the basis of experimental results.
18.	_____	Truth in research does exist outside of personal interpretation.
19.	_____	Nothing can be studied without changing it.
20.	_____	Research proceeds by drawing generalizable conclusions (which later become theories) from empirically collected data.
21.	_____	The subjective nature of the human mind inherently limits the possibility for complete objectivism during research.
22.	_____	The truth is still the truth regardless of who you are.
23.	_____	Research conclusions are implicit in data rather than a personal choice made by an inquirer.

24. _____ If a researcher is rigorous, their conclusions will likely be accurate.
25. _____ Having direct personal contact with research participants is optimal for reaching authentic research conclusions.
26. _____ Reality is something that is literally constructed through interactions between researchers and research subjects.
27. _____ Reality is determined by who holds power in society.
28. _____ Researchers construct knowledge by considering their own views in relation to others'.
29. _____ The generation of knowledge is dependent upon the imagination and creativity of researchers.
30. _____ Research reports should be presented using a neutral 'voice'.
31. _____ The best way to reach accurate research conclusions is through the use of research designs that feature systems for controlling variables.
32. _____ The eminence of quantitative methods in research is still warranted today.
33. _____ Reality is not an external entity, but rather something formed through individual cognition.
34. _____ Reality exists in different forms, where each is mediated by an individual's own experiences.
35. _____ The best way to understand something is to observe it directly in its natural setting.
36. _____ Researchers should have a strong background in quantitative research methods.
37. _____ Building a body of knowledge is itself valuable and reason enough to conduct research.
38. _____ It is the mission of an inquirer to act on behalf of research subjects.
39. _____ The quality of a research study is determined primarily by the action it prompts.
40. _____ Smaller, purposively collected samples provide the most meaningful data for research.
41. _____ Replicable findings equal valid research conclusions.
42. _____ Researchers are capable of making totally unbiased observations during the course of a study.
43. _____ Knowledge gained through research is primarily a consequence of personal interpretation.
44. _____ The object of research is to understand others' individual realities.
45. _____ There are societal truths in the world that transcend even cultural boundaries.
46. _____ Today's scientific facts are likely tomorrow's fiction.
47. _____ The discovery of truth is a product of quality research.
48. _____ Objectivism is a philosophy to which all researchers should adhere.
49. _____ Knowledge gained through impartial, systematic inquiry illustrates truth.
50. _____ Quality research is an objective process.
51. _____ Different research studies help to illustrate different realities.
52. _____ Knowledge obtained through research is morally neutral.
53. _____ Knowledge generated by research is mediated by the personal views of an inquirer.
54. _____ The strength of a theory is best determined by using data-based quantitative methods.
55. _____ It makes sense that researchers who use the same methods to study the same topic will often reach different conclusions.
56. _____ The process of research is value-free.
57. _____ Research is essentially a subjective process.

58. _____ A researcher should make every attempt to limit their contact with research subjects in order to control for biased research conclusions.
59. _____ In research, the object of study is something completely independent of the researcher.
60. _____ There is really only one truth, but there are lots of opinions.
61. _____ Knowledge is more a consequence of social, political and/or economic forces than of actual evidence.
62. _____ Research serves to describe and explain the world as it actually is.
63. _____ The research process is inherently laden with the values of an inquirer.
64. _____ Theories derived from research describe a real external world that exists independently of human perception.
65. _____ Researchers seek to uncover the truth.
66. _____ The best way for a researcher to understand something is to experience it themselves.
67. _____ The personal values of an inquirer should be separated from the research process.
68. _____ The eradication of personal views during the research process is unlikely because those are the views that prompted research to begin with.
69. _____ The process of research is divorced from the personal beliefs of an investigator.
70. _____ A good researcher should have a strong background in qualitative methods.
71. _____ A research design should emerge as a study unfolds.
72. _____ Quantitative methods are most appropriate for the type of research questions I am likely to ask.
73. _____ Larger, randomly selected samples provide the most useful data for research.
74. _____ Reality is something that exists regardless of whether a researcher chooses to study it or not.
75. _____ Quantitative research methods are ineffective for the research topics I am likely to explore.
76. _____ The failure of an inquirer to bracket their personal beliefs during the research process is detrimental to reaching accurate conclusions.
77. _____ Statistical tools are necessary for understanding and interpreting research data.
78. _____ It is naïve to think that an inquirer can separate their personal biases from the research process.
79. _____ There is such a thing as an independent reality, but it is only partially apprehendable because the human mind is imperfect.
80. _____ Interaction with people in their natural environment is more valuable than the statistical testing of hypotheses.
81. _____ The object of research is to reveal reality.
82. _____ Qualitative research methods are most appropriate for the types of research questions I am likely to ask.
83. _____ Although researchers can never prove something, they can determine how likely something is through objective analyses.
84. _____ True knowledge is garnered through objective analyses.
85. _____ The quality of a research study is not determined by the values of an inquirer, but rather through the application of conventional rules for rigor and validity.
86. _____ It is important to have the words of research participants present in a research report.
87. _____ Researchers create knowledge by attaching their own meaning to objects.
88. _____ Carefully constructed and validated instruments are the most effective tools for collecting quality research data.
89. _____ Knowledge gained through research is based upon subjective choices made by an inquirer.
90. _____ If researchers try hard enough, they can ultimately get to the truth.
91. _____ Facts gleaned through research are likely to change over time.

92.	_____	Any theory 'worth its salt' must be subjected to data-based quantitative tests.
93.	_____	An important part of research is determining how the world should be before a study begins.
94.	_____	Researchers can begin to approximate reality through ongoing critical analysis.
95.	_____	The personal ethics of an inquirer are inherently intrinsic to the research process.
96.	_____	Each step of a research study's design should be carefully structured from beginning to end before the process of data collection begins.
97.	_____	Researchers studying the same topic have access to the same reality.
98.	_____	A researcher's intuition plays a central role in the creation of knowledge.
99.	_____	Hypotheses are what are formulated at the end of the research process.
100.	_____	Truth is determined by the cultural, economic and/or political climate of the time.
101.	_____	Quantitative research methods are often inappropriate because they silence the voice of an individual.
102.	_____	Reality is a stable entity – it is only people's perception of reality that varies.
103.	_____	The most effective way to conduct research is to formulate a hypothesis and then test it using empirically collected data.
104.	_____	All researchers experience a different reality.
105.	_____	It is important for researchers to feel morally connected with research subjects.
106.	_____	The generalization of research conclusions is not reasonable beyond the immediate setting of a study.

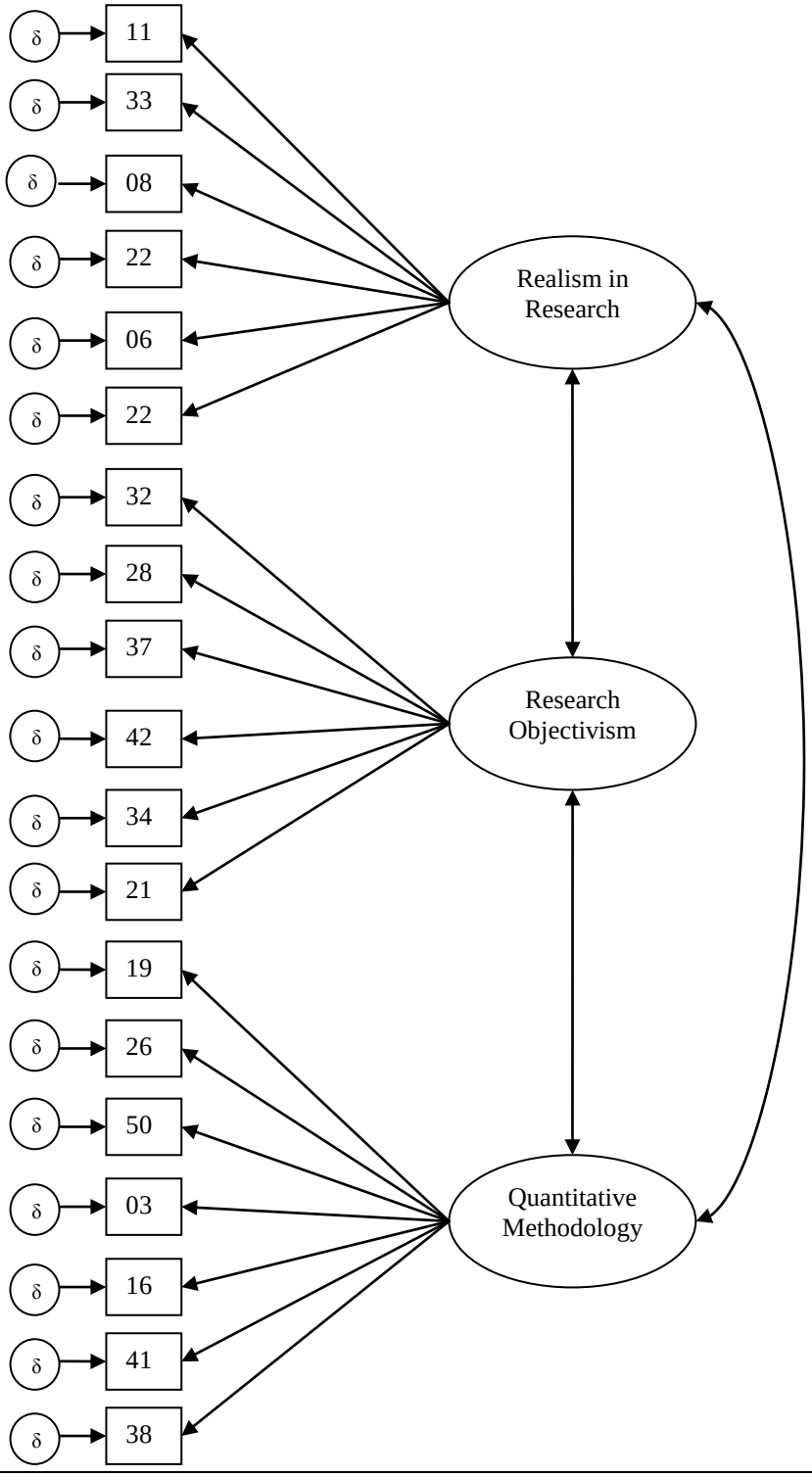
Part II

Directions: In the spaces below, please write any additional statements you feel are necessary for measuring the domains contained in Table 1, but were not included in the list of items presented for your review. In addition, please indicate the domain number from Table 1 you feel the statement represents.

	Content Domain	Philosophical Statement
1.	_____	
2.	_____	
3.	_____	
4.	_____	
5.	_____	
6.	_____	
7.	_____	

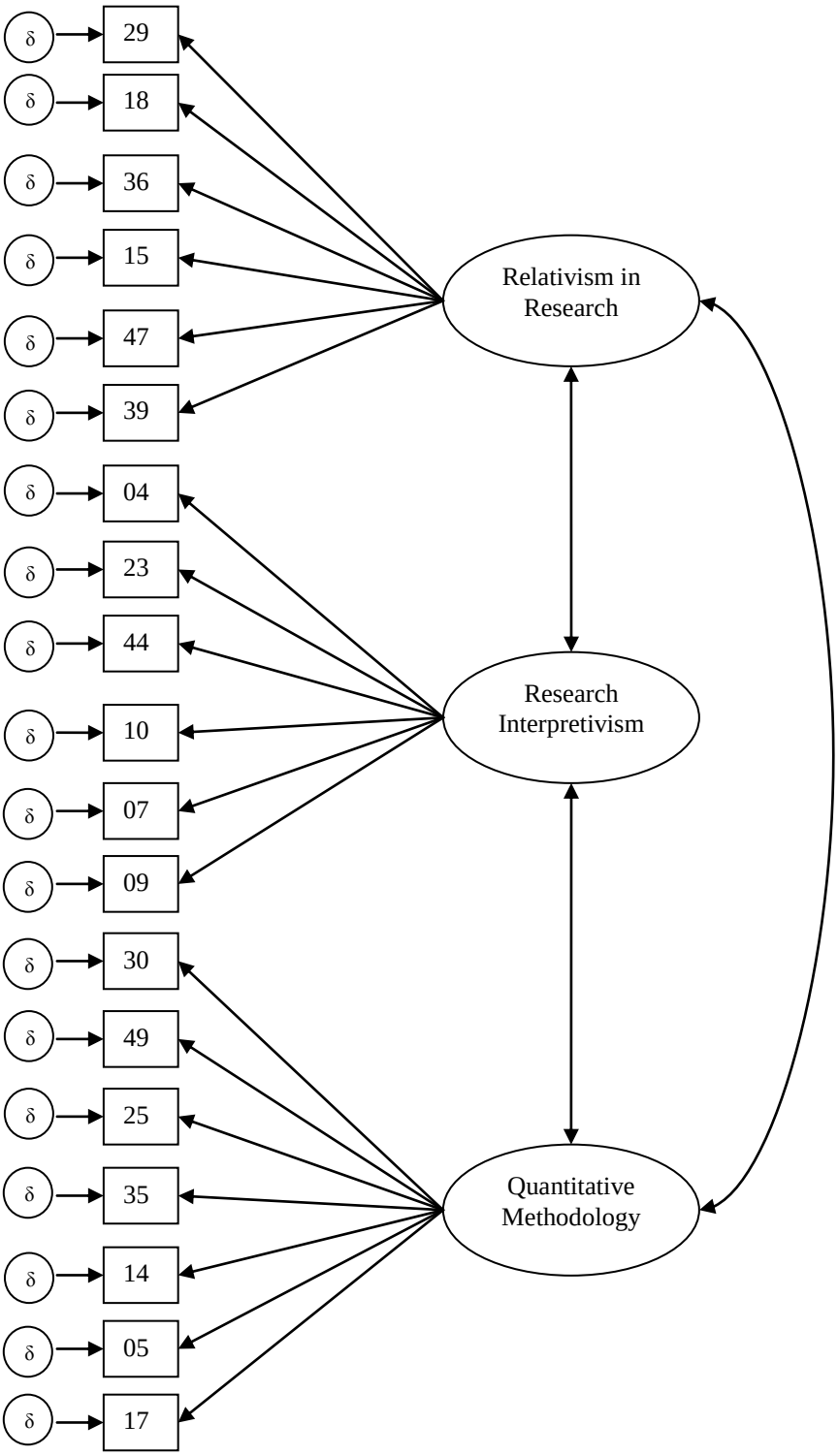
Appendix D. Conceptual Depiction of Instrument Dimensionality

Traditionalist Scales



NOTE: Boxes represent observed variables. Numbers represent sequence on field test survey.

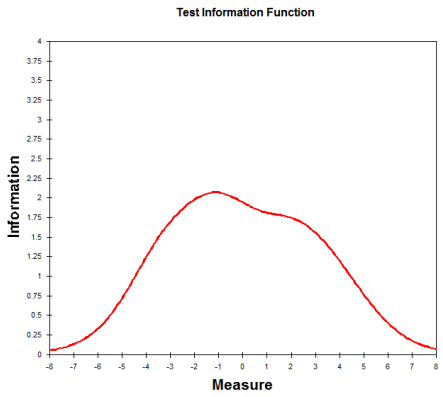
Interpretivist Scales



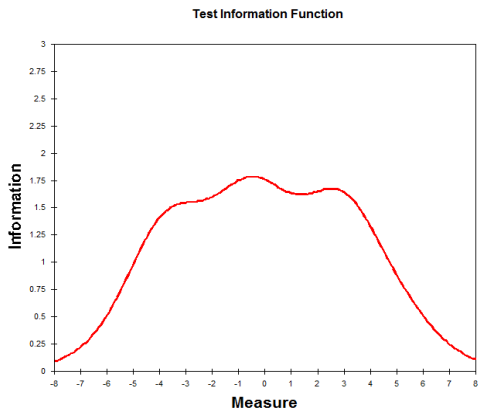
NOTE: Boxes represent observed variables. Numbers represent sequence on field test survey.

Appendix E. Scale Information Functions

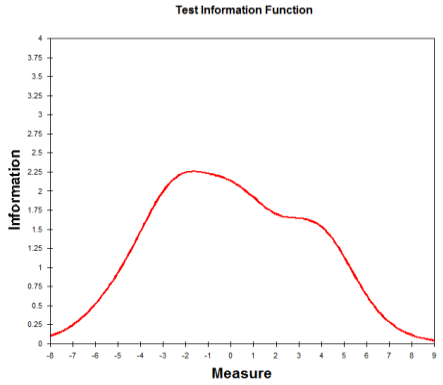
1. Realism in Research



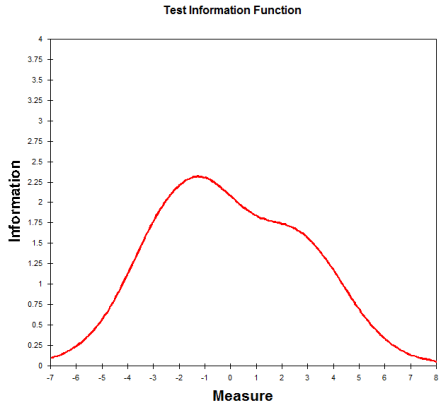
2. Research Objectivism



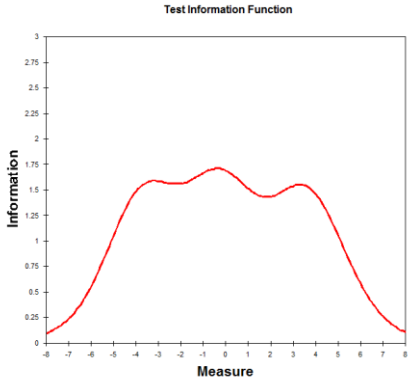
3. Quantitative Methods



4. Relativism in Research



5. Research Interpretivism



6. Qualitative Research