ENGAGING METHODOLOGICAL PLURALISM

8/10/15

Pamela A. Moss, University of Michigan

Edward H. Haertel, Stanford University

[Now PUBLISHED, with minor substantive edits, as:

Moss, P. A. and Haertel, E. H. (2016). Engaging methodological pluralism. In D. Gitomer and C. Bell (Eds), *Handbook of research on teaching* (5th Ed), (pp. 127-247). Washington, DC: AERA.

Feel free to message the senior author (pamoss@umich.edu) with questions regarding citations.]

ACKNOWLEDGEMENTS

In writing this chapter, we have freely sought feedback and advice from many colleagues. The handbook editors also solicited feedback on our behalf from several reviewers, some of whom remain anonymous to us. Without exception, these many commentators gave generously of their time, offering wise and insightful comments and suggestions. This chapter has benefited greatly as we have tried to incorporate their ideas and recommendations. We wish to acknowledge the handbook editors, Drew Gitomer and Courtney Bell, our two consulting chapter editors, Margaret Eisenhart and John Willett, and the following individuals who read and commented on portions of our draft, or in some cases on the draft in its entirety: Benjamin Alcott, Leah Bricker, Frederick Erickson, James Paul Gee, Andrew Ho, Diane Larsen-Freeman, Chauncey Monte-Sano, Daniel McFarland, Richard Murnane, Annemarie Palincsar, William Penuel, Donald Peurach, Matthew Ronfeldt, Mary Schleppegrell, Thomas Schwandt, Camille Wilson, and Leisy Wyman, as well as University of Michigan graduate students participating in the 2014 Complementary Methodologies class. We also extend a special thanks to Jeremy Golubcow-Teglasi, whose library research, article summaries, and comments on successive drafts were enormously helpful. We of course accept full responsibility for the many omissions and for any errors remaining.

ENGAGING METHODOLOGICAL PLURALISM

8/10/15

Pamela A. Moss, University of Michigan
Edward H. Haertel, Stanford University

This is the first time the *Handbook of Research on Teaching* has addressed research methodology in a single integrated chapter. Previous editions have included multiple methodology chapters, each more narrowly circumscribed by traditions or sets of methods, by substantive issues, or by aspects of the research process; and many (though not all) such chapters could be situated within one or the other of the constellations of methods/methodologies that have come to be labeled "quantitative" and "qualitative". While the terms qualitative and quantitative can be useful for characterizing "methods", they underrepresent the rich range of methodologies in education research, glossing over key differences and unique affordances of different methodologies. The growing prominence of "mixed methods" (Tashakkori & Teddlie, 2010) has drawn attention to the value of working across methodologies, but again, the broad terminology does not illuminate the many ways in which methodological pluralism might be engaged.

In this chapter, we explore a range of methodologies and ways of working across them—in complement and respectful challenge—so as to support our collective capacity to learn from our differences and to build more robust programs of research. One common approach to working with multiple methodologies has been to argue that the research question or problem should guide the choice of method(s). Our concern with this approach is that it is not simply a matter of questions leading and methods following. Rather, both the methods we know and the methodologies that ground our understandings of disciplined inquiry shape the ways we frame problems, seek explanations, imagine solutions, and even perceive the world (Bransford et al., 2006; Guile, 2011; Lamont & Molnár, 2002). Learning about other methodologies can illuminate the limitations and the taken-for-granted understandings and practices of the methodologies we know and let us imagine new possibilities (Moss, Phillips, Erickson, Floden, Lather, & Schneider, 2009).

Our major goals are (a) to review and illustrate an instructive range of methodological traditions to support readers in understanding "the potentially useful work" (Cobb, 2007) that different methodologies might do, but also the partialness of the perspectives on social phenomena that any one can offer; (b) to review and illustrate alternative approaches for drawing productively on multiple methodological traditions—engaging in "methodological pluralism" (Topper, 2005; Williams & Vogt, 2011)-- in addressing educational problems within and across local contexts; and (c) to provide a conceptual framework or "map" that functions as a set of heuristics to support readers in learning from these different traditions and in traversing the boundaries among them to build or participate in multi-methodological programs of research. In our conclusions, we consider the sort of social infrastructure that might sustain "integrated and coherent conversation[s]" (Bransford et al., 2006) across methodological traditions in addressing educational problems.

We illustrate the methodological traditions and approaches to methodological pluralism with concrete examples of research focused on a common theme relevant to the goals of the chapter and the *Handbook's* central focus on research on teaching: how teachers and other education professionals both use and engage in research about teaching and learning to inform their practice. We intend use and engagement to encompass both the use of data and research produced by others and engagement in the production of data and research locally to inform routine practice. This encompasses studies classified as focusing on data use, research use, or knowledge utilization by teachers or about teachers and teaching as well as studies that can be read with these questions in mind (e.g., studies of teachers' learning in communities focused on examples of students' work). By focusing on a common theme, we seek to illuminate the ways in which multiple methodological traditions can combine to enhance understanding and possibilities for action. The theme also allows us to foreground the increasingly crucial issue of how education research can be incorporated into educational practice. Ultimately, we intend the chapter to inform dialogue about how research can both contribute to general knowledge and improve education practice in local contexts.

Readers for whom we write include novice education researchers who are just learning about the field; experienced education researchers who have expertise in one or

more methodological traditions but are novices to others; readers who are responsible for research policy, funding priorities, or preparation of researchers; and readers whose professional practice is directly informed or shaped by education research. Recognizing that most readers of our chapter will be novices in some areas, our choices are largely informed by our pedagogical intent. While our representation of any single methodological tradition or approach to methodological pluralism will privilege the commonplace over the cutting edge, the distinctive contribution of this chapter lies in the synergies the comparisons offer. Our hope is to support readers in developing a professional vision (Bransford et al., 2006; Goodwin, 1994; Grossman et al., 2009) necessary to participate in a robust multi-methodological practice and to respond resourcefully as new circumstances and opportunities for learning arise.

In the remainder of this introduction, we bound the scope of the chapter, define key terms, sketch our methodology for producing this review of literature on methodology, and provide an outline for the rest of the chapter.

Scope and Definitions of Key Terms

This cannot be a review of literature in any traditional sense of the term, where a body of work is circumscribed and comprehensively reviewed. The scope of the field of education research methodology is too vast, varied, and dynamic. At best our chapter will prepare readers for future learning (Bransford & Schwartz, 1999; Schwartz & Bransford, 1998) by illustrating a selection of methodological traditions and of approaches to working across them. It will provide one map of the field of education research—unavoidably partial, perspectival and historically situated—that supports readers in understanding some of its key affordances, in using the comparisons to support learning, and in participating collaboratively in a collective research practice that draws productively on multiple methodologies.

Methodology can be defined most straightforwardly as discussion *about* methods (Abbott, 2004) with "about" signaling the centrality of theory and reflection in methodology. By methods, we refer to particular procedures through which research is enacted, including procedures of data collection or production, of data analysis and interpretation, of validation or evaluation, and so on. ***Methodologies*** encompass sets of methods within a "theory and analysis of how research *does or should* proceed"

(Harding, 1987). As this definition implies, methodology can be *prescriptive*, entailing normative statements of what should (or shouldn't) be done, or *descriptive,* entailing representations of what has been done or "research performed on research methods" (Alastalo, 2008). Further, methodology can be used as both a count noun (referring to different methodologies) and an abstract or mass noun referring to the discussion of methods more generally. We will use the term in all these ways.

Our focus will be on methodologies for empirical research with social phenomena relevant to education. By empirical research, we mean research drawing on observation and experience. By social phenomena, following Camic, Gross, and Lamont (2011), we include research focusing on human cognition, action, and interaction, including the interaction among people and their tools, alone or in collective units, often embedded within organizations or other social structures. Tools includes both conceptual tools like language and material tools like computers or research reports, that are inherited, taken up, modified, and/or produced in interaction. We focus primarily (but not exclusively) on methodologies that have been located within the social sciences and to a lesser extent within the humanities. We use the term "social research" to encompass methodologies in the humanities as well as the social sciences. Within the social sciences, we include methodologies that resemble those employed in the natural sciences and those that are distinct to the social sciences. While we have privileged methodologies that have been prominent in education research and research on teaching, we have also included methodologies that appear promising if not yet prominent in education research, including those in which teaching is relevant but decentered—incorporated as an interactive element of a larger unit within which teaching might be understood. Our many examples focusing on how teachers and other education professionals both use and engage in research about teaching and learning to inform their practice will keep the chapter well focused on the organizing theme of the volume.

We use the concept of "***methodological tradition***" to organize our discussion, acknowledging both the historically situated and dynamic nature of methodological theory and practice. As Abbott (2004) describes it:

> Methodological traditions are like any other social phenomena. They are made by people working together, criticizing one another, and borrowing from other traditions. They are living social things, not abstract categories in a single system. (p. 15)

Both methods and methodologies evolve over time, and overlap in complex ways with disciplines, philosophical perspectives, including philosophies of science, national borders, and substantive theories. Many of the disciplines that inform empirical research in education—anthropology, economics, history, linguistics, psychology, sociology—span multiple methodological traditions. Further, some methodological traditions have evolved and dispersed to encompass different philosophical perspectives; and some philosophical perspectives have been enacted through multiple methodologies. Further still, national cultures shape research practice; and relationships of research communities with their governments and other funding sources can lead to further differences. A tradition, for us, would be reflected in sustained, historically situated dialogue and mutual citation among those who locate their work within that tradition, and where the participants routinely acknowledge (embrace, revise, or resist) common understandings and practices.

With the term "***methodological pluralism***" (Topper, 2005; Williams & Vogt, 2011) we encompass a range of perspectives and practices for working productively across methodological traditions in studying social phenomena and changing social practice. This includes approaches where the goals are for methodologies to complement one another in deepening understanding of a complex social phenomenon and/or to challenge one another by illuminating taken-for-granted assumptions and limitations of any particular methodology; approaches focused on particular problems and on broader themes encompassing sets of problems; approaches intended to construct innovations in research practice that might incorporate what has come before into a novel methodology; and approaches involving generalizations and research syntheses of various sorts. While reflecting instructive differences, they all acknowledge that social phenomena are simply too complex to be adequately understood from a single methodological perspective.

Methodology for Review

The literature reviews that underlie this chapter can best be conceptualized as "scoping" reviews. Scoping reviews—a concept more prominent in health-related

research – are intended to produce "a high level map of an area of research that can inform future research…. The focus is on describing the breadth and key characteristics of research on a topic" (Scoping Studies, 2013; see also Arksey, 2005). Actually, we have conducted three scoping reviews—one focused on methodological traditions, one focused on explicit approaches to methodological pluralism, and one focused on conceptions of generalization, including how studies reflecting different methodological traditions contribute to general knowledge and support local practice. [1]

Our review process was further informed by the concept of "theoretical sampling" in choosing what methodological traditions, approaches to methodological pluralism, and conceptions of generalization to represent. Theoretical sampling, a concept borrowed from Grounded Theory (Charmaz, 2014; Corbin & Strauss, 2008; Glaser & Strauss, 1967), is an iterative process intended to support the development and refinement of categories—in our case, categories for mapping the field. The categories iteratively developed served two functions: (a) they helped us select particular methodological traditions, approaches to pluralism, and conceptions of generalization and (b) they helped us draw instructive comparisons among these illustrations while honoring the ways in which insiders represent their work. We will provide substantive particulars for each scoping review at the beginnings of the relevant sections.

In conceptualizing the chapter, we drew heavily on the increasingly popular metaphor of geography, cartography, or "maps" in representing fields of study. As a glance at the many handbooks, textbooks, and articles representing social or education research methods or methodologies suggest, there is no one correct or even widely accepted way to map the field. Different scholars foreground different sets of categories--philosophical perspectives, historical stages, disciplinary perspectives, stages of the research process, methodologies, substantive issues, and so on—at different grain sizes, and often with different ways of grouping and naming key elements. Different maps serve different purposes, foreground different distinctions, and honor or gloss over boundaries that are meaningful to different communities of researchers. As a number of

[1] We also conducted a literature search for empirical research and reviews relevant to our theme of how teachers and other education professionals use and engage in research (with the help of Jeremy Golubcow-Teglasi). But the goal of this review was to identify research reports we could use as examples throughout the chapter, not to represent the scope and dimensions of research on this topic.

scholars have noted, such maps do political work (with their potential to shape understanding and practice) and are best evaluated in terms of usefulness for their purposes and, ultimately, their social consequences (Clarke & Primo, 2012; Gieryn, 1999; Lamont & Molnár, 2002; Outhwaite & Turner, 2007; Pascale, 2011).

In their extensive review of literature on how boundaries are studied in the social sciences (including boundaries within social sciences themselves), Lamont and Molnár (2002) characterize *symbolic* boundaries as "conceptual distinctions made by social actors to categorize objects, people, practices, and even time and space….[They] separate people into groups and generate feelings of similarity and group membership" (p. 168). They draw an instructive analytical distinction between symbolic and *social* boundaries, which entail "more durable and institutionalized social differences" (Pachucki, Pendergrass, Lamont, 2007, p. 331) and can be "manifested in unequal access to and unequal distribution of resources (material and nonmaterial) and social opportunities" (Lamont and Molnár, 2002, p. 168). They note that examining boundaries "allows us to capture the dynamic dimensions of social relations, as groups compete in the production, diffusion, and institutionalization of alternative systems and principles of classifications" (p. 168), questions they show are as relevant to scientific communities as to other social groupings. The methodological traditions and approaches to pluralism we review all involve symbolic boundaries and, arguably, social boundaries as well. Lamont and Molnár draw on Star and colleagues to argue that while boundaries divide—and can create inequities in access to power and resources—they also create "important interfaces enabling communication across communities" (Bowker and Star, 1999, in Lamont and Molnár, p. 180)—a feature of boundaries that is central to our chapter. We will touch upon the social aspects of boundaries—and the potential for inequity—in our concluding remarks. Here we focus on the symbolic aspects of the boundaries that create these distinctions, dialogue across which enables methodological pluralism and affords opportunities for learning.

Given our purposes of supporting readers in learning from differences across methodological traditions, we have tried to locate and honor boundaries that are meaningful to the proponents of a given tradition and that have empirical referents in published literature of that tradition. Simultaneously, we have worked within and against

the maps that others have provided to develop categories that enable respectful comparisons and challenges to support a robust multi-methodological practice that our readers can then challenge and grow.

Organization of the Chapter

The organization of the chapter can be seen only partially in its major headings and subheadings; equally important are the cross cutting themes that support comparisons and synergies across different perspectives, all of which provide resources to support readers' learning. We narrate a synoptic overview here. Additional details will be provided at the start of each major section. An outline of the chapter is presented in Table 1 (at the end of this section).

The chapter is divided into four major sections in addition to this introduction and a conclusion: one section focusing on different Methodological Traditions (MTs); one section focusing on different Approaches to Methodological Pluralism (AMPs) or ways of drawing productively on multiple methodological traditions; one section focusing on Conceptions of Generalization (CoGs), with particular attention to generalizing across methodological traditions in supporting understanding and action; and one section focusing on integration of resources from the previous sections to provide heuristics for conceptualizing multi-methodological programs of research. While the first three of these sections foreground the perspectives of the theorists whose work we review, the fourth brings our own perspectives on methodology more directly and critically into the dialogue. Given the breadth and length of the chapter, we anticipate that readers may prefer to read the chapter in parts rather than straight through. Readers can get a substantive overview of the chapter by reading through this general introduction, the introductions to each of the next three major sections and then reading the fourth, "Heuristics for Engaging Methodological Pluralism" in full. That, taken together, with the outline in Table 1, will then support decisions about subsections on which they might want to focus in more depth.

The section on *Methodological Traditions* (MTs) presents 10 illustrative MTs, comparisons among which foreground key categorical distinctions in social research. Each MT is presented in its own subsection from the perspectives of one or more widely-

cited theorists of that MT, nearly all of whom have also engaged in education research. To facilitate comparisons, each MT is presented in terms of the same set of general categories, which we selected to be sufficiently ecumenical to allow representation of the MT in its own terms (i.e., aims and conceptions of social phenomena, design and enactment, and reporting and evaluation). Concrete examples of studies drawn from our theme of how teachers and other education professionals both use and engage in research about teaching and learning to inform their practice are used to illustrate each MT (and will be drawn on again in subsequent sections).

The section on *Approaches to Methodological Pluralism* (AMPs) presents a range of perspectives on how one might work across methodologies in social research. We begin with a brief history, followed by a review of the ways in which multiple methods have been used within methodological traditions and within particular fields of study. Then we focus on three quite distinct AMPs that provide conceptual resources (analytical language) for methodological pluralism, each with a somewhat different focus: the "Mixed Methods" (MM) AMPs tend to foreground ways of working across methods or methodologies, most often crossing boundaries named simply as qualitative and quantitative; AMPs under the rubric of "multi-, inter-, or trans-disciplinary" research (MD, ID, TD, respectively or *MIT*-D to encompass them all) tend to foreground ways of working across disciplines, including but not limited to their methodologies; and conversations under the rubric of "complex systems theory" (CST) tend to foreground systems of various sorts and the ways in which their parts dynamically interrelate, the study of which entails multiple methods and methodologies. Each of these focus prospectively on the design and enactment of research.

In the next section on "*Conceptions of Generalization* (CoGs), we consider different approaches to constructing generalizations across the particulars of different studies, both to contribute to general knowledge and to inform practice in particular contexts. Here, much of the literature we review considers how generalizations can be constructed across different MTs. And so this third section, while quite distinct from the AMPs we review, is also central to engaging methodological pluralism.

In the penultimate section, on "*Heuristics for Engaging Methodological Pluralism,*" we draw together resources from the earlier sections to suggest a set of

heuristics for conceptualizing or mapping programs of research informed by multiple methodologies. In our *Concluding Comments*, we consider implications of our review for research preparation, policy, and practice.

Table 1: Outline

1. [INTRODUCTION]
    - Scope and Definitions of Key Terms
    - Methodology for Review
    - Organization of the Chapter
2. METHODOLOGICAL TRADITIONS (MTs)
    - [Introduction]
    - Experimental and Quasi-Experimental Designs for Generalized Causal Inference (GCI)
    - Ethnography/Ethnographic Research (EthR)
    - Small N or Comparative Case Studies (CCS)
    - Discourse Analysis (DA): Critical Discourse Analysis (CDA) and Systemic Functional Linguistics (SFL)
    - Survey Research and Measurement (SR&M)
    - Design Based Research (DBR)
    - Social Network Analysis (SNA)
    - Complex Adaptive Systems Research (CAS): Agent Based Modeling (ABM)
    - Critical Race Theory (CRT)
    - Participatory Action Research (PAR)
    - Analytic Summary
3. APPROACHES TO METHODOLOGICAL PLURALISM (AMPs)
    - [Introduction]
    - Mixed Methods (MM) Research
    - Multidisciplinary, Interdisciplinary, and Transdisciplinary Research (MIT-D)
    - Complex Systems Theory (CST) as an Approach to Methodological Pluralism
    - Analytic Summary
4. CONCEPTIONS OF GENERALIZATION (CoGs)
    - [Introduction]
    - Theory-Based Logics of Generalization
    - Literature Reviews/Research Syntheses
    - Generalizing as Learning in Local Contexts
    - Analytic Summary
5. HEURISTICS FOR ENGAGING METHODOLOGICAL PLURALISM
    - [Introduction]
    - The Function of (Our) Heuristics
    - Illustrating our Heuristics
    - Alternative Approaches for Structuring Programs of Research
6. CONCLUDING COMMENTS
    - Situating Our Argument in On-Going Debates about Social Research
    - Implications of a Dialogic Approach to Methodological Diversity

APPENDIX
- Ethical Considerations

2. METHODOLOGICAL TRADITIONS (MTs)

> "The theory that we choose to work with, explicitly as researchers and perhaps implicitly as teachers, will dictate how we describe and investigate the world. It controls how we select, out of all that is possible, what to investigate or explain, what types of questions we ask, how data are collected, and what kinds of explanations of the data are considered valid." (Larsen-Freeman & Cameron, 2008, p. 16)

Each of the methodological traditions (MTs) we have selected brings unique affordances for understanding social phenomena; and comparisons among traditions illuminate key and enduring categorical distinctions in social research. We have chosen MTs that are prominently reflected in or that appear promising for research on teaching, although different choices could have been made. Thus we must emphasize that these are only a subset of the many traditions we could have selected. The categorical distinctions the traditions illustrate provide useful heuristics (Abbott, 2004) for designing multi-methodological programs of social research by prompting questions regarding what can we understand (differently) about an educational problem if we look at it using various methodological perspectives. By further illustrating each selected tradition with empirical studies focused on a single theme--how teachers and other education professionals both use and engage in research about teaching and learning to inform their practice--we show how methodological traditions can complement one another in addressing particular problems.

In the space of a single methodology chapter, we cannot represent particular methods or methodologies in sufficient depth to support readers in enacting them. Rather, our goals in representing any particular MT are to allow readers to understand its primary affordances, to consider key challenges it might raise for the MTs with which they are familiar, and to decide whether further study is warranted for their own work, but not to guide application or enactment. References will point readers to possible next steps for further study.

The MTs we have selected each have complex histories, representing internal differences and tensions, within and across national, disciplinary, and substantive boundaries, which are continuing to evolve. Boundaries are often ambiguous, and

hybrids develop. Most have multiple handbooks or encyclopedias devoted to them, which can better represent this complexity. To attempt to provide a summary characterization would risk misleading readers about the nature of methodological traditions. To avoid such essentializing, we have chosen instead to illustrate each tradition from the perspectives of one or more theorists who are recognized as leading scholars in that tradition and, wherever possible, who have engaged in research relevant to teaching. In describing each tradition, we function like historical ethnographers, representing our chosen theorists' perspectives in their own terms, and introducing these methodological perspectives to readers who may be novices to that tradition.

The framework we use to represent each MT focuses on categories/features we have selected to be sufficiently ecumenical to allow fair representations (of each MT in its own terms) yet still enable instructive comparisons. A fair representation requires that our comparative categories allow us to represent each tradition as insiders would. Further, we need to acknowledge that terms whose meanings are taken for granted in some traditions (e.g., validity, case, power), may take on quite different meanings in other traditions; and the features a term describes in one tradition may be labeled or grouped in different ways in other traditions. In our presentations of each MT we strive to acknowledge these differences. With those concerns in mind, we compare traditions in terms of the following features:

AIMS and CONCEPTIONS OF SOCIAL PHENOMENA

a) the aims or goals of research within the tradition;
b) the understanding or delineation of social phenomena that are the focus of the tradition;

DESIGN and ENACTMENT

c) the logic of inquiry or nature of research design/decisions as a study evolves from initial conceptualization through reporting of a completed study;
d) key methods and/or processes through which the research is enacted, including data collection or production, analysis, and interpretation;
e) the roles of researcher and researched and the relationship among them;
f) the theorists' sense of evolving variations within the tradition and how their work is situated among them

REPORTING and EVALUATION

g) how the research is represented when it is reported to others;
h) how validity, soundness, rigor, worth or quality are conceptualized; and
i) conceptions of how knowledge generalizes or is relevant outside the particulars of what was studied within the MT.

While we use the capitalized headings in our presentation of each MT, the categories underneath these headings may be presented in a different order or, possibly, under a different heading. Conceptions of generalization or relevance outside the particulars studied, for instance, might be most relevant to DESIGN and ENACTMENT in some traditions but to REPORTING and EVALUATION in others.

As we noted in the introduction, we engaged in a process of theoretical sampling (Glaser and Struass, 1967) to arrive at our final selection of MTs. This entailed iteratively drawing comparisons among candidate MTs and characterizing the dimensions along which they differed (informed by dimensions and MTs reflected in other maps of the field) to make sure the set of MTs illustrated an appropriately broad range of affordances for studying teaching. Examples of the dimensions we used include numbers and depth of cases studied, relationships between researchers and researched, and conceptions of change or progress. We postpone discussion of these dimensions reflecting differences among MTs until we have built the vocabulary necessary to understand them in the presentation of individual MTs. Interested readers can turn to the analytic summary at the end of this section or to Table 2 in our Heuristics for Engaging Methodological Pluralism section.

We have selected the following traditions that illustrate instructively different positions in terms of the categories named above. Some are established traditions with long histories in education research and some are less prominent in education research but are established in other social research traditions.[2]

---

[2] Not well represented are texts that describe general methods that might be relevant to a variety of MTs (e.g., Bryman, Bell & Teevan, 2009; Patton, 2002, Yin, 2009). Some such texts focus on quantitative or qualitative research, some on general types of methods named in other ways (e.g., case studies, discourse analysis), some on education or social research more generally. The range of variation is considerably wider within the constellation of methods that might be described as qualitative. These texts often have an early chapter that describes a range of traditions relevant to the methods presented but then move on to the discussion of methods considered more generally relevant across traditions. While these texts provide

1. Experimental and Quasi-Experimental Designs for Generalized Causal Inference (GCI)[3]
2. Ethnography/Ethnographic Research (EthR)
3. Small N or Comparative Case Studies (CCS)
4. Discourse Analysis (DA): Critical Discourse Analysis (CDA) and Systemic Functional Linguistics (SFL)
5. Survey Research and Measurement (SR&M)
6. Design Based Research (DBR)
7. Social Network Analysis (SNA)
8. Complex Adaptive Systems Research (CAS): Agent Based Modeling (ABM)
9. Critical Race Theory (CRT)
10. Participatory Action Research (PAR)

In illustrating a tradition, we have also selected one or two well-enacted examples of studies that, wherever possible, are relevant to our theme of how teachers and other education professionals both use and engage in research about teaching and learning to inform their practice. Our selections of examples were informed by the authors' categorizations of their work, the methodological references they cited, and our own judgments of consistency with our key theorists' perspectives (informed by feedback from our consulting editors and reviewers). Where an example directly relevant to our theme could not be located, we draw on as closely related an example as possible, and sketch the ways in which the theme might be studied from that methodological perspective.

In explaining our organizational choices, we should note that it is not uncommon to organize texts like this in terms of philosophical perspectives or "paradigms," often presented with distinct ontologies (what is seen to exist) and epistemologies (how knowledge is conceptualized and grounded). Methodologies are sometimes located within these broader themes. Philosophical perspectives one frequently encounters include positivist, post positivist, naturalist, interpretive, constructivist, critical, post-

---

readers with useful advice, given the purposes of this chapter, our choice has been to characterize methods in the contexts of how they work within specific traditions.

[3] Our name for this MT is quoted from the title of Shadish, Cook, and Campbell (2002).

structural, postmodern, realist, critical realist, standpoint, and so on. Each of these mega-concepts (Geertz, 1973) subsumes constellations of subconcepts.

Among the most enduring of the categorical distinctions is that surrounding the question of whether or not the social sciences should have the same goals and be practiced in the same way as the natural sciences (Jarvie, 2011; M. Martin & McIntyre, 1994). Those who adopt a "unified" conception of science (NRC, 2002; Phillips & Burbules, 2000) tend to argue that the primary goals of the social sciences, like the natural sciences, should be generalizable explanations and predictions. In contrast, those who take what is often called an "interpretive" (Flyvbjerg, 2001; Rabinow and Sullivan, 1987) approach to social science argue that social phenomena differ from natural phenomena because they are meaningful to the actors involved. Further, meanings are embedded in complex social contexts that shape what can be understood in ways that the actors involved may not perceive. From this perspective, a primary aim of social science is to understand what people mean and intend by what they say and do and to locate those understandings within the historical, cultural, institutional, and immediate situational contexts that shape them (adapted from Moss, et al., 2009). Even this distinction is perceived by some as controversial and not all the MTs presented below could be easily categorized on one or the other side of this distinction.

Moreover the same labels can be used to mean different things by theorists working in different traditions.[4] Sometimes such labels are (mis)used by their critics to name perspectives that insiders--those who use the terms in positioning their *own* work--would not endorse. The labels positivism and postmodernism are among the most misused, often by writers who have not taken the time to represent the history of the concepts from insiders whose work is being described.[5] Moreover, within any given discipline or methodological tradition, one can often find multiple philosophical perspectives. It's not that philosophical perspectives and methodological traditions are orthogonal, but they combine in complex ways. A responsible treatment of these concepts is beyond the scope of this chapter.

---

[4] See, for instance, Martin and McIntyre (1994) and Lincoln and Guba (1995) for two radically different uses of the term, naturalist or naturalistic.

[5] See, for example, St. Pierre's (2002) resistance to the way postmodernism is used to situate her work and Phillips's (1983, 2006) discussion of the many misuses of positivism.

The categories we've used to present the work of key theorists in each MT--aims and conception of social phenomena, design and enactment, reporting and evaluation—signal aspects of their philosophies, including their ontological and epistemological commitments. Where our key theorists use particular philosophical concepts, we will too, and we will gloss them as needed to support readers' understandings. Readers who are interested in further exploring the meaning of terms like those we've listed above can turn, for instance, to the Stanford Encyclopedia of Philosophy (plato.stanford.edu) which is freely available on line. Key publishers of methods texts maintain online search engines for encyclopedias, handbooks, and dictionaries, which can often be accessed through university libraries. We encourage any readers tracking down such philosophical concepts to be sure to read multiple texts on the topic to understand variations in how terms are used and to make sure at least some of the references consulted were written by authors who use the concepts to describe their own work.

We note as well that an adequate treatment of the topic of ethics in research methodology is beyond the scope of this chapter. While there are canonical documents regarding the protection of human subjects (e.g. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1987) as well as professional guidelines (e.g., American Educational Research Association , 2011) that are widely referenced, ethical issues are far broader. Different methodologies and research discourses illuminate different issues, and different relationships between ethics and epistemology. We have included a brief appendix to the chapter that points readers to some useful resources. With those caveats, we turn to our representation of ten instructively different MTs.

Experimental and Quasi-Experimental Designs for Generalized Causal Inference (GCI)

We take the term "Generalized Causal Inference" (GCI) from Shadish, Cook, and Campbell (2002), which serves as one of our key references for this methodological tradition. Methods used within this tradition are among the best known in educational research, including research on teaching. Campbell and Stanley's monograph on *Experimental and Quasi-Experimental Designs for Research*, an early, classic exposition of GCI, was initially published as a chapter in the first edition of the *Handbook of*

Research on Teaching (Gage, 1963), under the title "Experimental and Quasi-Experimental Designs for Research on Teaching," and later reprinted under the shorter title in 1966, 1973, and 1981. The third and fourth editions of the *Handbook* included chapters incorporating related methods, by Linn (1986) and by Crawford and Impara (2001), respectively.

Research in this tradition is often loosely referred to as relying on "quantitative methods," although that term has become freighted as part of the polarizing (and, we would argue, false) dichotomy of "quantitative versus qualitative" research. For that reason (among others), we try to avoid it. In addition to Shadish, Cook, and Campbell (2002), Murnane and Willett (2011) will serve as our second key theoretical resource for this tradition, and we will use elements of a study published by Slavin, Cheung, Holmes, Madden, and Chamberlain (2012) by way of illustration.

Aims and Conceptions of Social Phenomena

The distinctive feature of GCI is its concern with warrants for causal effects. These warrants are principled arguments, reasoning from observations and analyses, to establish justifications for causal inferences. As Murnane and Willett (2011, p. 12) put it, "[Causal] questions … concern the impact of a particular action on one or more outcomes." The logic supporting a causal inference may be complex, but the causal claim itself is generally simple. An "impact" is understood to be a quantifiable effect on a well-defined, measured variable (an outcome). The "action" (often, a "treatment") may be complex, but is, in principle, definable. It must be an action that could potentially be replicated in other times and places. Finally, the warrant addresses only the existence and the magnitude of a possible causal relationship, not the mechanism for that relationship. The inference in GCI is explanatory in the sense that the research design enables the researcher to rule out, to the extent possible, all competing explanations for a given effect, so that if an effect is found, it may be attributed to the potential cause under investigation. However, the patterns of reasoning and the supporting designs and statistical methods at the heart of GCI are explanatory only in that limited sense. The analysis itself offers no account of mechanisms whereby any causal effect arose. The warrant for any account of possible causal mechanisms must come from other sources (a topic we'll return to below in Reporting and Evaluation).

Slavin and colleagues (2012) sought to establish whether a district-level reform model focused on data use (the treatment), which was created by the Center for Data-Driven Reform in Education (CDDRE), had a measureable impact (effect) on elementary students' reading and mathematics scores on state-administered achievement tests (the outcome). Their paper reviews prior research showing that simply providing data to schools and districts generally appears to have little or no measureable impact on student test scores. In the CDDRE reform model, district personnel are assisted in their use and interpretation of data such as standardized test scores and district records of student attendance, disciplinary referrals, special education placements, or grade retentions, all summarized for specific sites and for specific student subgroups. These data are used to identify specific improvement goals, and district personnel are then encouraged to choose and implement targeted interventions to attain those goals, selecting from among instructional interventions for which there is prior empirical evidence of effectiveness. Thus, there is a plausible theory of action whereby the CDDRE reform model is hypothesized to have an effect. This theory guided the design of the district-level intervention and motivated the authors' investigation. However, the *causal* claim must be distinguished from this sensible, theory-grounded rationale. The causal claim itself merely holds that implementing the CDDRE model will lead to increases in test scores.

Stated more formally, GCI is concerned with estimating the effects of well-defined actions, often called *treatments*, on the *units* under study. These units might be students, classrooms, schools, school districts, or states, for example. In the simplest case, the logic of causal inquiry is based on a comparison between units that have and have not experienced the treatment.[6] The ideal comparison would be to compare the same units under both conditions at the same time! This ideal comparison is not possible, because a single unit can only be assigned to one of the two conditions. The outcome that would have been observed under assignment to the other condition is referred to as a "counterfactual." Thus, GCI researchers must find ways to approximate this ideal comparison. The simplest such approximation is to compare outcomes for "treatment" and "control" groups constructed to be as similar as possible, except that one group

[6] More complex cases may involve investigations of multiple, potentially interacting treatments, and may include treatments that vary in degree, as opposed to simply being present or absent.

receives the treatment and one does not. As we'll elaborate, the classic such design—the "true" experiment"—assigns units to treatment versus control groups by chance alone ("random assignment"). The goal is to be able to interpret post-treatment differences in the outcome variable to the treatment, and not to other pre-existing differences between these groups (often referred to as "alternative explanations" or "rival hypotheses") that might otherwise account for any effect found.

For purposes of analysis, the description of these units is quite sparse. The unique features of each particular unit—each student or classroom or school, for example—are mostly ignored within the GCI methodology itself, although specific dimensions of variation may be measured and incorporated into the statistical model as *covariates*[7]. Units are typically regarded as independent of one another, that is, of not interacting or influencing one another, although more complex statistical models may relax this assumption and may also explicitly account for the nesting of smaller units (e.g., students) within larger units (e.g., classrooms). Thus, it can be seen that the formal view of social phenomena from strictly within the GCI tradition is highly simplified and schematic. Sound and useful research employing GCI will almost certainly be informed by a deep understanding of the settings within which participants live and work, and within which treatments are implemented and data are collected. However, this knowledge is mostly brought to bear before the formal research actually begins, and in the discussion of the findings after it is completed.

Likewise, the understandings and perspectives of research participants themselves are generally outside the purview of the GCI methodology. Key design choices are made independent of, and prior to, the actual conduct of research with participants. To borrow a term from ethnography (discussed in the following section), the investigator's stance is *etic*, not *emic*, striving toward an objective determination of the magnitude of some effect that can be attributed to a specified cause.

Design and Enactment

Study design is likely to begin with the development of a clearly stated, falsifiable research question. One (or a small number) of potential causes for some phenomenon of

[7] Covariates are variables related to outcomes but not of primary interest. They are included in statistical models to improve accuracy and to reduce bias.

interest must be singled out for investigation. The causal variable (treatment) must be defined. This treatment variable must be *manipulable*, meaning it must be possible in principle to influence its level for each unit. So, for example, class size is manipulable (because students could be assigned to classes of different sizes), but students' first language is not (because a student cannot be assigned to a different first language than her own). Likewise, one (or a small number) of outcomes must be defined. The researcher may expect a treatment to have many effects, including some that may not be measurable, but for purposes of conducting a study, specific, quantifiable outcomes must be chosen for investigation. Finally, to improve accuracy, the researcher may identify additional quantifiable variables (covariates) plausibly related to the outcome or potentially influencing the relationship between the treatment and the outcome. These may be measured and used in statistical models to improve the precision with which causal effects can be estimated. In choosing or creating outcome variables and covariates, the researcher is likely to draw upon the models and methods of measurement theory to evaluate the reliability and validity of the measures chosen. Other things being equal, measures with greater precision (higher reliability, less measurement error) are preferable.

For instance, Slavin and colleagues (2012) framed their research question as follows: "In comparison to control groups, what were the effects of CDDRE participation (controlling for pretests) on state tests of reading and mathematics at the elementary and middle school levels?" (p. 380). Their units were districts and their outcome variable was students' performance on state tests. The districts were randomly assigned to treatment (received CDDRE) and control (did not receive CDDRE) groups. This is a multi-level design with schools nested within districts (and students within schools), such that the outcome was measured at the student level but the treatment effect was estimated at the school level. As the research question specifies, the students' scores on the state test prior to treatment (pretests) were used as covariates or controls to allow a more precise estimate of the treatment effects by taking into account some of the differences between groups left after random assignment. We'll elaborate these concepts a bit more below.

Choices of treatment variables, outcome variables, and covariates are informed by theory, prior research, and often by pilot studies, but they are specified in advance of the GCI study itself. Covariates (if any) must generally be measured before the treatment is administered, to assure that they are not impacted by the treatment, which would distort the estimate of the treatment's effect on the outcome variable. The fact that variables are specified in advance and are quantifiable distinguishes GCI from many other research methodologies in which research questions and variables may emerge or change in the course of the research.

Sample selection: The research design often, though not always, begins with a clear definition of some large, often very large, *population* of units that might be exposed to one or another treatment. The idealized study would proceed by drawing a random (or representative) *sample* from that population. The larger the sample, the more accurately population characteristics can be inferred from the sample. Statistical sampling can help assure that study participants (or more generally, the units sampled for inclusion in the research) are representative of a specified population. When this is possible, it greatly strengthens the statistical generalization from actual study participants to the larger population of potential study participants.[8] Note that *sampling* is distinguished here from the subsequent step of *random assignment* to treatment versus control groups, which can be used to help assure that the groups assigned to treatment and control conditions are similar in all respects.[9]

Statistical sampling from a well defined population makes possible a statistical inference from the sample to the full population. This is highly desirable, but actual studies often fall short of this ideal. Researchers must often pursue delicate negotiations to secure access to research settings, and are rarely able to secure a truly representative sample from a broad population. In the study by Slavin and colleagues (2012), random *sampling* from a large population was not possible. The CDDRE formed partnerships

---

[8]While simple random sampling is perfectly adequate, additional precision and cost saving can be obtained with additional methods of sampling. This complex and specialized topic goes beyond the scope of our chapter, but to offer some examples, cluster sampling may reduce costs by simplifying the logistics of data collection; stratified sampling may yield samples that more closely match the characteristics of the population; oversampling of subgroups may increase statistical power for testing additional hypotheses concerning treatment effects within those subgroups.

[9] The mechanism for assignment of units to treatment versus control conditions is of special importance in GCI, as discussed below in the section on "Identifying causal effects."

with education departments in seven states, through which a total of 59 high-poverty Title I school districts were recruited for the study. Thus, the formal warrant for the causal effect they reported extends only to these 59 districts. However, based on the authors' careful documentation of sample selection and the theory set forth to account for the observed outcomes, the findings of the study may reasonably be taken as a guide to likely effects of the same or similar treatments in these same districts in future years, in similar districts within these seven states, and elsewhere where sufficiently similar circumstances obtain. This further generalization is usually unavoidable. (We return to the topic of generalization later in this chapter.)

Identifying causal effects. At the heart of GCI is the challenge of disentangling the effect of just one potential cause from all the other influences, known or unknown, observable or not, that might also account for the treatment effect observed. This challenge is referred to as the problem of *identifying* the causal effect. To do so, GCI always relies on some element of *exogeneity* in the assignment of units (persons, students, teachers, classrooms, schools, and so forth) to treatments. That means that there must be some mechanism of assignment that is arguably independent of actions or decisions by the participants themselves.

Experiments: As noted above, the most straightforward way to ensure exogeneity, when feasible, is random assignment to treatment versus control conditions, and this is the defining feature of an experiment. Studies that do not rely upon random assignment of participants to treatment and control groups are not referred to as true experiments, but as quasi-experiments. Random assignment supports a statistical argument that treatment and control groups are equivalent (but for the vagaries of statistical sampling) with respect to all observable and unobservable characteristics. However, there are alternatives, used in quasi-experimental studies.

In the study of the CDDRE model by Slavin and his colleagues, exogeneity was provided by random assignment, with the 59 urban school districts recruited being randomly assigned to treatment and control groups. As stated, clearly and directly: "Pairs of volunteering districts within each state matched on demographic characteristics and prior achievement were assigned at random (by coin flip) to the experimental or control groups …." (Slavin, et al., p. 381). Agreement to participate was secured from all

districts *before* random assignment into treatment versus control conditions (Slavin, et al., 2012, pp. 380-381). (The control districts received the CDDRE intervention a year later.)

The assignment mechanism (a coin flip) is arguably independent of any potential cause, observable or otherwise, that might have given rise to the overall differences later observed between schools in the experimental and control districts. That is what is meant by exogeneity. Note that the units assigned here are districts, and the treatments are simply the experimental and control conditions. Of course, the experimental and control groups will almost certainly differ somewhat, but in the absence of a treatment effect (i.e., an effect attributable to the CDDRE reform model), the outcomes for groups formed by random assignment are said to be equal in expectation.[10] Thus, if after the treatment an outcome difference is observed between treated and untreated groups, and if that difference is larger than would have been likely to have arisen due to random assignment alone, then that observed difference is taken as evidence of a causal effect.

Simple random assignment is perfectly adequate, but greater precision may be realized by using more sophisticated versions of random assignment. For instance, random assignment within blocks or strata may result in treatment and control groups that are more similar on average than would be expected with simple random assignment alone. Slavin and colleagues (2012) used the process of matching districts prior to random assignment to increase the likelihood that treatment and control groups were similar, a particularly important extra step when sample sizes are small. If a study goes beyond the use of simple random assignment, then models for data analysis will probably need to be adjusted accordingly.

Quasi-Experiments with Exogeneity: Random assignment of units to treatments is not the only way to provide exogeneity. Quasi-experiments may capitalize on fortuitous sources of exogeneity to isolate the effect of a given treatment from other possible causes. Different mechanisms for doing so may be referred to as strategies for *identification* of treatment effects. So-called *natural experiments* may induce the application of a "treatment" to some units and not others in a manner arguably

[10] The phrase, "equal in expectation" is used to acknowledge that across groups created by random assignment, outcomes may still differ due to the vagaries of statistical sampling, but no systematic difference among them is expected. Absent treatment effects, on average across many replications, groups formed in this way should have the same outcomes.

independent of any rival explanatory factors that might otherwise account for observed treatment effects. (E.g., lotteries might be used by districts to select students into special schools or programs.) In *regression discontinuity* studies, treatment assignments are determined by a known, objective decision rule that creates an opportunity for comparing groups that are arguably close to identical apart from being assigned to treatment versus control conditions. For example, a policy might be enacted whereby students are tested and those below some cut off score are treated differently from those above that cut off score. It may then be argued that those testing just below versus just above the cut off score are essentially indistinguishable, and that an observed difference between these two groups can therefore be attributed to the different treatments associated with the cut off score classification. Additional statistical methods have been developed, which rely on stronger statistical assumptions.[11] Detailed discussions of these and other identification strategies may be found in Shadish, Cook, and Campbell (2002), in Murnane and Willett (2011), or in numerous other sources. These more complex models for GCI rely on stronger assumptions and so, other things being equal, offer weaker warrants for causal inference than do studies relying on random assignment. However, these methods can often be of great value when used to support causal inferences employing large data sets originally created for other purposes. They often enable causal inference in situations where random assignment is not possible.

Quasi-experiments without Exogeneity. In the absence of random assignment or some other exogenous assignment mechanism, researchers may seek to approximate random assignment by constructing a comparison (control) group as similar as possible to some treatment group by matching on a various observable characteristics of the units included in the two groups. Typically, a control group is assembled after the treatment group has been defined, in such a way as to match the characteristics of the treatment group as closely as possible. Matching may take account of student demographics, prior-year test scores, school climate, or other factors known to be related to the outcome of interest. The hope is that matching on many observable variables will also serve to render the groups comparable with respect to all other observable and unobservable

[11] In an analysis using an *instrumental variable*, for example, some third variable (the *instrument*) is found, which is related to the hypothesized causal variable of interest but is arguably independent of any observable or unobservable influences on the outcome that are not associated with that causal variable.

characteristics that might otherwise explain any difference in outcomes found between the treatment and comparison groups. Because precise matching is rarely possible, the matching strategy is typically supplemented by statistically controlling for these additional factors. So, for example, in comparing classrooms, variables representing students' prior-year test scores, demographic characteristics, or other attributes may be included in a statistical model as covariates. These statistical controls arguably reduce the plausibility of rival explanations for any observed group difference in measured outcomes, but as the GCI tradition has evolved, it has come to regard reliance on such covariance adjustments as at best inferior to exogenous treatment assignment, and ultimately insufficient. Referring to an earlier period of time when these strategies were more widely used, Murnane and Willett comment as follows:

> One common response was to include increasingly larger and richer sets of covariates … in the statistical models that were used to estimate the effect of treatment on outcome. The hope was that the presence of these control predictors would account for differences in the outcome that were due to all of the unobserved—and endogenously generated—differences …. Sociologists Stephen Morgan and Christopher Winship (2007, p. 10) refer to the period in which researchers relied on this strategy as "the age of regression." Seminal studies published in the 1980s threw cold water on this 'control for everything' strategy by demonstrating that regression analyses that contained a very rich set of covariates did *not* reproduce consistently the results of experiments in which individuals were assigned randomly to different experimental conditions. (Murnane & Willett, 2011, pp. 32-33)

Modern refinements of these statistical control strategies are useful for some purposes, but the principle still holds that only random assignment (or some other method of assignment that is arguably exogenous) can be relied upon to assure that, absent a treatment effect, outcomes for the treatment and control groups are equal in expectation. Covariates can play important roles in research studies aimed at GCI. They are often used to improve the accuracy with which treatment effects can be estimated. However, covariance adjustments are *not* relied upon as a substitute for exogenous treatment assignment. By themselves, covariate adjustments cannot offer a rigorous warrant for

inferences as to the particular causes for observed differences among nonequivalent groups.

Carrying out the study. After formation of groups, while the treatment is being implemented, observational methods may be used to monitor the fidelity of treatment implementation and to better understand the effects found. In the CDDRE study, for example, school walk-throughs by CDDRE consultants together with district leaders informed treatment fidelity.[12] The authors reported, "These structured walk-throughs provided insight for both the CDDRE consultants and the district administrators into the quality of instruction, classroom management, motivation, and organization of each school. They examined the implementation of various programs the schools were using and focused on student engagement. In addition to informing CDDRE consultants, these walk-throughs were intended to help district leaders understand the real state of education in their own schools, find out which of the many programs provided to their schools were actually in use, and create a sense of urgency to take action" (Slavin, et al., 2012, p. 378).

Analysis. The basic goals of statistical analysis are (a) to consider whether post treatment differences between groups (sometimes framed in terms of the relationship between the treatment and outcome variables) are larger than would be expected by chance alone and/or (b) to estimate the magnitude of the effect along with the precision (or equivalently, the likely range of error) to be anticipated given the variability in the data and the size of the sample. The analysis can be as simple as comparing the post-treatment means of the treatment and control groups with a statistical test known as a "t-test," which shows the extent to which the *observed* difference between the two groups is larger than would have been likely in the absence of any *actual* difference (often described in terms of statistical significance). Or, the researcher might report the difference between means along with a confidence interval that indicates the range of likely values of that statistic. More complicated designs, like that of Slavin and colleagues (2012), which included student pretest scores as covariates in a multi-level design, require far more complicated analyses, but the basic goal of estimating post treatment differences between groups remains.

---

[12] By influencing district leaders' perceptions and motivating their active involvement, these walk-throughs may also have served as an important part of the reform intervention itself.

The statistical interpretation of findings will be dictated by the models used. Many social scientists are trained in the use of statistical tests of null hypotheses, but it is also common to quantify reported results using effect size statistics, which describe the difference between group means relative to the degree of variability within the groups compared. In other words, an "effect size" expresses the degree of separation between the treatment and control groups on a scale that is independent of the various scales used for different measurements in different studies. (See Olejnik and Algina, 2000, for a helpful discussion of choices among alternative effect size definitions.) In the CDDRE study, findings were reported as effect sizes in part so as to enable comparison of the magnitudes of effects found to those from earlier studies of other data-use interventions.[13] Slavin and colleagues (2012) summarized their analysis and findings as follows:

> "Fifty-nine districts in seven states were randomly assigned to CDDRE or control conditions. A total of 397 elementary and 225 middle schools were followed over a period of up to 4 years. In a district-level hierarchical linear modeling (HLM) analysis controlling for pretests, few important differences on state tests were found 1 and 2 years after CDDRE services began. Positive effects were found on reading outcomes in elementary schools by Year 4." (p. 371)

These findings are elaborated in the body of the report in a series of tables, further described in the text in light of the research question. For instance, Slavin and colleagues (2012, p. 389) present the following table, indicating the effect size for CDDRE for each grade level in each year.

Effect Sizes for Cohorts 1 and 2

	Year 1 N = 33	Year 2 N = 33	Year 3 N = 33	Year 4 N = 20
Fifth-grade reading	+.17	+.11	+.24*	+.50***
Eighth-grade reading	+.17	+.31***	+.05	+.25
Fifth-grade math	+.30**	+.10	+.24*	+.32*
Eighth-grade math	+.09	+.16	+.01	+.31*

*p < .10. **p < .05. ***p < .01.

Slavin et al. (2012) p. 389

[13] The authors usefully distinguish between school-level and student-level effect sizes: "School-level effect sizes were considered educationally important if they were at least +.20 (equivalent to individual-level effect sizes in the range of +.07 to +.10, because school-level standards deviations are typically two to three times lower than individual ones)" (Slavin, et al., p. 384).

For each effect size, the table indicates the extent to which it is "significantly" different from zero, i.e., from what would be expected if there were no difference between the groups. Stated more technically, the year-four fifth grade reading effect size might be interpreted as follows: "If there were no difference in outcomes between the groups, we would expect to see an effect size as large as .50 less than one percent of the time, were the test repeated with equivalent groups." Effect sizes are standardized units that facilitate comparisons of findings across different studies. For purposes of showing the substantive magnitude and importance of treatment effects, they are often more useful than results reported using the original scale of the outcome variable.

Generalization. In the absence of random *sampling*, the formal statistical warrant for the causal claim extends only as far as the set of units randomly assigned to treatment versus control conditions. Stronger reliance must then be placed on substantive, essentially theoretical, arguments for the relevance or applicability of research findings in other settings or with other populations.

In practice, even with random sampling from a well defined population, there is almost always interest in generalization beyond the population sampled. If nothing else, the population sampled is limited by the time when the study was conducted, even though it may be justifiably assumed that similar treatment effects could be expected in the future. If a representative sample can be obtained from a broad population, the statistical warrant will be stronger and the need for theory-based extrapolation beyond that population will be lessened correspondingly. However, regardless of whether representative sampling is possible, the actual use and interpretation of research findings requires *causal generalization* (Shadish, Cook, & Campbell, 2002, pp. 24 ff.) beyond the conclusion reached on narrowly statistical grounds. Thus, the ultimate generalization may be thought of as having two parts. First is a statistical generalization from the units sampled to the population they represent. Second is a substantive (nonstatistical) generalization beyond that population, which is always needed if conclusions are to be extended beyond the originally sampled population. In addition to reaching beyond the population of units actually sampled, these generalizations will typically reach beyond the precise treatments actually employed, beyond the particular outcomes actually administered, and beyond the specific settings of the research study. The need for such

generalization was recognized earlier in Cook and Campbell's (1979) concern for "external validity," and has since received increasing, and more formal, scholarly attention. We return to this issue below and in our section on "Conceptions of Generalization" (CoGs) section.

Reporting and Evaluation

The methods of GCI are designed to support generalizable conclusions about lawlike cause-and-effect relationships involving variables that can defined in a common fashion across a large number of cases. To this end, the logic of GCI dictates decisions concerning study design, instrumentation, analysis, and interpretation. This is a worthy goal, but in itself the abstract, formal theory of GCI will fall short of conveying the meaning, implications, or importance of whatever may be found. Thus, reports of GCI studies will go beyond the bare statistical findings, including description as well as causal claims. Investigators using GCI will often have hypotheses and will offer plausible explanations for causal findings.

Findings are reported dispassionately and objectively, with the aim of conveying as clearly as possible the scientific conclusions that are warranted. Methods are documented to demonstrate to the reader that the research was soundly conducted, i.e., that the stated findings are well warranted. Readers will also rely on the study documentation to inform their judgments as to the generalizability of study findings to other situations. Explanations for causal findings and implications for policy and practice, as well as further research, are likely to be proposed. Nonetheless, although researchers using GCI methods may strive to convey a deep and sensitive understanding of the mechanisms responsible for any causal relationships they are able to establish, any substantive explanations offered are likely to go beyond the purview of the formal GCI methodology itself.

Slavin and colleagues (2012, pp. 390-391) begin their discussion of findings with a brief summary of statistical results. They then offer their own observations about the mechanisms through which the treatment appeared to have functioned: "What the findings imply is that helping school leaders understand student data is helpful but in itself does not produce educationally important gains in achievement. Schools must actually take action to change teaching and learning. … [T]eachers need to be using

effective practices every day, and consultations, benchmark data, and policies encouraging the use of proven replicable strategies may [encourage adoption of] such programs."

Evaluating GCI research. Campbell and Stanley (1963) organized questions of validity, soundness, or rigor in terms of "threats to validity"—a list of possible problems or confounds that could give rise to plausible rival interpretations of study findings. Their language suggested that these threats either could or could not be "ruled out," although in practice the concern was more often one of reducing the plausibility of any given threat than eliminating it entirely. Cook and Campbell (1979) elaborated these lists, but more importantly, they broadened considerations of soundness, rigor or quality under the rubrics of four types of validity. Internal validity addressed the design of the study, illuminating the degree of confidence that could be placed in causal interpretations concerning the participants in the study itself. Statistical conclusion validity addressed the appropriate use and interpretation of statistical methods. Construct validity raised questions as to the adequacy of measures used, especially for outcome variables. Finally, external validity was concerned with generalization beyond the particular study to similar populations or outcomes of interest. Shadish, Cook, and Campbell (2002) continued in this same tradition, going into much greater depth than previous authors in considering criteria for the soundness of nonstatistical warrants for generalization beyond the study itself to other populations, treatments, outcome variables, and settings. Drawing on the work of numerous scholars, they distilled five related principles that scientists rely upon to guide research generalization:

1. *Surface Similarity*. They assess the apparent similarities between study operations and the prototypical characteristics of the target of generalization.
2. *Ruling Out Irrelevancies*. They identify those things that are irrelevant because they do not change a generalization.
3. *Making Discriminations*. They clarify key discriminations that limit generalization.

4. *Interpolation and Extrapolation*. They make interpolations to unsampled values within the range of the sampled instances and, much more difficult, they explore extrapolations beyond the sampled range.
5. *Causal Explanation*. They develop and test explanatory theories about the pattern of effects, causes, and mediational processes that are essential to the transfer of a causal relationship. (Shadish, Cook, and Campbell, 2002, pp. 24-25)

By relying on such principles, users of GCI study findings must arrive at their own grounded judgments as to the applicability or implications of prior research for their own purposes in developing theory or improving educational practice. The key affordance and distinguishing feature of GCI is its provision of powerful design warrants to distinguish the effects of specific causal factors from all possible competing explanations for observed effects.

Ethnography/Ethnographic Research (EthR)

Ethnographic research has developed along distinct lines in anthropology, where it has been at the core of the discipline, and in sociology, where it has played an important role alongside other methodological traditions. Erickson (2011) dates the first use of the term ethnography to the last quarter of the 19th century. Ethnographic research now has a vital presence in multiple disciplines and fields of study including education research and research on teaching. It has been prominently represented in the last two *Handbooks of Research on Teaching (HRT)s,* with chapters by Erickson (1986) and Eisenhart (2001).

Among the difficulties in bounding the term "ethnographic," are differing constructions of the overlap among the terms "ethnographic," "interpretive," and "qualitative". Sometimes "qualitative" is used as a synonym for ethnographic and sometimes as a broader term that subsumes ethnographic research and much more. As the editors of the 2001 *Handbook of Ethnography* note, what makes research *ethnographic* is the "commitment to the first-hand experience and exploration of a particular social or cultural setting on the basis of (though not exclusively by) participant observation" (Atkinson et al, 2001, p. 5). Erickson's (1986) *HRT3* chapter title uses the

term "Qualitative Methods[14]", and yet he names ethnography as an alternative term and, within the first few paragraphs, he signals his preference for the term "interpretive" (p.119) or, more specifically, "interpretive participant observational research." Eisenhart's (2001) *HRT4* chapter, entitled "Changing Conceptions of Culture and Ethnographic Methodology," positions Erickson's (1986) chapter as introducing researchers "to interpretive scholarship, ethnographic methodology, and their potential for educational research" (p. 209). Consistent with our earlier description of interpretive social science as focused on understanding meaning in context, the use of the term "interpretive" signals the incorporation of these understandings into ethnographic practice.

Erickson (1986, 1998, 2011) and Eisenhart (1988, 2001, 2009) will serve as our key theoretical resources, supplemented by Hammersley and Atkinson (2007), who provide a introductory textbook consistent with an interpretive approach to ethnography; and we will use elements from *Reform as Learning: School Reform, Organizational Culture, and Community Politics in San Diego* by Hubbard, Mehan, and Stein (2006) to illustrate interpretive ethnographic practice. In the available space, we can't possibly do justice to Hubbard et al.'s presentation of their research. We have made choices to illustrate the methodology, but they reflect only a small portion of what can be learned from this text.

Aims and Conceptions of Social Phenomena

The approach to ethnographic research Erickson and Eisenhart call "interpretive" focuses on understanding the "immediate and local meanings of actions, as defined from the actors' point of view" (Erickson, 1986, p. 120) and the ways in which those meanings-in-action shape and are shaped by (co-constitute) the social contexts of which they are a part. As Eisenhart notes, attention to interpretive scholarship represented a distinctive turn in ethnographic research and in the conception of culture that had been its traditional focus. Ethnographic methods "were originally developed as the means to study 'culture,' defined as the lifestyle of a social group with clear boundaries and distinctive behaviors and beliefs" (p. 210). The interpretive turn in ethnography, made

[14] Erickson (1986) notes this incorporates "approaches to research on teaching that are alternatively called ethnographic, qualitative, participant observational, case study, symbolic interactionist, phenomenological, constructivist, or interpretive", which share a "strong family resemblance" (p. 119).

prominent by Geertz's (1973) *Interpretation of Cultures,* shifted the focus away from behavior toward *action*, where "meaning [is] actively appropriated, constructed, and manipulated in specific contexts" (p. 211).[15]

The distinction between behavior and *action*—which Geertz (1973) famously illustrated by contrasting a blink (behavior) with a wink (action)--is key to *interpretive* ethnographic research. The action entails the meaning given to the behavior by the actor *and those with whom the actor is engaged in interaction* (Erickson 1986 pp. 125-126). As this example implies, meanings-in-action are seen to have both local and non-local origins. Locally, they are embedded in immediate, moment-to-moment interaction among individuals whose understandings can shift as the interaction unfolds and in the "microculture" within regularly interacting groups of individuals as meanings come to be shared and patterns of action and interaction develop (Erickson, 1986, pp. 128-129). Non-locally, these meanings-in-action are constrained by larger social structures and cultural understandings that individuals in a group inherit, orient to (or not), and can participate in transforming over time. These include "learned and shared standards for perceiving, believing, acting, and evaluating the actions of others….and the perceptions that local members have of interests or constraints in the world beyond…their face-to-face relations" (Erickson, 1986, p. 129). Thus ethnographic researchers conceptualize and study social phenomena as meaningful to the actors involved, situated in particular social contexts, and continually evolving. They study the "relationships between actively constructed meaning systems (cultures) and externally imposed conditions (structures)" (Eisenhart, 2001, p. 212).

The researcher's task in ethnographic research is often framed as one of "*translation*" from one culture to another (Marcus, 1998 in Mitchell, 2007), recognizing that there is no single *correct* translation, given differences in culture and individual perspective on all sides. As Eisenhart describes it: "Accounts of what ethnographers learn from studying others are, as Geertz put it, "our own constructions of other people's constructions of what they and their compatriots are up to" (I973b, p. 9)." (Eisenhart, 2001, p. 219). This distinction between the points of view of participants or local actors

---

[15] Erickson (1986) credits Malinowski (1922) as the first ethnography to foreground meaning from the natives point of view (p. 45).

on the one hand and researchers on the other is often described as *emic* versus *etic* (borrowing analogically from linguistics), as experience-near versus experience-distant, or as first order versus second order constructs (Erickson, 2011; Geertz, 1973, 1983). Thus, "The accounts these ethnographers produce attempt to represent the meanings (symbolism) of acts, as they are understood by participants" (Eisenhart, 2001, p 219). However, just like the subjects whose (local) meanings they are representing

> These researchers consider themselves active, reflective subjects, who produce the images, concepts, and understanding represented in ethnographic accounts, based on firsthand knowledge of others and deliberate scrutiny of one's own viewpoint in light of others. (Eisenhart, 2001, p. 219)

A primary goal of ethnography is to represent the emic perspectives of those being studied, through the perspective of the ethnographer, in terms that the readers of the ethnography can understand. We will return to this issue, foregrounding controversies surrounding the respective roles of researchers and researched, later in the chapter.

The word "ethnography", as in "doing an ethnography," has typically been reserved for studies involving long-term participant observation and has tended to focus on a single site or setting. Participant observation refers to the on-site work of ethnographers as they (learn to) interact with participants in locally meaningful and socially appropriate ways and then record those interactions. However, interpretive ethnographic research—with its goals of understanding human meaning in social context and "'translation' of meaning from one culture to another" (Marcus, 1998, in Mitchell)--encompasses a broader set of research goals and types of studies. These include studies that focus on particular aspects of a group's interactions, perhaps over shorter time frames of study, and that trace individuals and cultural products across different groups and contexts of action.

As Eisenhart (2001) notes "it is no longer straightforward for anthropologists to plan to study 'cultural groups'" (p. 213) that are clearly bounded. Individuals participate in varied social settings that shape their identities and "widespread access to transportation, the mass media, and computer technology" further diversifies the social contexts in which they participate. This suggests the importance of attending to culture both within particular social groups (in intersubjective meanings and patterns of action),

and in the identities of individuals who regularly participate in multiple social contexts. Ethnographers, she argues "must find ways to learn about cultural forms and structural constraints that materially and symbolically organize people in and across times and spaces, as well as what can be 'seen' and experienced by a positioned researcher-actor." (Eisenhart, 2001, p.218). For ethnographic researchers of teaching, this means finding ways to get information about students' and teachers' interactions outside as well as inside school and to understand how the two are dynamically related.

Hubbard et al. provide a rich example of interpretive ethnographic research. They traced the design and enactment of a centralized reform effort in the San Diego City Schools (SDCS) intended to "improve student achievement by supporting teaching and learning in the classroom" (p. 63) from 1998-2002. They focused on the interactions among and between groups of stakeholders at different levels of the system as they encountered, made sense of, and enacted elements of the reform as they understood it, including the ways in which they used information relevant to teaching and learning. The question the book ultimately addresses is "why did a reform that began with so much promise and celebrated leadership not materialize to the degree envisioned?" (p. 4).

While the reform had multiple dimensions, we'll focus on one key element that Hubbard et al. also foreground, "Balanced Literacy", where data use relevant to teaching and learning played a key role. Balanced Literacy was an approach to teaching and learning that had been successfully implemented in New York City's Community District #2, under the leadership of Anthony Alverado, who was subsequently recruited to lead the SDCS reform as chancellor of instruction by its new superintendent, Alan Bersin.

> A Balanced literacy approach requires teachers to determine each student's capacity to read and understand text ["through a variety of means, including reading journals, conferring notes, and formal assessments" (p. 107], design instructional tasks that will challenge each student to reach the next level of reading," and provide just the right kinds and appropriate amounts of assistance as the student tackles that task (p. 90).

A great deal of professional development for teachers and their leaders focused on "preparing teachers to assess students and use that information to plan instruction" (p.

107). And that, in turn, involved data use on how teachers were teaching, principals were leading, and so on.

Hubbard et al. studied the SDCS as a set of nested learning communities, focusing in particular on what they called "intersection encounters," where educators working at different levels of the system interacted with one another to support individual and organizational learning. Thus Hubbard et al. traced the enactment of the reform as it was handed off from district leaders to instructional leaders, from instructional leaders to school based principals and coaches, from principals and coaches to teachers, and from teachers to students. Consistent with the interpretive ethnographic perspectives described by Erickson and Eisenhart, they viewed all the participants as active agents who *co-constructed* the reform, grappling with what the reform meant as they tried to incorporate their understandings of it into their routine practice (or not). "In the act of trying to balance new demands and established routines, educators transform polices and the innovations that embody them" (p. 9). Hubbard et al. noted "every school has its own culture," shaped by everyday routines, by individuals' (sometimes conflicting) values, beliefs, and assumptions, and by political forces in the surrounding environment. Understanding how the local culture shaped the ways the reform was (differently) understood and taken up was an important goal of the study. As they note: "Organizations are messy. Therefore, when change is introduced in any one part of the system, it reverberates throughout the system in ways that cannot be anticipated" (p. 8). We'll illustrate their methods and the ways in which they describe and theorize their learnings, in the sections below.

Returning to the more general aims and understandings of social phenomena in ethnographic research, Eisenhart (2001) describes the ways in which ethnographers, informed by alternative theoretical perspectives, have developed alternative directions within the field (see also Erickson, 2011). Critical theorists from a Marxist perspective call for a focus on the way culture "camouflages or 'mystefies' social inequities in a class based society" (p. 212). They frequently seek to "empower participants to take greater charge of their own lives" (p. 219). Feminist and ethnic studies scholars call for ethnography to address concerns about the ways in which conventional studies of "culture" can essentialize and stereotype, ignoring within group variations and leaving

little room for understanding agency: "not all people of color, all women, all men, or all people socially identified as members of any group have the same histories, experience the world in the same way, face the same problems, or construct the same meanings" (p. 214). Scholars whose work can be located within the broad constellation of perspectives that have been labeled postmodern call for illumination of the ways in which any given representation of "reality" filters or screens ways of seeing the world. Feminist, postmodern, and ethnic studies scholars raise concerns about the one-sidedness of ethnographic accounts and the issues of power and perspective entailed. These concerns point to more collaborative relationships between researcher and researched; and to strategies of representation that allow multiple interpretive voices to be heard, to address issues of perspective and power. Hubbard et al. take up some of these concerns explicitly. They routinely shared their plans and evolving learnings with participants (p. 25). Their questions attended to multiple perspectives including "competing and conflicting" understandings and enactments of the reform. They also engaged in occasional enactments of "design research", for instance, where one of the authors and a school principal collaborated to support the principal's coaching of teachers to enact the Balanced Literacy Reform, with care in documenting the researcher's role alongside those of participants.

Design and Enactment

In ethnographic research, design and enactment of the research are engaged iteratively: The development of research questions, site selection, data collection/production, analysis, and even write-up can and do evolve iteratively throughout an ethnographic study. As Eisenhart (1988) describes it, "The collection of new material and subsequent analysis may raise new research questions or lead to insights that become incorporated into, or sometimes radically redirect, the study itself as well as later data collection and analysis procedures" (Eisenhart, 1988, p. 107). Erickson (1998) goes so far as to suggest he is suspicious if researchers' questions have not evolved, as this raises questions about the extent to which they have allowed their preconceptions to be challenged by what they are learning about the site (p. 1167). We can see this sort of evolution operating in Hubbard et al.. For instance, the issue Hubbard et al. use to frame their book—"why… a reform that began with so much promise and

celebrated leadership [did] not materialize to the degree envisioned" (p. 4) is not and could not have been the issue orienting their initial design.

Site selection and entry. Neither Erickson nor Eisenhart say much about the initial selection of the site(s), individual(s), or cultural product(s) that provide the focus for the study. This may be, in part, an artifact of selection being based on intrinsic interest in the site and the anticipation that research questions will evolve as researchers become more familiar with the context of the research, and thereby able to ask more specific questions. As Hammersley and Atkinson note, "sometimes the setting itself comes first—an opportunity arises to investigate an interesting situation or group of people; and…problems spring from the nature of that setting" (pp. 28-29). But even where the setting is selected based on a "foreshadowed problem" in which researchers are interested, the "nature of the setting may still shape the research questions" (pp. 28-29). Hammersley and Atkinson advise researchers to "case" possible research sites, to consider their suitability and feasibility for study, including the sorts of access that the researcher might accomplish. The researchers' entry into the site, including establishing relationships of trust with the local actors, is seen as crucial for a successful launch of an ethnographic project. Hubbard et al. spent six months exploring the reform implementation before developing their initial research questions and strategies for data collection to share with key stakeholders.

Fieldwork refers to the frequently extended part of the research cycle during which the researcher is in the field--engaging with actors in the settings where the research is taking place—and making or obtaining records from data sources. Fieldwork includes on-going analysis to support researchers in making a number of sampling decisions: in deciding what to observe (when, how often, and for how long), with whom to talk, what documents or other artifacts to review, and so on. Fieldwork has been thought by some to be radically inductive, but both Erickson and Eisenhart push back against this notion, in favor of making fieldwork as deliberate as possible. While categories for observation are not known in advance, researchers enter a site with conceptual issues of research interest, foreshadowed problems (Hammersley & Atkinson, 2007), and frames of interpretation or theories that inform their work and that they seek to develop or modify. "In fieldwork the induction and deduction are in constant

dialogue…. The central issue of method is to bring research questions and data collection into a consistent relationship, albeit an evolving one" (Erickson, 1986, p. 121, 140).

Data collection. Common methods of data collection or production[16] include observation with varying degrees of participation, interviewing, and document collection, and researchers' introspection (Eisenhart, 1988). As our theorists note, other methods of data collection, from other methodologies, can and do supplement ethnographic methods, an issue we'll return to in our section on methodological pluralism.

Participant observation requires the researcher to learn how to act and interact appropriately in the local context as participants go about their daily lives and to record those experiences. As Eisenhart describes it, participant observation:

> is a kind of schizophrenic activity in which, on the one hand, the researcher tries to learn to be a member of the group by becoming part of it and, on the other, tries to look on the scene as an outsider in order to gain a perspective not ordinarily held by someone who is a participant only. (Eisenhart, 1988, p. 105)

Erickson (1986) describes three issues as crucial at the outset:

> (a) identifying the full range of variation in modes of formal and informal social organization … and meaning-perspectives; (b) collecting recurrent instances of events across a wide range of events in the setting, so that the typicality or atypicality of certain events types with their attendant characteristic social organization can be later established; and (c) looking at events occurring at any system level (e.g., the classroom, the school, the reading group) in the context of events occurring at the next higher and next lower system levels. (p. 143)

Eisenhart (2001) notes, however, that newer multi-sited ethnographies must depend on alternative methods of data collection allowing ethnographers to explore structure and meaning across time and space in ways that don't depend upon first hand observation (p. 218). Interviews and informal conversations with participants give researchers access to participants' perspectives, however all our key theorists highlight the importance of understanding and analyzing the interview as another form of social

[16] While all of our key theorists use the term "data collection", others worry about the implication that data are there to be found and collected, rather than being co-constructed by researchers who filter data sources through their interpretive frames.

interaction, co-constituted by interviewer and interviewee, and *not* as allowing direct access to participants' perspectives. Observations and interviews can be recorded in detailed fieldnotes, in audio-visual recordings, or both. The construction of fieldnotes is a book-length topic in its own right (Emerson, Fretz, and Shaw, 2011) and there are substantial variations in the ways researchers inscribe their experiences. The goal with fieldnotes and audio-visual recordings is to create a sufficiently robust record to support analyses from various perspectives as research questions evolve. Researcher introspection refers to the ways in which researchers' own interpretive frames, their participation in the setting, and their reactions to their on-going experiences are shaping their learnings. As our theorists note, other methods of data collection, including surveys and design experiments can supplement conventional ethnographic methods.

Hubbard et al. engaged in all of these data collection activities across the four years of their study beginning with a six month period of observation where they learned enough about the reform to develop their initial plans for data collection. Their data collection required them to make a number of sampling decisions as the study unfolded: selection of case study schools within the district, of teachers and other staff to interview, of classroom lessons to observe and so on. Consistent with advice of our key theorists, they wanted to maximize variations in the perspective on and enactment of reform that were available to them. In terms of observations, they chose to focus on what they called "intersection encounters" where participants from one level of the education system interacted with participants from another, so they could trace the implementation of the reform over time. They observed repeated instances of classroom lessons, staff development sessions, principals' conferences with instructional leaders, principals' observations of and interactions with teachers ("walk throughs"), professional development meetings with instructional leaders, and various community interactions. They engaged in interviews with the superintendent and chancellor of instructor, senior district staff, instructional leaders, principals, teachers, and community members. They routinely collected artifacts along the way, beginning with the "Blueprint for Student Success" that articulated the theory of action espoused by the superintendent and chancellor of instruction. And they theorized their own roles as the research unfolded.

Data analysis: The distinction between data analysis and collection is somewhat arbitrary since both proceed simultaneously throughout much of the research cycle, although both typically become increasingly focused as the study evolves and more specific research questions can be asked. As our key theorists present it, there are no formulas or recipes for data analysis although there are practices that have grown up over time that can be represented as general themes in ethnographic analysis. Two general approaches to analysis that our key theorists consistently cite are associated with the concepts of "analytic induction" (Becker,1998; Hammersley, Gomm & Foster, 2000; Znaniecki,1934) and Grounded Theory's "constant comparison" (Glaser and Strauss, 1967). While units of analysis vary, they often involve emically bounded events or situations in which participants are acting and interacting. Eisenhart (1988) describes a general process of analysis consistent with "constant comparison".

> Generally, the procedures involve defining "meaningful" units of the material (meaningful to participant or researcher) and comparing units to other units. Like units are grouped together into categories. Categories are compared to all other categories and relationships between them posited. Categories and the relationships among them are considered and reconsidered in light of old material and as new material is gathered" (Eisenhart, 1988, p. 107).

Erickson (1986) foregrounds a process of analytic induction. Analytic induction involves "reviewing the data corpus repeatedly to test the validity of assertions that were generated, seeking disconfirming evidence as well as confirming evidence" (p. 146) and revising the assertions accordingly. Erickson notes that the "deliberate search for disconfirming evidence is essential" (p. 147). The researcher is looking for patterns of generalization within the case at hand or "internal generalization". Evolving interpretive frames or theory play a role in both approaches—as Erickson noted, induction and deduction are in constant dialogue--and will be addressed in illustrations based on Hubbard et al.. In their book-length introductory text, Hammersley & Atkinson provide a more detailed set of potential analytic strategies.[17]

---

[17] Eisenhart (personal communication, 8/6/13) recommends readers consult Spradley (1979, 1980) for a more cognitive structure approach to data analysis and Agar (1996) who recommends that researchers iteratively identify "rich points" that reflect happenings they find surprising, and "frames" that explain these rich points.

Hubbard et al. are not fully explicit about their analytic methodology although readers can draw some relatively straightforward inferences based on the way they present their learnings. For instance, for each type of "intersection encounter," they provide examples of enactments that are consistent with the intent of the reform and then a range of enactments that are not (distinguishing between staff who were faithfully trying to enact the reform and those who were actively resisting it, based on what they learned from interviews). They also trace elements of the reform across levels of the system, for instance, how the professional development for principals was taken up (or not) in their interactions with teachers or how the professional development for teachers was taken up (or not) in their interactions with students. Thus the sorts of categories supporting their analysis and the relationships among them become clear as one reads through their presentation of their learnings.

Reporting and Evaluation

As Hammersley and Atkinson (2007) suggest, "There are always many ways we could write about the social phenomena with which our research is concerned" (p. 191). Writing is not a "transparent medium of communication," (p. 191); rather, just like data collection and analysis, it is another way in which "researchers shape the phenomena that they study" (p. 191) as they transform their "experience of the social world into a social science text" (p. 191). In fact they see writing as closely related to analysis. Further, it is in the process of writing up the research that the final research questions are often formed in light of the researchers' specific experiences and challenges to evolving interpretive frames. The goal of the write up is to "demonstrate relationships between data and concept or theory" (p. 193). Erickson (1986) suggests a general set of questions a report might address: what actually happens in a particular site at a particular time, what these happenings mean to the actors involved, how these happenings are organized in patterns over time, how these patterns relate to what is happening elsewhere (e.g., at other levels of a school system), and how the patterns in this site compare to those in other sites at other times. And Hubbard et al. address each of these general questions.

No single report can describe all the researchers' experiences. "The aim is to persuade the audience that an adequate evidentiary warrant exists for the assertions made, that patterns of generalizations within the data set are indeed as the researcher claims they

are" (Erickson, 1986, p. 149). All our theorists describe research reports as moving back and forth between concrete details of particular happenings or dialogue to more general descriptions of the patterns that the concrete examples illustrate. Erickson advises that a report of interpretive ethnographic research include:

- the researchers' assertions, claims, or conclusions describing observed patterns and exceptions;
- concrete examples or "particular descriptions"—for example, narrative vignettes of events and quotations from fieldnotes, interviews, or documents--to illustrate the patterns,
- more "general descriptions" of the breadth of evidence, including the typicality or atypicality of the observations reported,
- interpretive commentary of various sorts, which serves at least three roles:
    - it situates the concrete examples, drawing readers' attention to what it is an example of;
    - it explains the significance of the patterns observed, connecting them to more general theory developed within and beyond the study; and
    - it shows how the researchers' preconceptions evolved throughout the study; and
- a description of the history of the inquiry.

Again, one can see all these features reflected throughout Hubbard et al.'s book, beginning with the early history of the inquiry. The opening chapters describe the theoretical frameworks on which they were drawing as they undertook their research (including how they were conceptualizing reform as "learning") and relevant research from elsewhere, especially NYC District #2 where Alverado had led a similar effort; their professional backgrounds and commitments, including their on-going relationships with the SDCS and with NYC District #2; and their general ethnographic methodology and data sources (including numbers of observations by type, interviews by role, etc.). They then draw on written documentation and interviews with SDCS leaders to unpack the leaders' plans and expectations for the reform, which serves as a point of comparison in the chapters focused on enactment of the reform.

Hubbard et al.'s presentation of their learnings about the reform's enactment is then divided into three chapters, each focusing in detail on a different intersection encounter: teachers with students, principals and coaches with teachers, leaders with principals and coaches. For instance, in the chapter on teachers' interactions with students in the classroom, Hubbard et al. outlined the intended features of a balanced literacy reading workshop lesson, showed concrete illustrations—narrated events and quoted dialogue—in which teachers taught the lessons, and described general patterns in the teachers' and students' interactions. The goal was to support students in moving toward independence in their reading. We focus here on their report of teachers' enactment of conferring with individual students during independent reading time where a key goal is to "gather rich information about students' current strengths and weaknesses" to allow timely intervention to support students' development. The concrete examples focus first on exemplary enactments of this practice and then on enactments that deviate in various ways. In their presentation, they distinguish between structural and interactional features of the intended curriculum. They observed multiple enactments of *structural* features—physical arrangements and distributions of time—consistent with the reform's intent, although they did note some deviations in time distribution, with some teachers devoting more time to the didactic components and less time to the interactive components of the lesson. Deviations with respect to the *interactional* features of the lesson were considerably more prevalent:

> Patterns of deviation…occurred in the questioning strategies teachers used. The framework encouraged teachers to devise increasingly difficult questions with several possible answers, eventually requiring students to make sophisticated judgments and proffer varying points of view. The majority of questions we heard posed, however, remained at the level of simple comprehension. Similarly, teachers often supplied the answers to the questions they asked, rather than waiting for students to answer. …. Teachers also tended to use questions that were either too specific, and thus did not encourage conversation, or too vague, and thus did not promote students' use of complex comprehension strategies. (p. 112)

Hubbard et al. speculated about these differences in terms of the improvisational nature of these interactions: "teachers must construct them, including questioning strategies, in the moment responding to the information they receive from students as they confer" (pp. 115-116). Tying these observed patterns to what they had learned from other data sources in other contexts (including knowledge of the previous curriculum) and to their theories of professional and organizational learning, they speculated that teachers "interpreted reform components through the prism of already existing norms and routines" (p. 109).

This pattern of reporting successful enactments and deviations is repeated in subsequent chapters focused on encounters at other levels of the system. The subsequent chapters allow readers to trace the ways in which the learning opportunities for teachers were shaped by the learning opportunities for others in the system. For instance, following their presentation of evidence on interactional encounters between teachers and principals, they noted striking parallels between teachers' and principals' struggles: "School leaders tended to have difficulty with—or skip—the part of their leadership that involved interacting with teachers as learner: diagnosing their (changing) level of expertise, negotiating goals for further skill acquisition, and planning appropriate types and amounts of assistance to help teachers reach the agreed-on goals" (p. 130).

Drawing together all the evidence presented in light of their theoretical framework, their initial conclusions address their orienting question about why the reform did not materialize as anticipated and as it had evolved, more successfully, in NYC District 2.

> Stepping back from our analyses we can see how the intent of the reform was altered as it was handed over from district leaders to ILs, from ILs to principals, from principals to teachers, and from teachers to their students. A reform that began as conceptually driven was proceduralized; an approach to learning that began as student centered became teacher-centered; and a framework with many openings for the application of professional judgment became understood as scripted. We think the explanation for these transformations lies in the overlapping and sometimes contradictory relationships among several factors: the challenging nature of the reform; the capacity, norms, values and standard

> operating procedures of the district…; and the exceedingly rapid pace with which all principals, teachers, and students were brought into the reform. (pp. 239-240)

They then turn to a discussion of the relevance of their learnings for organizational theory and for leaders of other reform efforts (which we describe in the section on generalization).

While we have chosen a book-length report to represent an ethnography, there are many examples of ethnographic research that are presented in articles (and we will refer to others when we overview the literature relevant to our theme of data use later in the chapter). We should also note that evolving visions of ethnography foreground alternative relationships between researcher and researched and alternative textual forms that represent responses to concerns about power and perspective in conventional ethnographic work. We will provide examples in subsequent sections on design based research, and participatory action research.

Validation and Generalization: While both Eisenhart and Erickson use the term validity to characterize the soundness of ethnographic research, not all ethnographic researchers do. Some (e.g., Wolcott, 1994) see it as too closely tied to research in the tradition of the natural sciences and to methodologies that privilege general laws. The approach to validity Eisenhart and Erickson describe considers "the 'reality' of others' worlds to be the constraint which checks bias and assures social science" (Eisenhart, 2001, p. 219) or "the immediate and local meanings of actions, as defined from the actors point of view" (Erickson, 1986, p. 119) as the criterion on which validity judgments are based. Validity issues and strategies for addressing them are relevant throughout the research process. They begin with site selection and entry, including establishing of trusting relationships necessary to for researchers' learning, and continue through the writing-up of the research report, which should allow readers to evaluate the validity of researchers' conclusions for themselves. Strategies for enhancing validity mentioned by our key theorists include:

- long term involvement in the site and the development of trusting relationships with local actors;

- triangulation or comparison across data sources and methods, researchers' and local actors' perspectives, and even theoretical frameworks, both to establish coherence and consensus *and* to illuminate differences;
- the ongoing search for disconfirming as well as confirming evidence and the tryout of alternative interpretations that might explain the patterns observed, with assertions evolving accordingly;
- "member checks" which involve local actors in the review and/or generation of interpretations of what was observed;
- sufficient evidence to warrant key assertions;
- documentation of evolution in the researchers' perspectives in response to experiences in the site;
- audit of data sources and analyses leading to conclusions by other researchers; and
- a write-up that provides sufficient evidence for readers to serve as co-analysts.

For ethnographic researchers who are concerned about issues of power and perspective, some of these validity strategies are problematic, because they privilege the researchers' perspectives and risk misrepresenting a complex and multi-voiced social reality. Some researchers drawing on ethnographic methods address these concerns in more transgressive and performative ways. Lather and Smithies (1997), for instance, in their study of a support group for women with HIV/AIDS, resist the construction of coherent narratives. They maintain visually distinct streams of text throughout the pages of their book, presenting not just participants' distinct voices (quoted uninterpreted except for the interpretation entailed in the selection), but also their own voices, and voices from other publications. Lather (2007) theorizes her approach in *Getting Lost: Feminist Efforts Toward a Double(d) Science*.

As we noted earlier, generalization internal to the case—a warrant that the patterns described account for all the available data —is central to the validity of ethnography for our key theorists. They each also consider issues of external generalization or relevance of the ethnography beyond the particulars of the case at hand. While these warrants cannot depend on random sampling, they do incorporate some of the same strategies named by Shadish, Cook, and Campbell (2002) in their section on

external validity with non-random samples, and more. Conceptions of generalization offered by Erickson (1986) and Eisenhart (2009) include the following and research reports can and do draw on multiple approaches:

- readers' generalizations, where researchers support readers in drawing their own conclusions about generalization or transfer of learnings to their own contexts, which requires researchers to provide sufficient detail about the studied context to support such comparisons;
- researchers' logical generalizations (of the sort Shadish, Cook, and Campbell, 2002 proposed) that build the argument for generalization based on the similarity of the studied case to either another case of interest or to a set of similar cases (based on evidence of similarity);
- research syntheses or reviews of relevant literature, which support both logical generalizations (above) and theoretical generalizations (below);
- theoretical generalizations which show how evidence from a particular case contributes to an evolving theory that draws on multiple cases and types of studies.

Eisenhart (2009) provides an accessible overview. She pays particular attention to theoretical generalization, citing a classic quote from Geertz:

> "What generality it contrives to achieve grows out of the delicacy of its distinctions not the sweep of its abstractions…. Studies do build on other studies, not in the sense that they take up where the others leave off, but in the sense that, better informed and better conceptualized, they plunge more deeply into the same things…. One can, and this in fact is how the field progresses, conceptually, take a line of theoretical attack developed in connection with one exercise in ethnographic interpretation and employ it in another, pushing it forward to greater precision and broader relevance….The theoretical framework in terms of which such an interpretation is made must be capable of continuing to yield defensible interpretations as new phenomena swim into view" (Geertz, in Eisenhart, p. 62)

We will have more to say about generalizations that rely on theory in our section below on "Conceptions of Generalization".

Hubbard et al. arguably engage in all these approaches to support the relevance of their research beyond the particulars of the case studied, with rich details of the

circumstances surrounding the enactment of the SDCS reform, comparisons to other reform efforts, including an extended comparison with NYC District #2, comparisons with other studies of the SDCS reform by other teams of researchers (including an instructive comparison of diverging interpretations of the same student achievement data), and interpretations of the ways in which the learnings from this case reflect and contribute to organizational and learning theory. They also raise instructive cautions about the risks of over generalizing, as reformers did from the successful enactment of a similar reform effort elsewhere:

> not only do reforms differ from place to place, they also differ from time period to time period. What is understood as constituting the best approach to teaching and learning changes over time, as does consensus over what the goal of K-12 education is or should be…. It is important to be sensitive to both the temporal and the spatial dimensions of context. (p. 253)

Small N or Comparative Case Studies (CCS)

We use the label "Small N or Comparative Case Studies" (CCS) to describe a somewhat heterogeneous set of methodologies, in some sense intermediate between large-N quantitative studies and analyses of single cases. Despite the label "Small N," as further discussed below, it would be a serious mistake to infer that this methodology can be distinguished merely on the basis of sample size. Research studies in this category share some general features, but there is no simple list of essential characteristics that distinguishes them from studies employing other methodologies. CCS employ more than one case, but typically fewer than fifty, purposively chosen so as to illuminate the question or phenomenon of interest.[18] Typically, cases are chosen so as to contrast with respect to some set of key features. In CCS, within-case analyses are supplemented by cross-case comparisons, which help to support generalization.

There are various approaches to CCS across the social science disciplines. Significant theoretical development has occurred within the discipline of political science

[18] As discussed below under "Design and Enactment," cases may be defined and characterized in many ways, and conceptions of cases may evolve as the research proceeds. For present purposes, as a rough and ready notion, cases may be thought of simply as what is being observed, investigated, or compared and contrasted.

and in the area of policy analysis. From within these perspectives, George and Bennett (2005)[19] and Byrne and Ragin (2009) will serve as our key theoretical resources for this tradition. Our illustrative case for this methodology is a study by Coburn, Russell, Kaufman, and Stein (2012) of twelve teachers in a single school district, investigating the configurations of factors that distinguished those who were versus those who were not able to sustain new mathematics instructional practices after district supports for reforms were withdrawn. Their study illustrates the CCS method of Qualitative Comparative Analysis (QCA), as explained later in this section. We note that together with QCA, this example draws on (a qualitative approach to) social network analysis.

George and Bennett (2005, pp. 18-19) offer a useful perspective on comparative case studies, which they term, "the comparative method," as follows:

> In one view, the comparative method (the use of comparisons among a small number of cases) is distinct from the case study method, which in this view involves the internal examination of single cases. However, we define case study methods to include both within-case analysis of single cases and comparisons of a small number of cases, since there is a growing consensus that the strongest means of drawing inferences from case studies is the use of a combination of within-case analysis and cross-case comparisons within a single study or research program (although single-case studies can also play a role in theory development). (George & Bennett, 2005, p. 18)

(Note that we generally avoid the phrase "case study methods," simply because it has been used in so many different ways across various methodological traditions.)

Aims and Conceptions of Social Phenomena

Researchers in this methodological tradition strive for an in-depth understanding of the particularities of each case chosen for study and then proceed to generalizations, across studied cases and beyond, within the framework of a configurational, as opposed to a variable-based, conception of social phenomena. A "configurational conception" may refer to sets of attributes or conditions that are either present or absent for each case, various combinations of which may be necessary and/or sufficient to bring about some

---

[19] An instructive contrast to George and Bennett (2005) within the discipline of political science can be found in Flyvbjerg (2001), *"Making Social Science Matter: Why Social Inquiry Fails and How it Can Succeed Again"* (for examples, see Flyvbjerg, Landman, and Schram, 2012).

outcome of interest. In elaborations of CCS methodology, these attributes or conditions may be present to varying degrees. Ragin (1992, p. 5) distinguishes this case-centered orientation from variable-centered approaches, explaining that:

> [variable-centered] investigators begin by defining the problem in a way that allows examination of many cases (conceived as empirical units or observations); then they specify the relevant variables, matched to theoretical concepts; … the language of variables and the relations among them dominate the research process. ... The alternative, case-oriented approach places cases, not variables, center stage.

CCS methods may sometimes be chosen because the phenomena of interest are infrequent, and of interest "precisely because the N of cases is small" (Ragin, 2007, p. 68). In other words, cases with rare but important (either positive, perhaps exemplary, or negative, perhaps catastrophic) outcomes may be examined to better understand the configuration of attributes that led to those outcomes. Variables may be defined in the process of this work, and used to characterize significant attributes of cases. The goal of the investigation, however, is to characterize sets of cases sharing particular configurations of attributes, not to describe relationships or dependencies among the variables themselves. Whereas GCI seeks to understand relationships among small sets of variables chosen during the planning stages of the study, the CCS investigator seeks to describe the constellations of conditions under which some outcome may be expected:

> Case study researchers are more interested in finding the conditions under which specified outcomes occur, and the mechanisms through which they occur, rather than uncovering the frequency with which those conditions and their outcomes arise. (George & Bennett, 2005, p. 31)

CCS researchers typically acknowledge multiple contributing factors leading to a given outcome, and allow for the possibilities that more than one configuration of conditions may lead to the same outcome and that the same configuration of conditions may lead to different outcomes. As stated by George and Bennett,

> Case studies are much stronger at identifying the scope conditions of theories and assessing arguments about causal necessity or sufficiency in particular cases than

> they are at estimating the generalized causal effects or causal weight of variables across a range of cases. (George & Bennett, 2005, p. 25)

CCS may be used for theory building (as with typological theorizing, described below), and also to test or examine prior theories and assumptions. Provisional theories may be developed and then tested against additional, purposively chosen cases as the research proceeds.

The "degrees of freedom problem." Our theorists speak directly to concerns likely to be raised by researchers in GCI and survey traditions about the number of variables exceeding the number of cases, the so-called "degrees of freedom" problem, characterized by George and Bennett (2005, p. 28) as a "misapplication of a statistical version of underdetermination." Unlike studies in some other methodological traditions, which may rely on the logic of statistical inference from a sample to a population, CCS research grounds generalizations on internal analyses of several cases, followed by an examination of commonalities and differences in the patterns discovered. Indeed, the number of variables considered in CCS may easily exceed the number of cases. As explained by George and Bennett (2005, p. 25):

> In statistical methods—we focus for purposes of illustration on the example of multiple regression analysis—the term "degrees of freedom" refers to the number of observations minus the number of estimated parameters or characteristics of the population being studied (such as mean or variance). In a multiple regression analysis, the number of observations is taken as the number of cases (or the sample size) and the number of parameters is the number of independent variables and one additional parameter for the value of the intercept (the point at which the estimated regression line intercepts the axis on a graph). Thus, a study with 100 cases and 6 variables would have 100 – (6 + 1) or 93 degrees of freedom.

George and Bennett go on to explain that the apparent "degrees of freedom problem" is the result of a naïve mapping from the problem of statistical inference to problems of inference in case study research. In support of this argument, they cite Campbell (1975), who in turn cited several published variations of this argument, including a quotation from his own earlier work (Campbell and Stanley, 1966). In his

1975 article, Campbell rejects this degrees of freedom argument as a caricature of case study research. As he notes, this logic would imply that a case study researcher should be able to find a multiplicity of explanatory models, any one of which affords a perfect fit to the available observations. This is simply not found in practice, because each case in fact affords a large number of data points of different kinds, providing ample opportunity to examine internal consistency and to test observations against theoretical predictions.

In his edited volume *What is a Case?* (Ragin & Becker, 1992), Ragin (1992, p. 1) contrasts "a study that uses interviews of employees to construct a picture of the informal organization of a firm" versus one "that uses interviews of employees to address variation in job satisfaction." The two are superficially similar, but the second would count employees as cases, affording ample degrees of freedom, whereas the first would regard the firm as a case, nominally just a single observation. A firm with 1,000 employees might be one case or 1,000 cases. George and Bennett (2005, pp. 29-30) summarize the issue as follows:

> We have criticized … the definition of a case as a phenomenon in which we report only one measure on any pertinent variable. It is this definition that leads to the conclusion that case studies suffer from an inherent degrees of freedom problem. In fact, each qualitative variable has many different attributes that might be measured. Statistical researchers tend to aggregate variables together into single indices to get fewer independent variables and more degrees of freedom, but case study researchers do the reverse: they treat variables qualitatively, in many of their relevant dimensions.

Our illustrative study by Coburn, et al. (2012) may serve to clarify this distinction. Their study employed just 12 teachers, and the number of attributes characterizing teachers' social ties, expertise, depth of interaction, and instructional quality over multiple years exceeded twelve. Nonetheless, analysis of configurations of these variables yielded a parsimonious model characterizing the differences between teachers who were versus were not able to sustain reform.

Design and Enactment

Selecting cases. A CCS investigation begins with some conception of the question or phenomenon of interest, which guides the definition and selection of cases for

study. The researcher must attend carefully to the questions of (1) what constitutes a case, and (2) what cases are cases *of*. Cases are often generic, conventionally defined units (e.g., lessons, teachers, classrooms, or school districts), but an observation of teachers and students during a lesson might serve as a case of classroom management methods, discourse patterns, enactments of specific curricular or instructional approaches, or individualization of instruction for different learners, for example.

Research questions may evolve as a study proceeds. As Ragin and Becker (1992) acknowledge, one view holds that:

> To begin research with a confident notion of "What is a case?" (or, more precisely, what this - the research subject - is a case of) is counterproductive. Strong preconceptions are likely to hamper conceptual development. Researchers probably will not know what their cases are until the research, including the task of writing up the results, is virtually completed. What it is a case of will coalesce gradually, sometimes catalytically, and the final realization of the case's nature may be the most important part of the interaction between ideas and evidence. (Ragin, 1992, p. 6)

Having conceived of cases and having arrived at some preliminary notion of what they are cases of, the investigator must next, of course, select specific cases for study. Flyvbjerg (2006, p. 230) summarizes several alternative rationales for case selection. While not ruling out selection by various statististical sampling strategies, Flyvberg focuses on purposive sampling criteria, in support of what he refers to as "information-oriented selection." These include (1) selection of extreme or deviant cases, (2) selection so as to achieve maximum variation in cases, (3) a focus on critical cases, and (4) selection of paradigmatic cases. Extreme or deviant cases may be studied "To obtain information on unusual cases, which can be especially problematic or especially good in a more closely defined sense." A set of cases that vary along specified dimensions can offer "information about the significance of various circumstances for case process and outcome…." Critical cases can support "logical deductions of the type, 'If this is (not) valid for this case, then it applies to all (no) cases.'" Paradigmatic cases may be chosen to exemplify theories, processes, or phenomena, offering prototypical illustrations to support communication and understanding.

It should be noted that a focus on critical or paradigmatic cases may entail deliberate selection of cases that share a particular outcome. Ragin (2009) focuses on selection of cases where the goal is to understand how a particular type of outcome emerges. He suggests selecting a set of cases to include those that reflect the outcome in question and those that do not; but in selecting cases that do not reflect the outcome, he suggests focusing on those that have a possibility of the outcome. This requires a "*possibility analysis*" which he defines as "a method that directly examines the degree to which the members of a given set of cases are candidates for the outcome under investigation" (p. 523).[20] He acknowledges that case selection requires substantial knowledge about potential cases and enough theory to identify "the kinds of settings capable of sustaining a process that could lead, potentially, to the outcome in question" (p. 526). He worries that random selection would fail to distinguish between negative cases that are and are not plausible candidates for the outcome: "in a conventional quantitative analysis, the inclusion of implausible cases simply inflates theory-confirming correlations" (p. 523) . QCA, which allows preliminary comparison of the features of cases with and without the outcome, provides a useful tool for possibility analysis. While he sees purposive sampling of negative as well as positive cases as focusing on a population of interest, he acknowledges the population is harder to define in the negative case.

However, in CCS (as is also true with GCI), an examination limited solely to cases sharing a common outcome cannot establish the causal antecedents of that outcome. That said, such an examination could certainly disconfirm the putative necessity of a given cause. Likewise, an examination solely of cases in which some outcome did not occur could disconfirm the putative sufficiency of one or another antecedent condition. George and Bennett (2005, p. 23) state that "case study researchers sometimes deliberately choose cases that share a particular outcome. Practitioners and analysts of case study methods have argued that selection on the dependent variable

[20] The example Ragin offers entails cases of countries where the outcome is ethnic political mobilization and where the criteria for candidacy might include features like ethnic diversity, ethnic inequality, or a non-repressive political system, as absent these features, the outcome of ethnic political mobilization is likely impossible.

should not be rejected out of hand. Selection of cases on the basis of the value of their dependent variables is appropriate for some purposes, but not for others."

The conception of what cases were cases *of* evolved in our illustrative study. As described below, Coburn, et al. (2012) set out to study "how district reform strategies interacted with human and social capital in the development of teachers' capacities to enact new approaches to the teaching of mathematics" (p. 144). To that end, they began with a purposive selection of four schools:

> Consistent with the exploratory, theory-building purpose of our study, we used purposive sampling … to select elementary schools and teachers within those schools for our study. Because the overall study was interested in how schools with contrasting organizational conditions— different levels of social and human capital—implemented the new mathematics curriculum, we sought four schools that varied along these two dimensions. … The final sample thus included … one school with strong professional community and strong teacher expertise, one with strong professional community and weak teacher expertise, one with weak professional community and strong teacher expertise, and one with weak professional community and weak teacher expertise. (Coburn, et al., 2012, p. 145)

However, as the study evolved and the focus shifted to the teacher level, these school-level contrasts had little relevance to the final analysis. Instead, the twelve chosen teachers became cases of the relationships between teachers' social networks and their ability to sustain reforms after supports were withdrawn.

Data Collection and Preliminary Analysis. Coburn, et al. (2012) illustrates how data were collected to support the derivation of variables enabling comparison. Note that their variables were conceived as binary attributes of teachers (cases), as opposed to the kinds of variables that take on a range of numerical values (e.g., the presence or absence of high-level expertise in teachers' social networks). Such variables may be created via thoughtfully developed decision rules for coding specific case attributes as present versus absent. Note also that Coburn, et al.'s (2012) variables are analyzed not in isolation. Rather, the authors look for patterns across these binary variables (configurations) that can provide parsimonious explanations of alternative outcomes for different teachers. The authors capitalized on an unexpected development in the course of a longitudinal

research study tracking the scale-up of a complex elementary school mathematics reform effort in a mid-size urban school district. The district had adopted a new curriculum, *Investigations in Data, Numbers, and Space*, which focused on "big ideas" rather than algorithms and memorization, but which placed considerable demands on teachers to implement well. Substantial professional development was provided and mathematics coaches were on hand in each school. Then, in the third year of implementation, there was a leadership change in the district and new priorities were established. Supports for the new reform were scaled back dramatically as the district-level focus shifted from mathematics to English learners. This provided an unexpected opportunity to study factors related to sustainability of reform-related instructional approaches after supports were largely withdrawn. In their study, the authors focused on 12 teachers in four purposively chosen elementary schools. Of these 12, 7 were able to sustain the new instructional strategies in the absence of formal support during year three of the study, and 5 were not. The researchers were interested in qualities of the teachers' social support networks that might account for these different outcomes.

Analysis. CCS analyses are distinguished by a logic of configurational analysis, relying on thorough investigation of each separate case together with cross-case comparisons. A broad range of research methods may be employed to this end, spanning methods conventionally referred to as both "qualitative" and "quantitative." There are, however, several distinctive research methods that have been developed specifically within this methodological tradition. In this section, we focus on two foregrounded by our key theorists: "typological theorizing" as described by George and Bennett and Qualitative Comparative Analysis (QCA) as described by Ragin and colleagues (1987; Rihoux & Ragin, 2009).

Typological Theorizing. A broadly applicable method, termed *typological theorizing* by George and Bennett (2005, pp. 233-262), is defined as follows:

> We define a typological theory as a theory that specifies independent variables, delineates them into the categories for which the researcher will measure the cases and their outcomes, and provides not only hypotheses on how these variables operate individually, but also contingent generalizations on how and under what conditions they behave in specified conjunctions or configurations

> to produce effects on specified dependent variables. We call specified conjunctions or configurations of the variables "types." (George & Bennett, 2005, p. 235)

Thus, the typological theory organizes cases into types, each characterized by a distinct configuration of values of the independent variables. The variables in typological theories are often nominal classifications, sometimes taking on just two values (e.g., present versus absent), although there are exceptions. The goal of the typological theory is to account for some dependent variable, which likewise may be a nominal classification. The authors emphasize that "the careful characterization of the dependent variable and its variance is often one of the most important and lasting contributions to research" (George & Bennett, 2005, p. 248).

The systematic construction of a typological theory requires an "iteration between theory and data and between within-case and cross-case comparisons" (George & Bennett, 2005, p. 254). The research may begin with a somewhat ad hoc selection of cases and independent variables, guided by prior theory and the investigator's own working hypotheses. Once an initial typology is constructed, it may become evident that additional independent variables are needed or that some already included are unnecessary. Examination of the configurations present in the initial set of cases can guide the selection of additional cases to be added to the analysis. The authors acknowledge that "Adding variables increases the complexity of the research design, and each new variable requires additional observations if it is to be tested, but new variables do not raise an inherent problem of indeterminacy as long as they generate additional independent observable implications on causal processes and outcomes. This is true whether these independent observable implications are in the same case or in a separate case" (George & Bennett, 2005, p. 247).

Four patterns (research designs) that may be pursued are (1) examination of cases within the same type to confirm that they have similar outcomes or to search for those with deviant outcomes for further investigation; (2) examination of highly similar cases ("adjacent types") with differing outcomes; (3) a focus on most-likely, least-likely, and crucial cases (defined below); and (4) examination of least similar cases, i.e., those differing with respect to all but one independent variable but nonetheless sharing a

common outcome. Most-likely, least-likely, and crucial cases are "single cases that may be particularly informative for theory development" (George & Bennett, 2005, p. 253), explained as follows:

> In a most-likely case, a single variable is at such an extreme value that its underlying causal mechanism, even when considered alone, should strongly determine a particular outcome. If … other… variables… point toward the same outcome as the extreme variable, then this is a crucial case. … Conversely, if a case is weakly determined or least likely for a single causal mechanism, and alternative hypotheses offer different predictions, but the causal mechanism still correctly predicts the outcome, then this constitutes a crucial case that offers the strongest possible support for the mechanism. (George & Bennett, 2005, p. 253)

Thus, typological theorizing represents a distinct method, or a set of closely related methods, within the CCS tradition. The method guides the investigator's selection of cases, characterization of those cases according to independent and dependent variables, organization of cases into types, and theorizing about the causal mechanisms accounting for outcomes within each type. The resulting theorized typology may allow for multiple paths to a common outcome, but often the analysis will suggest refinements to the dependent variable, when what was at first regarded as a single outcome turns out on closer inspection to differ in theoretically significant ways from one type of case to another.

Qualitative Comparative Analysis. The word "qualitative" in Ragin's (1987) original formulation of QCA referred to a focus on qualitative differences among cases—attributes that were either present or absent, not a matter of degree. Based on a careful study of individual cases, the investigator defines multiple potentially relevant attributes. In the original formulation of QCA, what has since come to be termed crisp set QCA (csQCA), each of these attributes is dichotomized—represented in a binary variable indicating whether that attribute is present (1) versus absent (0) for each case. The outcome of interest is likewise dichotomized as either having occurred (1) or not having occurred (0). Attributes and outcomes are then arrayed in a data table with a row for each case and a column for each attribute together with a column for the outcome. Ragin (1987) demonstrates how Boolean algebraic operations can be applied to this table to

arrive at parsimonious descriptions of those combinations of attributes associated with the presence versus absence of the outcome of interest. The relations captured by the data table can be expressed in set theoretic terms. If a given condition A is present for all cases in which some outcome O is observed, then the set of cases with outcome O is a subset of the set of cases with condition A. More complex patterns, involving multiple attributes, may be expressed similarly. Note that unlike a description in terms of correlation, the set-subset relation between two variables is asymmetric, and so is naturally suited to the problem of identifying necessary and sufficient combinations of conditions for some outcome. QCA is far from a mechanical algorithm yielding a single outcome. Rather, it is a flexible tool the analyst uses, guided by substantive theory, to investigate configurational patterns in the data at hand. There is substantial scope for judgment in the initial selection of attributes, in the reduction of each attribute to a binary variable, and in the selection of a final QCA model specification. Along the way, the investigator may engage in thought experiments, considering the likely outcomes associated with patterns of attributes not actually observed. Discrepant cases may be identified for more intensive study and new attributes may be identified as the analysis proceeds iteratively toward a satisfactory data summary. Statistical software is available to implement these analyses once the data table has been constructed.

Ragin (2000, 2008) has extended his original QCA formulation to include what is now termed fuzzy-set Qualitative Comparative Analysis (fsQCA). This approach relaxes the requirement that a given case be "fully in" or "fully out" of the set sharing a particular attribute. Instead, set membership is treated as a matter of degree ("fuzzy"), with 0 designating fully out, 1 designating fully in, and .5 designating a cross-over point of maximal ambiguity. Thus, each attribute is represented by "a continuous variable that has been purposefully calibrated to indicate degree of membership in a well defined and specified set." Thus, with fsQCA, the data table contains values between 0 and 1. Statistical software is likewise available to calibrate data of this kind, yielding membership scores for each case. [This description is informed by Whittington, et al., 2013, p. 286.]

Rihoux and Lobe (2009, pp. 229-237) describe a series of analytical and procedural choices in Qualitative Comparative Analysis, framed as "dialogues along the

way." Some of their terminology is specific to QCA, but a brief summary of their presentation may nonetheless give a sense of how such a study might progress. Their dialogues (or "operations") are organized into three stages, although they should not be regarded as a strictly linear sequence. The first stage, "Before the 'analytic moment'," encompasses deliberations around (1) research design and case selection, (2) gaining case knowledge, (3) defining the outcome of interest, (4) model specification, i.e., selection of conditions, and (5) visualizing/synthesizing cases and the models. The second stage, "The 'analytic moment'," addresses processes of synthesis and minimization via the operations of (6) dichotomization or threshold-setting, (7) truth-table exploration and contradiction solving, (8) minimization, (9) solving 'contradictory simplifying assumptions', and (10) selection among alternative minimal formulae. The final state, "Downstream: Interpretation," encompasses (11) factoring out conditions in the minimal formulae, (12) case-by-case interpretation, (13) interpreting cross-case patterns, (14) 'limited historical' generalization, and (15) culmination.

Coburn, et al. (2012) used QCA in their research. The binary outcome variable (high-quality reform-related instruction in year 3) as requiring "high" observation-based ratings on all three of (1) maintenance of cognitive demand, (2) attention to student thinking, and (3) intellectual authority (i.e., vesting authority in mathematical reasoning rather than the teacher or the text). Based on surveys and interviews, including interviews with network members identified by target teachers, they coded a series of binary indicators describing teachers' support networks. Their initial data table included just "the presence or absence of high-depth interaction, high level of expertise, and strong ties in all 3 years" (Coburn, et al., 2012, p. 152). This coding proved inadequate, because some configurations of conditions characterized sets of cases including both positive and negative outcomes. The authors then refined their data table to include as separate attributes the presence versus absence of high-depth interaction, high expertise, and strong ties in each of years 1, 2, and 3 separately. Some of these variables were subsequently dropped. Their final resulting data table was as follows:

Presence or Absence of Social Network Conditions and Outcome of Interest

CONDITIONS							OUTCOME	
Y1 Depth	Y1 Expertise	Y1 Tie Strength	Y2 Depth	Y2 Expertise	Y2 Tie Strength	Y2 High-Quality Instruction	Y3 HIGH-QUALITY INSTRUCTION	TEACHER
0	1	0	1	1	1	1	1	Xandria, Denise
0	0	0	1	1	1	1	1	Nina
0	0	1	0	1	0	1	0	Sarah
0	1	0	0	0	1	0	0	Tara
0	1	0	1	0	0	1	0	Larissa
0	1	1	0	0	1	1	0	Winona
0	1	1	1	0	0	1	1	Kathy
0	1	1	1	1	0	1	1	Laura
1	0	0	1	0	1	0	0	Don
1	0	1	1	1	1	1	1	Quinn
1	1	1	0	1	0	0	1	Florence

NOTE.—1 = presence of a given condition or outcome; 0 = absence of a given condition or outcome.

(from Coburn, et al., 2012, p. 152)

In the final analysis, three distinct constellations of conditions were found, each of which was sufficient for sustainability in year 3. Different constellations of conditions leading to both positive and negative outcomes are thoughtfully described. The authors note (Coburn, et al., 2012, p. 154) that with 7 attributes in their final model, there were 2^7 or 128 possible configurations, of which just 11 were observed across the 12 teachers. Nonetheless, the use of QCA to guide a dialogue with the data supported the construction of a compelling and internally consistent account of what Ragin (2008, cited by Coburn, et al., 2012, p. 151) called "causal recipes" giving rise to the outcome of interest.

Reporting and Evaluation

Coburn, et al. draw conclusions with implications for policy grounded in their in-depth analysis, noting the way their research contributes to our understanding of the dynamics of sustainability:

> Teachers' social networks in the first 2 years of the initiative influenced their ability to sustain reform-related instructional approaches after supports for reform were withdrawn. Social networks with combinations of strong ties, high-depth interaction, and high expertise enabled teachers to adjust instruction to new conditions while maintaining the core pedagogical approach. (p. 137)

As suggested by this quotation, the conception of validity within the CCS tradition includes attention to the utility of analyses for significant stakeholders. Ragin (2007) also highlights criteria of parsimony and plausibility of counterfactuals as criteria for selection among alternative models. CCS methods are used to support what may best

be described as bounded generalizations. "Social constructivists, cognitive theorists, and historical institutionalists may welcome the comparative advantages of case studies for addressing qualitative variables, individual actors, decision-making processes, historical and social contexts, and path dependencies" (George & Bennett, 2005, p. 9). In addition, as noted by Ragin and Rihoux (2004, p. 10),

> … policy researchers, especially those concerned with social as opposed to economic policy, are often more interested in different kinds of cases and their different fates than they are in the estimation of the net causal effect of independent variables across a large, encompassing population of observations. After all, a common goal of social policy is to make decisive interventions, not to move average levels or rates up or down by some small fraction.

In their presentation of findings, Coburn, et al. (2012) present three distinct profiles of social networks that were sufficient for sustained innovation, and draw implications for the ways teachers' social networks might be developed or supported to make this favorable outcome more likely. They highlight the fact that no single variable they examined was sufficient to account for teachers' success versus failure in sustaining mathematics teaching innovations in year 3. Instead, it was the configuration of related variables that accounted for the patterns observed: "social networks with a combination of a high level of expertise, strong ties, and high-depth interaction focused on substantive issues related to mathematics and pedagogy sufficient to support sustainability (p. 139). There is no explicit statement as to the range of cases to which their findings might generalize, but their compelling discussion, drawing on prior theory and supported by consistencies observed across their 12 cases, lend credence to the proposition that their findings are of more general import.

Discourse Analysis (DA): Critical Discourse Analysis (CDA) and Systemic Functional Linguistics (SFL)

Discourse analysis (DA) is the study of language-in-use, of language as it shapes and is shaped by its social context. Work in various traditions of discourse analysis convincingly shows that language does much more than merely communicate information. In addition to representing ideas or experiences, language also constructs interpersonal relationships, including the structuring of power, authority, and identity.

Each clause also relates to what has come before and moves the text or conversation along in some direction. Linguists have developed various models and methods for discourse analysis, which can be highly technical, and various of these tools have been used in education research. Gee (2011, p. 9) describes his own approach to DA as "critical discourse analysis," as distinguished from "descriptive" forms of discourse analysis. He explains that the goal of descriptive approaches is to theorize and explain how language works, whereas the goal of "critical" approaches is "not just to describe how language works … [but also] to speak to and, perhaps, intervene in, social or political issues, problems, and controversies in the world." Our first key theoretical resource is Gee's (2011) textbook setting forth his model for critical discourse analysis, *An Introduction to Discourse Analysis* (3rd ed.). Gee's approach "looks at meaning as an integration of ways of saying (informing), doing (action), and being (identity), and grammar as a set of tools to bring about this integration" (Gee, 2011, p. 8).

In addition to Gee's Critical Discourse Analysis (CDA) we have chosen to focus also on an approach to discourse analysis, Systemic Functional Linguistics (SFL), which Gee has employed in his work.[21] SFL was developed by Halliday (1978, 1994) with elaborations and extensions due to Martin (1992, 1993) and others. Our second key theoretical resource is a chapter by Schleppegrell (2012), which provides an overview of SFL and its applications in education and other areas. As explained by Schleppegrell (2012, p. 21), "SFL facilitates exploration of meaning in context through a comprehensive text-based grammar … SFL describes three abstract functions (*metafunctions*) … the *ideational*, *interpersonal*, and *textual*, as every clause in our language simultaneously construes some kind of experience (ideational metafunction), enacts a role relationship with the listener or reader (interpersonal metafunction), and relates our messages to the prior and following text and context (textual metafunction)."

Our two illustrative cases for this methodology are each analyses of mathematics lessons. Zolkower and Shreyar (2007) apply SFL to explicate a teacher's skillful mediation of a "think-aloud" whole-group discussion in a sixth-grade mathematics classroom. O'Halloran (2004) uses SFL to contrast oral discourse in three mathematics

---

[21] Gee (2010) points the readers of his introductory text to Halliday's functional approach to grammar as a supplement to his book, citing Halliday and Hasan (1989) and Halliday and Matthiessen (2004), confirming the relevance of functional linguistics to his CDA approach.

lessons, drawn from classrooms in three different high schools in Perth, Australia. These include two elite private schools, one for boys and one for girls, and "a comparable Year 10 working-class Government school" (O'Halloran, 2004, p. 195). Her analysis, which is taken from a larger study, investigates differences in how mathematics is represented, how students are positioned as learners, and how they engage with mathematics across the three classrooms. Her 2004 chapter focuses on oral discourse, but in the larger study from which it was taken, O'Halloran (2000) theorized mathematics discourse as "a multi-semiotic[22] construction which involves the use of language, mathematical symbolism and visual display" (O'Halloran, 2004, p. 191). Discourse analysis may also include gesture and eye gaze (for spoken discourse) and elements of context that are relevant to understanding what an utterance means. Both Schleppegrell (2012) and Gee (2011) highlight as a strength of SFL the fact that it can be extended in this way to examine multi-modal texts.

Aims and Conceptions of Social Phenomena

Researchers in this methodological tradition seek to explain, "How does this text mean what it does?" (Schleppegrell, 2012, p. 22). ("Text" as used here is a meaningful stretch of discourse of any length, spoken or written and accompanied by meanings made in other modalities as well.) This explanation involves precise characterization of the choices speakers and writers make from among the many resources language provides. SFL uses categories of ideational resources (*field*), referring to the ways language conveys topic or content; interpersonal resources (*tenor*), referring to the ways language constructs or enacts relationships and conveys attitudes; and textual resources (*mode*) referring to the ways language relates to context, including the text that has come before. These categories are subdivided further. Taken together, sets of choices within these categories define alternative linguistic *registers*.

Gee (2011, p.8) illustrates this notion of register by contrasting two sentences:

1. Hornworms sure vary a lot in how well they grow.
2. Hornworm growth exhibits a significant amount of variation.

[22] Gee defines semiotics as "a term for the study of sign systems" (p. 211) of which language, mathematical symbolism, and related visual displays are examples.

> Sentence 1 is in a style [register] of language (called the "vernacular") we use when we want to talk as an "everyday person," not as a specialist of any kind. This is the identity (being) it expresses. … Sentence 2 is in a specialist style of language, one we would associate with biology and biologists.

Gee goes on to point out how "hornworms" serves as the subject of sentence 1, whereas the nominalized process "hornworm growth" is the subject of sentence 2. Similarly, sentence 2 uses the nominalization "variation" instead of the verb "vary." Whereas sentence 1 conveys the speaker's emotional investment with the word "sure," sentence 2 is dispassionate. These and related linguistic features (e.g., the quantification "significant" in sentence 2 versus "a lot" in sentence 1) make each of these utterances more suitable for some contexts of use and less suitable for others.

The perspective of DA offers a powerful lens on social phenomena, extending well beyond the text itself. In discussing aspects of context that could potentially be relevant to the meaning of a text, Gee (2011, p. 100) includes "the physical setting in which a communication takes place and everything in it; the bodies, eye gaze, gestures, and movements of those present; what has previously been said and done by those involved in the communication; any shared knowledge those involved have, including shared cultural knowledge" (Gee, p. 100). Of course, it is never possible to analyze context exhaustively. The discourse analyst must decide what to include according to the purpose of a particular study.

In line with Gee's (2011) aspiration for "critical discourse analysis," various researchers used SFL to illuminate education phenomena so as to guide wiser or more effective practice. For example, Zolkower and Shreyer (2007) show convincingly that students may be afforded agency and may engage in authentic mathematical discourse practices even in a lesson that appears superficially to be teacher-directed. As they state in their conclusion (Zolkower and Shreyer, 2007, p. 200):

> We intended this analysis to illustrate the power of SFL for studying the inner grammar of classroom interactions so as to illuminate the complexities and subtleties in the teacher's mediating role. At the surface level, the authoritative manner in which the teacher mediated the whole-class discussion may have seemed un-conducive to engaging her students in genuine mathematical

exchanges. Our analysis of Ms. L.'s moves in the conversation allowed us to demonstrate that the contrary was the case, highlighting how through her very lexico-grammatical choices, the teacher guided these sixth graders in thinking aloud together about algebraically significant patterns in a number array.

Design and Enactment

Discourse analysis begins with the selection of one or more research sites and then of specific texts for focused analysis. Thus, Zolkower and Shreyer (2007, p. 184) chose for close analysis a classroom discussion "captured during a nine-month study of a 6th grade classroom attended by African-American, Asian, Caucasian, and Latino students" in the course of which "data were collected from 10 non-consecutive periods of mathematics instruction." They state further that, "The school is located in a large urban setting in the United States. This 6th grade class of 26 students was highly heterogeneous in regard both to socio-economic class, racial, ethnic, cultural, and linguistic background and level of mathematical performance. At the time of this study the teacher (Ms. L.) had 15 years of teaching experience."

Having chosen an illustrative text or a set of contrasting texts for analysis, the researcher will next prepare a detailed transcription for analysis. For spoken texts, the transcription would typically include more than just the words spoken. Pauses, stress, hesitations, and other details may be included. Different features of the text and context may be coded, depending on the purpose of the analysis. Beyond the transcriptions, Zolkower and Shreyer (2007, pp. 184-185) go on to describe their procedure as follows: "Lessons were observed by one of the researchers and detailed field-notes were taken to capture both verbal and non-verbal features of the interaction, the latter including facial and body gestures as well as the use of diagrams, tables, graphs, and other artifacts. Lessons were also audio-taped and transcribed verbatim."

Coding the text will involve segmentation into units, which will then be coded according to one or more linguistic features. Different features may be chosen for analysis, again depending on the purpose of the study. Zolkower and Shreyer (2007) first segmented their transcribed text into "episodes" and then selected episodes of whole-class discussion for more fine-grained analysis. One of their coded segments is shown below:

#	Speaker	Clauses	Mood	Speech Function
16	Teacher	[40]. Leslie	Min	Of
17	Leslie	[41]. Do you see at the corners? [42]. Starting with 26, it's odd, even, odd, even and so on	Int(p) Dec	Co St
18	Teacher	[43]. Can you please elaborate?	Int(p)	Co
19	Leslie	[44]. I'm looking at the columns [45]. It's odd, even, odd, even	Dec Dec	St St
20	Teacher	[46]. *(Writes alternating "o" and "e" on the top of each column)* e o 1 o e 2 3 4 e o 5 6 7 8 9 o e 10 11 12 13 14 15 16 e o 17 18 19 20 21 22 23 24 25 o 26 27 28 29 30 31 32 33 34 35 36		
21	Alex	[47]. Every time you go slanted going up is minus an odd number	Dec	St
22	Teacher	[48]. *(Uses arrows to notate Alex's contribution)*		

Zolkower & Shryer (2007, p. 191)

The "Mood" codes (part of SFL's *interpersonal* meta-function) classify clauses as Declarative (Dec), Imperative (Imp) or Interrogative (Int). Minor clauses (Min) are not coded for mood. The "Speech Function" codes in SFL classify clauses as Statement (St), i.e., giving information; Offer (Of), i.e., giving goods or services; Question (Qu), i.e., demanding information; Command (Co), i.e., demanding goods or services; and Check (Ch), i.e., checking to make sure the listener is attuned.

In their article, Zolkower and Shryer (2007) provide eight such detailed coded excerpts, of which just one is reproduced here, which are woven into a narrative account of the entire class discussion. Their careful analyses of the teachers' linguistic choices show how student ideas both relevant and irrelevant to the assigned task are all treated in an affirming, impartial way, with no hint from the teacher as to which of the students' various ideas are, in fact, key to solving the mathematical problem posed.

The descriptions of each of these studies include further details as to coding conventions used. Some DA studies also report evidence of inter-rater agreement in transcript segmentation and coding. Although neither of our two illustrative studies provides this information, each provides excerpts like the one shown above, enabling the reader to better understand and to some extent even to verify the coding. Coding may be

facilitated by the use of special-purpose software. O'Halloran (2004, pp. 191-192), for example, made use of "a prototype version of *Systemics* (O'Halloran and Judd 2002), software for systemic analysis which contains pre-programmed options for each of [various metafunctions] documented in Halliday (1994)."

O'Halloran (2004) engaged in a similar coding process, organizing her transcript at the successively finer-grained levels of paragraph, clause complex, clause, and word group. In addition to SFL, she segmented the lesson according to a system of 46 different "microgenres" defined by Christie (1994), including such categories as "Homework Check," "Lecture," "Diagnostic Activity," and "Test/Exam." These were somewhat more fine-grained than Zolkower and Shreyer's episodes. O'Halloran (2004, p. 196) illustrates the microgenre of "Teacher disciplinary interruption" as follows:

The teacher interrupts the genre-in-progress for disciplinary purposes.

Field: Related to student behaviour.

Tenor: Participants: teacher and student(s)

Affect: Low-high/neutral to negative.

Power: Unequal relations with full exposure of the teacher's positioning as dominating.

Mode: i. oral (two-way, turn-restricted, turn-controlled, quasi-dialogue).
ii. constitutive (construction/reconstruction).

Analyses may include tabulations of the relative frequencies of various codes as well as graphical displays. For example, O'Halloran's (2004, pp. 196-198) includes a diagram showing the sequence and duration of successive microgenres over the course of the lesson in each of her three classrooms. The diagram points up a striking contrast, which she summarizes by pointing out, "The clear and progressive structure of Lesson 1 [in the elite boy's school] may be contrasted to Lessons 2 [elite girl's school] and 3 [working-class government school], which consist of a series of disjointed movements between different Microgenres… the results of the discourse analysis indicate that interpersonal relations are a major area of concern in Lesson 3 and, to a lesser extent, Lesson 2. In contrast, the monofunctional tendency … in Lesson 1 appears to be the ideational meaning, or the mathematics content of the lesson."

Discourse analysis can both highlight telling details and reveal general patterns that would be otherwise difficult to perceive. In O'Halloran's Lesson 3, for example, the teacher refers to algebraic symbols as "letters and numbers" whereas in Lesson 1 they are referred to as "constants and variables."

After fine-grained analyses of a series of lesson episodes in which the teacher elicits and responds to individual students' mathematical ideas, Zolkower and Shreyer show how, taken together, these episodes support their general conclusion. Their discussion of the excerpt reproduced above may convey some sense of this form of fine-grained analysis.:

> Ms. L. uses a minor clause with the speech function of an offer in order to nominate a student to speak (16.40). In response to Leslie's contribution (17.41-42), the teacher chooses to use a polar interrogative as an explicit command for the student to further elaborate her idea (18.43). After Leslie does so using a more precise wording ("columns" replaces "corners") (19.44-45), Ms. L. enhances her contribution with a non-verbal move in which she inserts symbols for "odd" and "even" on the array (20.46).

Note that in last sentence just quoted ("Ms. L. enhances her contribution …") refers to something the teacher wrote on the board, rather than words spoken.

In summarizing their findings, Zolkower and Shreyer (2007, p. 198) point out how:

> … the teacher structured the whole-class discussion as an interpersonal gateway for expanding her students' potential to mediate their emergent algebraic activity via verbal thinking. … [She] guided her students to interpret the task at hand as an open-ended search for patterns, in part, by refraining from explicitly evaluating their contributions. She never let on that the best fitted pattern had been found, or even that it existed at all. … This sheds light on an important lesson to be learned by the students in this class, namely, how to conduct an open-ended search and verbalization of patterns so that it successfully [functions] within the larger context of an algebra inquiry.

Reporting and Evaluation

As reflected in our illustrative studies, the findings from linguistic analyses tend to be stated in a somewhat technical vocabulary that may be unfamiliar to nonspecialists, but major conclusions may be restated in more accessible terms. Thus, Zolkower and Shreyer (2007) point out that "Ms. L. [refrained] from explicitly evaluating [students'] contributions," but then go on to support that conclusion with a more technical account of Ms. L.'s specific linguistic choices. Similarly, one of O'Halloran's (2004, p. 222) general conclusions is her assertion that, "The analyses of Lessons 1 and 2 suggest that stereotypical gender constructions are reinforced in single-sex classrooms in elite private schools." This is then supported by careful tabulations showing that:

> The male students operate from a position of power, the social relations are involved but not intimate, there is minimal expression of affect and few instances of reacting. Interpersonal relations are maintained as formal, constant and steady. In contrast, the female students operate from a more deferential position, express less certainty, utilize politeness strategies, enter into less formal and more intimate relations, express euphemistically a range of attitudes and comments and react with an increased volume of interpersonal meaning.

DA has found many applications beyond those shown in our two illustrative articles. One example, presented by Gee (2011, pp. 176-192), suggests how these methods can also be used to clarify and, hopefully, to help resolve sources of tension and misunderstanding between academic researchers and school personnel. This is a case study of a history professor at a university seeking to engage local middle-grades school teachers in having their students collect oral histories by interviewing their relatives and neighbors. Another example is provided by Schleppegrell (2013), showing how SFL meta-language can equip teachers with tools to support second-language (L2) learners. Among numerous further examples, Schleppergrell (2012, pp. 26-27) summarizes a study by Christie (2002) that used SFL to show "how instructional content and regulation of students are simultaneously managed by teachers," illustrating how teachers' use of content (instructional) and regulative (pedagogical) registers determines what information is made available to students and how they are positioned as learners. In summary, DA has brought to light significant, potentially mutable aspects of educational practice with

implications not only for students' understanding of curricular content, but also for their identities as learners and their appropriation of different subject matter discourses. It offers a powerful and distinctive lens on the processes of teaching and learning.

In discussing validity, Gee (2011, pp. 122-124) begins with a reminder that human interpretations of reality are mediated through language, and so validity cannot simply be a matter of correspondence between a discourse analysis and some objective "reality." That said, it is not the case that judgments of validity are entirely subjective. Gee refers back to the analytical tools of SFL, which are used to warrant claims or hypotheses (e.g., that conflict or awkwardness in a given situation arose because of unacknowledged differences between the discourses being used by different participants). He goes on to suggest four criteria for validity: convergence, agreement, coverage, and linguistic details. Convergence refers to the internal consistency of supporting evidence adduced from analyses of different linguistic resources. Agreement refers to consensus among linguists and also to the appearance of reasonableness of results to "native speakers" or members of the discourse communities implicated. Coverage refers to perceptions of explanatory power beyond the specific discourse samples analyzed, generalizing to the corpus from which the analyzed samples were drawn or to discourse in similar settings. Finally, linguistic details refers to the extent of tight connections between specific grammatical resources and the inferences made. Sound analyses will be extensively supported by detailed, specific examples from the discourse examined.

Survey Research and Measurement (SR&M)

The field of educational measurement is concerned with tests (or assessments),[23] primarily of student learning outcomes (achievement). Tests may be used to support classroom instruction; to monitor, compare, and evaluate education actors and entities; to inform decisions concerning licensure and certification; in admissions to higher education; and for other purposes. The field of survey research is concerned with the

[23] Different authors over the years have recommended various distinctions between the terms "test" and "assessment" but usage varies. The term "assessment" may connote a broader range of testing formats (beyond paper-and-pencil testing), a focus on testing for the purpose of guiding instruction rather than comparing individuals to one another, an evaluation of value or worth, or the integration of multiple sources of evidence to arrive at a more holistic or contextualized interpretation of an individual's capabilities or performance than can be obtained from test scores alone.

design, administration, analysis, and interpretation of instruments used to measure various kinds of knowledge, attitudes, values, or behavior. Surveys may be used to obtain information about school safety, classroom climate, students' perceptions of classroom instruction, or their health-related behaviors, for example.[24] These two fields may best be regarded as two distinct methodological traditions, as we are using the term. Each has its own graduate training programs, textbooks, research literatures, journals, professional organizations, and conferences. The roots of educational measurement may be traced to psychology, and its fundamental concern is with the measurement of individuals. Survey research has evolved mainly within the disciplines of sociology and political science, and although surveys are administered to individuals, the fundamental concern of survey research is with the measurement of populations represented by survey samples.

That said, the two fields share many related practices and concerns. Surveys may be used to describe individuals as well as populations. For example, a home language survey may be used to determine which entering students are administered an English language proficiency test. Conversely, achievement tests may be used to describe populations as well as individuals. For example, distributions of achievement test performance may be compared across teachers, schools, states, or nations. Both fields rely upon statistics, and practical problems from each area have spurred advances in statistical theory and methods. Scientific sampling methods have been developed primarily within the survey research field, but are also used in some achievement testing applications. Methods for studying single test items as well as test scores, constructing alternate forms of the same test, linking and equating different tests, among other matters, have been developed primarily within the tradition of educational measurement, but are also used in the design and analysis of surveys.

We have chosen two key theoretical resources and one large-scale application of these related methodologies. Additional information about testing and measurement is also available in chapters by Penuel and Shepard (this volume) and Rowan and Raudenbush (this volume). Our first key resource is *Survey Methodology* (2nd edition),

---

[24] See, e.g., the *California Healthy Kids Survey* modules described at https://chks.wested.org/administer/download.

by Groves, Fowler, Couper, Lepkowski, Singer, & Tourangeau (2009). Robert Groves, Floyd Fowler, and their co-authors wrote *Survey Methodology* in part to serve as the text for a semester-long graduate-level course, intended for students who have had at least one prior course in statistics. Following an overview of the process of designing, conducting, and analyzing surveys, most of the remaining chapters take up in turn the broad steps in that process. Our second key resource is *Educational Measurement* (4th edition), edited by Robert Brennan (2006). Since the first edition of *Educational Measurement* was published in 1950, successive editions have served as major handbooks for the measurement field. Following the editor's introductory chapter, the 2006 edition includes chapters on validation, reliability, item response theory (an important class of statistical models for analyzing test data), and on equating, linking, and scaling (primarily statistical models for improving the comparability of scores from different tests or from different forms of the same test). The remaining chapters address a range of more specific topics and applications.

Our illustrative application is TALIS 2013, the 2013 *Teaching and Learning International Survey* conducted by the Organisation for Economic Co-operation and Development (OECD, 2014). TALIS is "an international, large-scale survey that focuses on the working conditions of teachers and the learning environment in schools. TALIS aims to provide valid, timely and comparable information to help countries review and define policies for developing a high-quality teaching profession" (OECD, 2014, p. 26). Approximately 34 countries and economies participated in TALIS 2013, of which about three quarters were members of the OECD.

Aims and Conceptions of Social Phenomena

Both surveys and tests are sometimes used in the course of research within one or another methodological tradition, including but certainly not limited to Generalized Causal Inference and Social Network Analysis. In other applications, including our example of TALIS 2013, they are used independent of any specific research application, often under the auspices of governments or other organizations, to provide data that can then be used to inform policy or for various research purposes. Tests and surveys may be used repeatedly (perhaps annually) to monitor population trends and for other purposes (e.g., school accountability). Survey data may also be used in studies of the relationships

among different variables, such as the relationship between income and years of education (Groves, et al., 2009, p. 2).

Design and Enactment

Major activities in both testing and survey research may be organized into four broad stages of measurement, sampling, administration and scoring, and reporting. First, test or survey questions must be constructed to engage intended cognitive processes and to evoke meaningful and useful responses. Related questions may be assembled into tests or scales. These activities are referred to here as measurement. Second, respondents must be identified, a process referred to here as sampling. Third, the tests or scales must be administered, or survey questions must be posed, to the chosen respondents, typically under controlled (standardized) conditions; respondents' answers must be recorded; data must be analyzed to create scores; and, if information about groups of respondents is desired, these scores must be aggregated (summarized) for one or more groups or subgroups. These activities are treated under the heading of administration and scoring. Finally, findings must be reported, at the individual level, group level, or both. For most published tests, for large testing programs, and for some small-scale testing applications, additional statistics are calculated to describe the technical quality of the reported scores or score summaries. These are often presented along with the test or survey results. To make this brief discussion parallel to those for other methodological traditions, these reporting activities are discussed in the subsection on "Reporting and Evaluation."

Measurement. The construction of both tests and surveys starts from a specification describing the content to be covered, test length, question (item) formats, and related concerns. Guidelines for item writers may also be developed. Typically, more items will be written than are planned for the final version of a test, so that items may be discarded if they fail to function as intended during pilot or field test administrations. Items may go through various reviews and revisions even before they are first tried out with actual respondents. During some item development activities, respondents may be asked to "think aloud" while responding to test questions or may be debriefed after responding concerning their reasoning, so as to learn more about the cognitive processes items are eliciting. For achievement tests intended for large-scale applications, items are then assembled into test forms and administered to examinees like

those for whom the test is intended.[25] This "field testing" stage is used to study the statistical properties of the items, not to produce scores intended for some practical use (operational scores). Finally, based on both content and statistical considerations, test items will be assembled into operational test forms.

For achievement tests and for many surveys, the specification will describe one or more *constructs* the test or survey is intended to measure. Groves, et al. (2009, p. 41) explain that "'Constructs' are the elements of information that are sought by the researcher." They go on to state that "'Measurements' in surveys are ways to gather information about constructs" (p. 43). They do not offer any extensive treatment of "mismatches between a construct and its associated measurement" except to say that "Measurement theory in psychology (called "psychometrics") offers the richest notions relevant to this issue" (p. 50). The chapter in *Educational Measurement* on validation (Kane, 2006) devotes many pages to evolving uses of the term "construct" over the past sixty years. Suffice it to say that in measurement theory, a "construct" is some attribute, variable, or dimension a test or survey is intended to measure (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, & NCME], 2014, p. 11). In the models most often used for educational measurement, each test item is regarded as a partial, fallible indicator of the underlying construct. Administering a number of items each designed to measure the same construct provides more accurate information than would be provided by any single item concerning an examinee's status with respect to that construct (e.g., the examinee's level of ability, developed proficiency, self-efficacy, or whatever construct the test is intended to measure). The test questions actually administered may be regarded as a sample from some larger set of questions that could have been used instead. In some cases, several different sets of questions are assembled to create alternate forms of the same test, each form measuring the same construct.

Patterns of responses across items may guide the refinement of construct definitions. In TALIS 2013 (pp. 214-216), for example, twelve items designed to measure "teacher self-efficacy" were included. Using a statistical technique called factor

---

[25] In ongoing testing programs, separate field tests may not be required each time a new test form is developed. Alternatively, new items may be inserted into operational forms and field tested unobtrusively during regular administrations. Performance on these new items would not affect examinees' scores.

analysis, these were then grouped into three subscales of four questions each, labeled "self-efficacy in classroom management," "self-efficacy in instruction," and "self-efficacy in student engagement." Because TALIS was an international survey, it included forms adapted and translated for use in different national settings. Factor analysis was also used to examine the degree to which these different versions appeared to measure the same thing. The assembly of final test forms and the organization of items into subscales are usually guided by considerations of both content coverage and the statistical properties of the items, as determined after they have been administered.

Sampling. Survey research begins with consideration of both the population(s) the final survey statistics are intended to represent and the measurements (responses) that will be elicited (Groves, et al., 2009, ch. 2). Whereas theories and methods of measurement are most fully developed within the measurement field, theories and methods of sampling have been most fully developed within the survey research field. The requirement of *representation* holds that the sample of respondents must be representative of the population, and of sufficient size to yield estimates that are accurate enough to be useful.

Sampling proceeds from a general description of the target population to a precise definition of the *sampling frame*, which is developed to match the target population as closely as possible. A *sample* is then drawn (selected by an essentially random process) from the sampling frame so as to meet certain statistical criteria. The first principle in statistical sampling is that each potential respondent in the sampling frame must have a known, finite probability of being included in the final sample. The simplest sampling plan is referred to, appropriately, as a "simple random sample." If there are N potential respondents and the target sample size is some smaller number, n, then each potential respondent is sampled with probability n / N. In other words, each respondent has the same chance as any other of being included. Most large surveys use more complex sampling plans

In TALIS 2013 (pp. 208), for example, "The objective of the Teaching and Learning International Survey (TALIS) in 2013 was to obtain, in each participating

country, a representative sample of teachers for each ISCED[26] [or comparable educational] level in which the country participated." The sampling frame was defined to include "those teachers who teach regular classes in ordinary schools…. Teachers teaching to adults and teachers working in schools exclusively devoted to children with special needs are not part of the international target population and are deemed out of scope." Although it would be ideal if the sampling frame could include teachers in all of a country's schools, some adjustments were permitted:

> For national reasons, participating countries could choose to restrict the coverage of their national implementation of TALIS 2013 to parts of the country. For example, a province or state experiencing civil unrest or an area struck by a natural disaster could be removed from the international target population to create a national target population. Participating countries were invited to keep these exclusions to a minimum.
>
> TALIS 2013 recognised that attempting to survey teachers in very small schools can be inefficient and difficult. For each ISCED level, surveying teachers in schools with no more than three teachers at a specific ISCED level and those teaching in schools located in geographically remote areas could be a costly, time-consuming and statistically inefficient exercise. Therefore, participating countries were allowed to exclude those teachers for TALIS 2013 data collection, thus creating a national survey population different from the national target population. (p. 208)

TALIS 2013 did not use a simple random sample. Instead, two-stage cluster sampling was used, with schools as primary sampling units (PSUs) and teachers as secondary sampling units. This means that schools were randomly chosen first, and then teachers were randomly chosen within each of the selected schools. The sampling plan specified a target minimum sample size of 200 schools, with a target minimum sample of 20 teachers from each school sampled. Surveying more teachers in each of fewer schools is obviously more efficient than obtaining data from a simple random sample. Note, however, that each teacher is still sampled with a known probability, which could be

[26] ISCED, the "International Standard Classification of Education," is a uniform system of terminology for describing educational levels in different countries. TALIS 2013 included surveys of teachers at several ISCED levels, but not all countries surveyed teachers at all of these levels.

calculated as the probability of the teacher's school being chosen multiplied by the probability of that teacher being among those 20 actually surveyed, in the event that her or his school were in fact chosen. The TALIS sampling plan was actually even more complex than this. In order to help assure that the schools actually chosen were representative of the country as a whole, all schools might first be grouped into categories (strata), perhaps geographic regions, and designated numbers of schools might then be chosen within each stratum. Thus, the probabilities of inclusion might differ from one school to another, but it would still be possible to calculate the probability of any particular teacher being sampled. This is referred to as stratified two-stage cluster sampling.

Administration and scoring. Survey researchers have developed various data collection methods, including among others, mailed surveys, face-to-face interviews, telephone surveys, and surveys conducted online using the internet. Different methods may be combined. In TALIS 2013, for example, data were collected using "Separate questionnaires for teachers and school leaders, each requiring between 45 and 60 minutes to complete," which were "filled in on paper or on line" (OECD, 2014, p. 27).

The administration of many educational achievement tests is highly standardized. Tests are often carefully timed, and instructions may be read verbatim, following a script. Test materials are typically kept secure. In large testing programs, each copy of a test may have a unique identifier, and elaborate procedures may be instituted to assure that all copies are accounted for following the test administration. Online testing can simplify these procedures. In addition, with "adaptive testing" methods, typically computer-based, test forms are not created in advance. Instead, as the testing proceeds, items or blocks of items are selected for an individual examinee based on that examinee's previous responses. A statistical model is then used to generate scores on a common scale for all examinees even though all have not responded to the same items. In other educational testing applications, especially tests used for day-to-day instructional purposes, administration conditions may be less highly specified.

Test scoring typically begins with a determination of whether each separate question was answered correctly. (It may also be possible to award partial credit.) Item-level scores may simply be summed to yield a total raw score, or more complex statistical

methods may be used to generate scale scores based on the pattern of item responses. These initial scores are often converted to different score scales as aids to interpretation or to improve the comparability of scores across different test forms. Survey scale scores may also be placed on different scales to aid interpretation. In TALIS 2013, for example, the initial scores for each scale were calculated using weights derived via factor analysis. Next, these initial scale scores were transformed to a new set of scales:

> [Factor scores] were rescaled to have a standard deviation of 2 on the pooled sample used for estimating the loadings and intercepts (ISCED 2 countries), and the value 10 on the scale was made to coincide with the midpoint of the scale in the response options for the questions that make up the scale. In the cases of scales made up of items with response options ranging from 1-strongly disagree to 4-strongly agree, a score higher than 10, even if below the empirical scale average, indicates agreement with the items in the scale. A score below 10 indicates disagreement with the items in the scale. (OECD, 2014, pp. 214-215)

In many testing applications and in most surveys, individual respondents' scores are summarized to estimate averages or other statistics for populations and subpopulations. The details of the sampling plan will determine the precise formulas used to calculate estimates of population statistics from the sample data. With a simple random sample, an unweighted average of the test scores or survey responses for each respondent provides an estimate of the mean for the population they represent. With more complicated sampling designs, *sampling weights* must be used. These weights in effect tell how many members of the population each sampled respondent represents. They are inversely proportional to the probabilities of each respondent having been selected. In effect, this means that respondents representing larger groups receive more weight in the calculations. A population mean is estimated by a weighted average of all responses, calculated using the sampling weights. More complex formulas may be used to calculate additional population statistics.

Reporting and Evaluation

At the individual level, test results, and occasionally survey results, may be provided in the form of score reports designed to aid interpretation. These may indicate performance relative to some relevant reference group (e.g., a percentile rank),

performance relative to performance bands (e.g., "proficient"), relative strengths and weaknesses across different subareas tested, and some indication of score accuracy (e.g., margin of error). At the level of populations or subpopulations, survey results, and sometimes test results, are expressed initially as statistics describing the groups represented by the respondent samples. These statistics might be average responses on some measurement scale, for example, or proportions of respondents giving each possible answer for a single survey item. Statistical significance tests might be reported, indicating whether differences observed between subpopulations or across different points in time were larger than would be expected as a result of random fluctuations in the absence of any true subpopulation differences or trends.

In the TALIS (2013) survey report, presentations of results ranged from simple reporting of percentages of teachers providing different responses (in tables or charts), to those that presented comparisons across participating countries, to far more complicated representations depicting relationships among variables. So, for instance, in one line of related analyses we see bar charts reflecting the frequency with which teachers report engaging in different practices, then we see differences in relative frequency for particular practices disaggregated by country, and finally we see a two-by-two table depicting the relationships between particular practices and professional development opportunities by showing the numbers of countries where statistically significant relationships existed between a particular teaching practice and particular professional development opportunities. These analyses led to the following conclusion.

> Teachers who report participation in professional development activities involving individual and collaborative research, observation visits to other schools or a network of teachers are more likely to report using teaching practices that involve small groups, projects requiring more than a week for students to complete and information and computer technology (ICT). (p. 150)

In a subsequent set of analyses, indices of teachers' job satisfaction and sense of efficacy are reported, overall and by country, and then explanations for the differences are sought in different professional development opportunities and environmental features. This led to conclusions like the following:

> Teachers who report that they are provided with opportunities to participate in decision making at a school level have higher reported levels of job satisfaction in all TALIS countries and higher feelings of self-efficacy in most countries. The relationship between job satisfaction and teacher participation in school decision making is particularly strong for all countries. . . .
>
> Teachers' perception that appraisal and feedback lead to changes in their teaching practice is related to higher job satisfaction in nearly all countries, whereas the perception that appraisal and feedback is performed merely for administrative purposes relates to lower levels of job satisfaction in all TALIS countries. (p,182)

Thus survey findings like these can be used for both descriptive and explanatory purposes. Researchers further supported the generalizability of the findings by grounding them in existing literature.

Two principle concerns in evaluating the quality of scores are reliability and validity. Reliability refers to the stability, consistency, or replicability of test scores, and validity refers to a complex set of concerns having to do with the meaning and appropriateness of test scores in the context of some particular use or interpretation. Reliability and validity are relevant to the use and interpretation of individual scores as well as to the use and interpretation of score summaries for groups of respondents, as reported in the TALIS study. At the group level, however, the stability or replicability of survey and test results is affected by the sampling of respondents from a population (sampling error), as well as by the inherent imprecision of individual respondents' scores (measurement error). These two sources of imprecision both contribute to *standard error* (further discussed below), which is the statistic most often used to describe the accuracy of surveys. In addition to affecting accuracy, sampling can also affect the validity of survey results. Threats to validity associated with statistical sampling are discussed here under the heading of *bias*.

Reliability. Any examinee's score may be higher or lower on one testing occasion than another. The difference between an examinee's score on some single testing occasion and that examinee's (unknowable) long-run average score across imagined repeated testings is referred to as measurement error. That measurement error arises first, because the responses of an individual to any specific item are somewhat unstable.

Answers to some kinds of questions may be influenced by mood or the examinee's physiological state at the time of testing. Answers may vary due to lucky or unlucky guessing, momentary lapses of attention, errors in recording responses, or other causes. In addition, many educational measurement models treat an examinee's score on a test as an estimate of the score that examinee might have earned on any randomly chosen alternate form of the same test, i.e., different tests constructed according to the same test specification. In that view, measurement error also reflects the sampling of the particular questions (items) on the actual test from a larger domain of questions that might have been chosen or written instead. Measurement error can be reduced by increasing the number of items on a test. A longer test will yield more precise estimates of individuals' scores.

When tests are used to rank examinees according to their scores or to compare examinees to one another, or when patterns of association between test scores and other variables are investigated, it is often useful to describe the degree of measurement error using a *reliability coefficient*. This is a number ranging from zero to one. It is equal to the proportion of variance in scores that is attributable to true examinee differences, as opposed to errors of measurement.[27] Because the reliability coefficient depends on both error variance and observed score variance, it may change depending on the group tested. Other things being equal, the greater the variability among a group of examinees, the higher the reliability of that group's test scores.

Validity. Evaluation of the quality of information from tests and surveys must of course include more than precision. A precise answer to the wrong question or a distorted answer to the question of interest is of little or no value. Reliability coefficients and standard errors describe score accuracy in narrowly statistical terms, without regard for the meaningfulness or appropriateness of the scores for any particular application. The many considerations beyond statistical accuracy are all included in the general term *validity*. Modern treatments of validity (e.g., Kane , 2006) regard it as a matter of degree, and as pertaining to a use or interpretation of scores, not a test or survey itself. A test or

[27] The terms "true examinee differences," "true scores," and "true score variance" refer here to hypothetical error-free scores representing each examinee's long-run average score across many repeated testings. There is no implication that these scores are "true" in the sense of measuring only what is intended, or all of what is intended. Any influence on an examinee's score that is consistent across repeated testings is automatically part of that examinee's "true score."

survey valid for one use or interpretation may be invalid for another. The evaluation of validity for a test use or interpretation (called validation) is best approached by constructing an *interpretation/use argument* (IUA, see Kane, 2013), which is a set of propositions describing the chain of reasoning justifying a given score use or interpretation, and then investigating the support for each of these propositions. It is not possible to do justice to this complex topic in the space of this chapter, but validity may be supported by evidence based on response processes, the internal structure of the a test, relations to other variables, and in some cases consideration of both intended and plausible unintended consequences of testing, among other considerations (AERA, APA, & NCME, 2014).

Bias. Two major criteria used in evaluating the quality of survey findings are bias and standard error. Bias refers to systematic error. An estimate is biased if, over many imagined replications of the entire survey, the value calculated would average out to a value different from the true value in the population. If, for example, low-scoring students were somewhat more likely to be absent on testing days than high-performing students, then average scores for tested students would tend to be slightly higher than the average for all students. As another example, if principals tended, on average, to report somewhat fewer student disciplinary actions than actually occurred in their schools, then a summary statistic, based on their responses, describing the overall frequency of such actions would be biased. Obviously, if bias could be quantified precisely, then survey findings could simply be adjusted to remove it. Unfortunately, while it is often possible to determine the direction of potential sources of bias, it is rarely possible to eliminate its effects. Special studies may provide indirect evidence concerning specific sources of bias. For example, vigorous follow-up efforts aimed at a subsample of initial nonrespondents may yield data making it possible to compare the overall group of initial nonrespondents to initial respondents. Findings from such a follow-up study can then be used to adjust survey results to reduce nonresponse bias. Survey methodologists have identified numerous potential sources of bias arising throughout the many stages of constructing the sampling frame, drawing samples, collecting data, and adjusting for nonresponse. It is not possible to do justice to this complex topic in the space of this chapter, but Groves, Fowler, Couper, Lepkowski, Singer, & Tourangeau (2009) discuss

various sources of bias and include references to further readings where more detailed treatments are available.

Standard error. Every sample statistic has a standard error, which is typically reported along with the statistic. The standard error quantifies random error arising due to the vagaries of statistical sampling as well as day-to-day variability in individuals' responses. It is expressed in the same units as the statistic it pertains to, and provides an estimate of the standard deviation[28] of the values that would be obtained for that statistic if the entire survey were repeated again and again, with different random samples of respondents. The details of the sampling plan will determine the precise formulas used to calculate standard errors, but in general, if the sample is randomly drawn, then the standard error is closely related to the sample size. Larger samples yield more precise estimates of population values. Other things being equal, if the sample size is doubled, the standard errors of all survey statistics will be reduced by about thirty percent.[29] The standard error is also affected by clustering, stratification, oversampling of subgroups, and other features of complex survey sampling designs. The standard error also reflects the imprecision inherent in each individual respondent's answers to survey questions, that is, the measurement error. However, for purposes of estimating standard errors for surveys, there is often no need to distinguish between imprecision due to sampling and imprecision due to measurement.

Summary. In summary, both random error and bias may be introduced at almost every step in the processes of designing and executing surveys (Groves, et al., 2009, Chs. 3 and 4). Bias and random error arise in both the sampling process and the measurement process. With regard to the sampling process (representation), some segments of the target population may be difficult to reach; cost considerations may limit the size of the sample, and of course a smaller sample will provide less accurate information than a larger sample. Efforts to follow up with nonrespondents may be costly and ultimately only partially successful. Some responses may be recorded incorrectly, and so forth.

---

[28] The standard deviation is a measure of variability in some distribution. It is often reasonable to assume that about two thirds of all values fall within plus or minus one standard deviation of the mean, and about 95 percent of all values fall within plus or minus two standard deviations of the mean. Thus, a 95% confidence interval is often taken to be plus or minus two standard errors, i.e., an interval extending from two standard errors below to two standard errors above an estimated value of a statistic.

[29] More precisely, the standard error will be inversely proportional to the square root of the sample size, so doubling the sample size will reduce the standard error by a factor of one over the square root of two.

With regard to the measurement process, the questions posed may be interpreted differently by different respondents, and the processes by which respondents arrive at their answers may differ from those envisioned by the researchers. People may respond inconsistently to the same question from day to day. Much of the art and science of survey design and use is concerned with the problem of obtaining adequate precision to support intended uses and interpretations.

Design Based Research (DBR)

Design based research refers to a family of methodological approaches (variously called design research, design-based research, design experimentation, development research) in which instructional design and theory development are interdependent. The aim of design research is simultaneously to implement a theoretically grounded instructional approach and to test and refine the theory guiding that implementation. This is done in the course of a small-scale enactment featuring careful monitoring, active reflection, and ongoing revision to both the theory and the planned implementation, as researchers learn in active collaboration with the participants in the instructional setting.

Our first key theoretical resource is an article by Cobb, McClain, and Gravemeijer (2003) in which they describe and illustrate their methodology for enacting this form of research in classrooms (i.e., conducting a classroom design experiment). We also draw upon a brief article by the Design-Based Research Collective (2003), which sets forth some features common to design experiments by various investigators. Although early applications of design research methodology were used to study student learning in classrooms, DBR methods have also been used in studying teacher learning and problems of implementation at multiple levels of the education system. Cobb, Jackson, Smith, Sorum, and Henrick (2013) report on the first five years of a planned eight-year study conducted in partnership with four urban school districts implementing middle-grades mathematics instructional reforms. In that project, they are collaborating with district leaders in accordance with the tenets of DBR. We have chosen for our illustrative example a design experiment reported by Cobb, Zhao, and Dean (2009) that focused on the learning of a group of middle grades mathematics teachers in five schools within the same district as they sought to implement research-based instructional practices.

Aims and Conceptions of Social Phenomena

The Design-Based Research Collective (2003) outlines the core characteristics of design based research as follows:

> First, the central goals of designing learning environments and developing theories or "prototheories" of learning are intertwined. Second, development and research take place through continuous cycles of design, enactment, analysis, and redesign…. Third, research on designs must lead to sharable theories that help communicate relevant implications to practitioners and other educational designers…. Fourth, research must account for how designs function in authentic settings. It must not only document success or failure but also focus on interactions that refine our understanding of the learning issues involved. Fifth, the development of such accounts relies on methods that can document and connect processes of enactment to outcomes of interest. (p. 5)

In design research, "instructional design and research are interdependent. On the one hand, the design of learning environments serves as the context for research; on the other hand, ongoing and retrospective analyses are conducted in order to inform the improvement of the design" (Cobb, et al., 2009, p. 169). Thus, the aims of the design experiment include both successful implementation and advances in theoretical understanding: "As is the case in all design experiments, our primary research objective when working with the teachers was to improve the design for supporting learning that we had formulated at the outset. To this end, our overriding objective while collaborating with the teachers was to test and revise our conjectures about both the trajectory of the teacher learning and the specific means of supporting and organizing that learning … [through] successive cycles of design and analysis" (Cobb, et al., 2009, p. 170).

The researchers' conception of teaching and learning is reflected in their statement that, "Our overall goal for the experiment was that teaching would become a collaborative, knowledge-generating activity in which the teachers would come to place students' reasoning at the center of their instructional decision making …. More concretely, we attempted to support the teachers' development of forms of instructional practice in which they would adapt, test, and revise instructional materials and

approaches that had proven effective elsewhere to the contingencies of their specific classroom settings by focusing on their students' interpretations and solutions" (p. 169).

They began from the premise that the work of the teachers with whom they were collaborating could only be understood within its institutional setting, including "district and school policies for instruction, … both the adoption of curriculum materials and guidelines for the use of those materials, … the people to whom teachers are accountable and what they are held accountable for, [and] social supports that give teachers access to new tools and forms of knowledge" (p. 166). Indeed, they held that, "teachers' instructional practices are not merely influenced by but are *partially constituted* by the materials and resources that the teachers use in their classroom practice, the institutional constraints that they attempt to satisfy, and the formal and informal sources of assistance on which they draw" (p. 166).

Invoking a distinction made by Wenger (1998) among others, Cobb, et al. (2009) conceived of the institutional setting as the *lived organization* experienced by teachers and other participants in the setting, not the *designed organization* "that is inscribed in official organizational charts, policy statements, and manuals" and "includes formally designated roles and divisions of labor together with official policies, procedures, management systems, organizational units, and the like" (p. 175). While documenting the organizational context of the teachers' work, they identified three distinct groups within the district, each "pursuing agendas for how mathematics should be taught and learned. [These were] the district mathematics leadership group, the leadership group in each of the schools in which the teachers worked, and the teacher group" (pp. 175-176). Alignment of the goals and understandings of these three groups, and the development of brokers who could move across groups to improve communication, emerged as goals in the course of this particular design experiment. Another key goal was to support the evolution of the group of teachers into a learning community.

A 2013 *Yearbook of the National Society for the Study of Education* was devoted to Design Based Implementation Research, which in common with DBR is based on "a focus on persistent problems of practice from multiple stakeholders' perspectives [and] a commitment to iterative, collaborative design" as well as "a concern with developing theory and knowledge related to both classroom learning and implementation through

systematic inquiry" (Fishman, Penuel, Allen, Cheng, & Sabelli, 2013, p. 136). Design Based Implementation Research is also concerned with "developing capacity for sustaining change in systems" (p. 136). In applications like the Cobb, et al. (2013) collaboration with leaders in four urban school districts, Design Based Implementation Research creates "a 'middle ground,' a space between practice on the one hand and policy on the other … [which helps to promote] a collaborative culture with shared processes and responsibilities rather than placing the onus for change solely on individuals" (Sabelli & Dede, 2013, p. 465).

Design and Enactment

Preparation. Cobb, McClain, and Gravemeijer (2003) divide the design experiment into three phases—preparation, carrying out the experiment, and retrospective analysis—which may be repeated over successive instructional cycles. The preparation phase begins with an "anticipatory thought experiment," in which prior research and theory are used to envision a hypothetical learning trajectory together with "specific means that might be used to support and organize that learning" (p. 2). For Cobb, et al. (2009), this theory was based on their own earlier design experiments at the classroom level, on the teaching of similar content.

Carrying out the experiment. Carrying out the design experiment involves active, ongoing adjustment and revision of the hypothesized learning trajectory. When design experiments are carried out in classroom settings, "all members of the research team meet after each classroom session to discuss their interpretations of classroom events and to propose possible modifications to the learning trajectory" (Cobb, et al., 2003, p. 3). Thus, the planned intervention evolves through *minicycles*, which are "tightly integrated cycles of design and analysis" (p. 3).

In carrying out this five-year experiment with teachers, the researchers began early in the first year by documenting teachers' perceptions of their institutional context via audio-recorded interviews, then using a snowballing process, interviewing formal and informal instructional leaders identified by the teachers at both the school and district levels, and then additional persons identified in these second-round interviews. It was through this process that the three groups of actors with their distinct agendas were identified. An unexpected discovery was that the process of interviewing teachers about

their own support networks, resources, and institutional constraints proved in itself to be a powerful aid in building trust and rapport. However, while the researchers "were eventually able to negotiate a productive relationship with the teachers in which authority became increasingly distributed …, this was a protracted process" (p. 173).

One distinctive feature of this design experiment, focusing on teacher learning rather than student learning directly, was the need to conceptualize the relations between teachers' activities in the distinct settings of their classroom teaching versus their professional development activities (work sessions) with the researchers (p. 168). Thus, data collected in carrying out the experiment included video recordings and copies of artifacts produced in all of the work sessions the researchers conducted with the teachers, as well as "teaching sets" collected at least twice per year in each teacher's classroom. Each "teaching set" consisted of a classroom observation with an accompanying video recording, as well as an audio-recorded structured interview in which the teacher was debriefed about the lesson.

The researchers describe and analyze occasions when work sessions did not go as planned. For example, a work session activity during the third year of the study involved teachers bringing in student work from a common instructional activity they had each conducted in their classrooms. By this time a level of rapport and trust had been established and the teachers' instructional practices had been deprivatized, so the researchers expected that the student work samples would foster teacher engagement with different patterns of student reasoning. They reported, however, that while teachers found the activity engaging and discussed their students' work openly, they nonetheless failed to see its relevance to their own classroom instruction. This unexpected finding led the researchers to reconceptualize the relation between the teacher professional development setting and the classroom setting:

> We found Beach's (1999) notion of *consequential transitions* particularly useful as we attempted to rethink the relations between teachers' activity in the two settings. … This perspective implies that design conjectures should not merely focus on the movement of artifacts between the two settings [but should instead] anticipate the ways in which teachers might use artifacts in professional

> development sessions given how they use them in their classroom practices. (p. 188)

Retrospective analysis. As explained by Cobb, McClain, and Gravemeijer (2003), the primary goal while carrying out the design experiment is typically pragmatic, focused on supporting participants' learning. It is during the retrospective analysis phase that attention turns to the goal of developing "a domain-specific instructional theory that can feed forward to guide instruction in other classrooms" as well as "guide the formulation of a new hypothetical learning trajectory for a follow-up design experiment." This is the "local instructional theory" that makes the findings of the design experiment potentially generalizable, "by means of an explanatory framework rather than by means of a representative sample" (Cobb et al., 2003, p. 3).

Similarly, Cobb, et al. (2009) explain:

> One of our primary goals in conducting retrospective analyses of these data is to document the actual learning trajectory of the teacher group over a 5-year period and, on this basis, to propose a revised conjectured learning trajectory for a teacher group that is situated in the institutional setting of the school and district. These analyses delineate the process of the group's collective learning, the means by which it was supported, and the institutional setting in which it was situated. … The revised trajectory … encapsulates what we learned in the course of the teacher development experiment and indicates how it might be possible to accomplish in 3 years what took 5 years in the teacher development experiment" (pp. 171-172).

In support of this goal, Cobb, et al. (2009) go on to provide just such an instructional theory, setting forth a framework for documenting the collective learning of a teacher group in terms of four types of norms, including an unanticipated category of *institutional reasoning* norms that emerged as significant in their research.

Reporting and Evaluation

The Design-Based Research Collective (2003, p. 6) cites and rejects a suggestion that design research might be used to generate causal accounts that could then be tested in rigorous randomized trials. They argue against this view on several grounds, notably that each design research experiment is aimed at a deep understanding of learning in a

specific local context. In contrast, "randomized trials may systematically fail to account for … phenomena that are contextually dependent or those that result from the interaction of dozens, if not hundreds, of factors" (p. 6). They go on to state, however, that "design-based research can generate plausible causal accounts because of its focus on linking processes to outcomes in particular settings, and can [assist] in the identification of relevant contextual factors [and] of mechanisms (not just relationships), [as well as] enriching our understanding of the nature of the intervention itself" (p. 6). (Alternative conceptions of generalization are addressed later in this chapter.)

The Design-Based Research Collective (2003, p. 7) observes that "Objectivity, reliability, and validity are all necessary to make design-based research a scientifically sound enterprise, but these qualities are managed in noticeably different ways than in controlled experimentation … [relying instead] on techniques used in other research paradigms, like thick descriptive datasets, systematic analysis of data with carefully defined measures, and consensus building within the field around interpretations of data. … In particular, design-based research typically triangulates multiple sources and kinds of data to connect intended and unintended outcomes to processes of enactment." They caution that, "design based research can generate usable knowledge about educational practice, but even usable knowledge will not make complex educational problems simple," and go on to "stress that design-based research should not become a euphemism for 'anything goes' research or oversimplified interventions." Rather, DBR should be viewed as a disciplined process of research in situ, which fully embraces the complexity of learning environments.

Social Network Analysis (SNA)

As a methodology, social network analysis (SNA) is neither variable-centered nor case-centered, focusing instead on the interdependence among individuals in groups or organizations, and only secondarily on the individuals themselves. The objects of study in SNA are fundamentally relational. "Of critical importance for the development of methods for social network analysis is the fact that the unit of analysis in network analysis is not the individual, but an entity consisting of a collection of individuals and the linkages among them" (Wasserman & Faust, 1997, p. 5). The individuals ("actors") may be persons or organizations; the linkages ("ties") may be friendships, flows or

exchanges of information or material resources, or reporting relationships within an organization, for example. The origins of SNA may be traced back to Moreno's invention of the *sociogram* in the 1930s (Moreno, 1953). The sociogram is a graphical representation in which each person in a group is represented as a point, and relationships between pairs of persons are represented by lines connecting pairs of points. SNA still makes extensive use of graphical displays, some far more sophisticated than these early sociograms.

Formal methods for SNA draw upon theoretical work in mathematics, especially graph theory, and in statistics. This methodology has been used to address problems within various social science disciplines, including anthropology, social psychology, economics, and political science, but our key theoretical resources are drawn from sociology. The first, Wasserman and Faust (1997) is an updated version of Wasserman and Faust's classic compendium *Social Network Analysis: Methods and Applications*, first published in 1994. The second, Carrington, Scott, and Wasserman (2005), is an edited anthology describing advances in a range of specific topics and techniques in the field.

The first of our illustrative cases for this methodology is an investigation by Penuel, Riel, Krause, and Frank (2009). This was an exploratory study using SNA as well as interview methods to investigate teacher communities in two elementary schools in California. These two schools served similar student populations and had similar resources, and both were designated "underperforming schools" at the beginning of the study. In both schools, principals identified collegial interaction as a key factor for enhancing literacy instruction. Three years later, however, student achievement in one of the schools had increased dramatically, whereas in the other school, achievement was essentially flat. SNA revealed striking contrasts between the teacher networks in the two schools, which supported a plausible account of some reasons for the observed outcome differences, in terms of differential access to expertise and uptake of innovative teaching practices, as well as differences in teachers' mutual trust and respect.

Our second illustrative case is a study by Spillane and Kim (2012), which used SNA to investigate the positioning of formal school leaders within teacher networks in each of the 30 elementary schools in a mid-size urban school district. "Instructional

leaders" included the principal (a full-time school leader) as well as instructional coaches and teachers with other formal mentoring roles (most of whom were part-time formal school leaders). In addition to documenting commonalities and differences among the schools, their analysis showed that part-time instructional leaders were, in general, more centrally situated in networks and more frequently sought out for advice than were principals.

Aims and Conceptions of Social Phenomena

Researchers in this methodological tradition seek to explain individual behavior and social phenomena in terms of organizational structures, including but certainly not limited to formal organizations. The way in which an individual is situated within a network may influence or constrain that individual's informational and other resources, possible actions, or even attitudes and values. Various kinds of social ties may be considered in SNA, including "evaluations of one person by another …, transfers of material resources …, association or affiliation …, [or] behavioral interaction …," as well as "movement between place or statuses …, physical connection …, formal relations (for example, authority), [or] biological relationship …" (Wasserman & Faust, 1997, p. 18). The structures modeled may be stable or may change over time. Networks typically involve actors of only one sort (e.g., school staff), but a *two-mode network* would include actors of two sorts (e.g., school staff and students). Likewise, the ties between actors may be of just one kind or of several kinds. Furthermore, ties may simply be present versus absent for each dyad (i.e., either connected or not), or may be a matter of degree (i.e., variable tie strength). Ties may be directional or not. Thus, it may be seen that the formal machinery of network analysis is varied and highly flexible.

Wasserman and Faust (1997, p. 17) emphasize that, "[SNA] is far more than an intuitively appealing vocabulary, metaphor, or set of images for discussing social, behavioral, political, or economic relationships. Social network analysis provides a precise way to define important social concepts, a theoretical alternative to the assumption of independent social actors, and a framework for testing theories about structured social relationships." Because the interdependence among individuals is of explicit concern in SNA, statistical methods that ignore this interdependence, or that assume statistical independence, are inappropriate for network analysis. A rich body of

theory and statistical methods has been developed for SNA, supported by special-purpose computer software (Huisman & van Duijn, 2005).

Wasserman and Faust (1997, p. 9) also contrast SNA per se with studies in which "network measurements, or statistics calculated from these network measurements" are used as variables characterizing individual actors.[30] While recognizing that such variables may be useful as "explanatory factors or variables in understanding individual behavior," they emphasize that incorporating such variables into analyses that use individuals as the basic modeling unit is not the same as SNA.

Design and Enactment

Having identified a research question for which SNA might be useful, the researcher must first specify the network or networks to be examined. This entails first, selecting the site or sites of the investigation and second, delineating the boundaries of the networks to be studied within each site. Site selection might be accomplished by first specifying a population (e.g., some set of classrooms, schools, or school districts) and then drawing a statistically representative sample. Alternatively, the researcher might sample purposively, following a logic similar to that of other methodological traditions discussed in this chapter. Once research sites are chosen, the boundaries of each network must be specified. If a researcher is interested in teachers' support networks, for example, a purposive decision must be reached as to whether resource personnel outside the school building will be included in the analysis. In many cases, administrative units such as schools will suggest natural boundaries for networks, but in the real world of schooling, such boundaries may not be entirely clear. "Actors may come and go, may be many in number and hard to enumerate, or it may be difficult even to determine whether a specific actor belongs in a set of actors" (Wasserman & Faust, 1997, p. 31). In Penuel, et al. (2009), for example, across the two schools, a computer teacher, an inclusion specialist, some arts teachers and pullout program staff members, as well as a few regular classroom teachers, failed to respond to surveys and were omitted from the network analyses. In some situations, the researcher might need to consider whether network

[30] The study by Coburn, et al., (2012), which served as one of our illustrative case studies in the discussion of Small-N Comparative Case Study (CCS) research, used concepts from SNA in this way, employing measures of "tie strength" and other attributes of individual teachers' social support networks as explanatory variables.

boundaries should be defined according to the perceptions of the actors themselves (a *realist* approach) or according to the theoretical concerns addressed in the research (a *nominalist* approach). Regardless, Wasserman and Faust (1997, p. 32) emphasize that, "Somehow, in order to study the network, we must enumerate a finite set of actors to study."

Network analysis is generally most straightforward if data are available on linkages between all pairs of actors in a network, but techniques have been developed for inferring the properties of networks using information from only a sample of actors. This literature generally focuses on the estimation of various network properties (e.g., average number of ties per actor, or technical attributes like degrees of reciprocity, density, or transitivity). Depending on the situation, the units sampled may be individual actors, pairs, triples, or larger groups of actors. "Snowball sampling," in which an initial set of actors (the "first-order zone") list others with whom they have ties of a certain kind and these "second-order zone" actors are then contacted to elicit their ties, and so forth, is perhaps the best known of various special-purpose "chain methods" of sampling designed specifically for network analysis. Further citations to the extensive literature on this topic are provided by Wasserman and Faust (1997, ch. 2).

Once actors are identified, the next step is network measurement. Typically, information is obtained by observing, interviewing, or questioning actors about their ties to other actors in the set.[31] The SNA researcher must often attend to the same sorts of limitations of self-report data as researchers in other traditions, including imperfect memory, interview and interviewer context effects, and various potential informant biases (Marsden, 2005). On a survey asking actors to list their friends or rank-order a set of people in terms of closeness of friendship, for example, different understandings of the word "friend" may affect responses. Test-retest studies and other methods for quantifying reliability and investigating validity of network measurements may be employed.

For Spillane and Kim, for instance, within each school, advice and information networks were constructed based on two waves of staff surveys, administered in spring of

[31] Note that with chain sampling methods, determining the set of actors and measuring their ties would happen simultaneously.

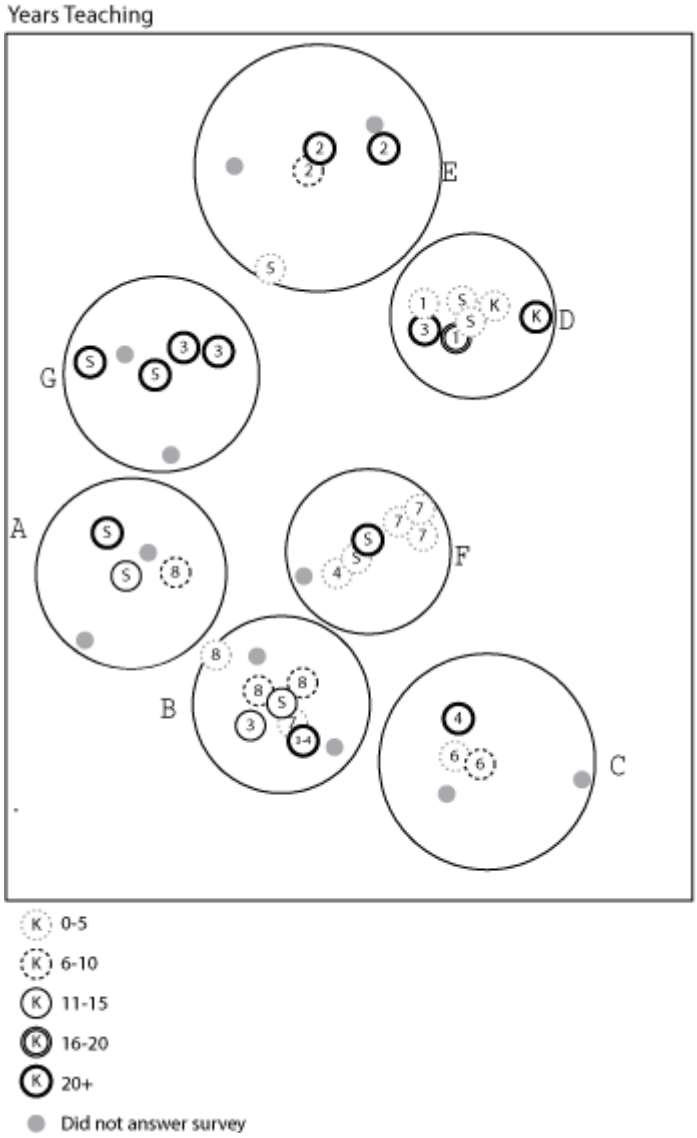
2005 and 2007. Survey respondents indicated whom they turned to for advice and information, separately for reading/language arts and for mathematics. Three measures were derived for each staff member. *Degree centrality* refers to counts of the number of other actors to whom that person turned for advice (out-degree centrality) and the number of other individuals who turned to that person for advice (in-degree centrality). *Betweenness* is a statistical summary of the degree to which an actor links or stands between other pairs of actors. *Closeness* is related to degree centrality, but serves as a more global measure of connectedness within the network. Each of these variables is precisely defined and explained in the article (Spillane & Kim, 2012, pp. 79-80).

Penuel, et al. (2009) relied on staff surveys to construct their network. In addition to collegial ties, respondents were also asked about their access to different kinds of resources and expertise, participation on committees or teams where reform was discussed, and perceptions of collegiality in the school as well as the degree to which reforms were being implemented. The researchers also interviewed six informants in each school, including school leaders and instructional coaches. As noted above, Spillane and Kim (2012) constructed school networks using a more limited set of questions from district-wide staff survey data. As with other research methodologies, finite resources may enforce a tradeoff between the number of sites studied and the richness of data collected at each site.

A range of techniques are available for analysis of network data. Penuel, et al. (2009, pp. 135-136), for example, describe key steps in their measurement and analysis of collegial ties as follows:

> Teachers were asked to identify the people they considered to be their closest professional colleagues using a numbered roster provided them with the names of all the staff …. For each colleague selected, the respondents were asked to indicate the frequency of interaction …. These responses were used to identify cohesive subgroups and then embed subgroup boundaries in a "crystallized" sociogram (Frank & Yasumoto, 1998; Frank & Zhao, 2005). These responses also form the basis for two measures we report on here in our analyses: (1) weighted in-degree … and (2) weighted out-degree.

The "crystalized sociogram" for one of the two schools appears as follows:



(Penuel, et al., 2009, p. 146).

In this figure, each small circle represents a teacher, and that teacher's grade level (or range of grade levels) is indicated by the number inside the circle. The darkness of the circle indicates years of teaching experience. Resource teachers and instructional leaders are represented by gray dots. The larger circles represent subgroup boundaries, and the smaller the circle, the more cohesive the group. Within the circles, teachers situated closer together interact more. Interpreted in the light of interview data and comparison to the school's formal ("normative") structure, and contrasted with the corresponding diagram for the second elementary school studied, this crystalized

sociogram proves to be a highly informative interpretive tool, illuminating the social structure of the staff within the school.

Spillane and Kim (2012) also analyzed teacher subgroups. In addition to deriving measures of degree centrality, betweenness, and closeness for each staff member, they partitioned the staff networks in each of their 30 elementary schools into nonoverlapping subgroups of teachers, so as to analyze the locations of different kinds of school leaders within subgroups:

> To identify formal school leaders' membership in and distribution across subgroups, we identified nonoverlapping subgroups of language arts and mathematics networks in 2005 and 2007 using the network clustering algorithm software KliqueFinder (Frank 1995, 1996…). In addition, we calculated the proportion (%) of language arts and mathematics subgroup members and floaters (who could not be assigned to any subgroup) in each school. Finally, we compared proportion (%) of formal leaders' subgroup connection and floaters across schools. (Spillane & Kim, 2012, p. 81)

The researchers then tabulated distributions of degree centrality, betweenness, and closeness for full-time leaders, part-time leaders, and other teachers, as well as distributions of subgroup members, floaters (staff who could have been assigned to more than one subgroup), and isolates (members of no subgroup), separately for English language arts and for mathematics. Taken together, these various tabulations offered an informative summary of the ways different kinds of instructional leaders were situated in the social networks of the 30 schools.

Reporting and Evaluation

There is no one set of criteria by which to judge the soundness of a study using SNA methods. SNA researchers must attend to the accuracy of the information used to construct networks as well as its validity—Do measures of ties in fact capture what they are purported to measure, reasonably free of bias and distortion? Once actors and ties have been measured, various analyses of the network are possible. One common goal is to partition the actors into subgroups. There is nearly always some arbitrariness in such a partitioning, however. One is unlikely to find a network in which actors are neatly separated into subgroups such that all pairs within subgroups are connected and no pairs

across subgroups are connected. Consequently, various criteria have been developed for defining subgroups, and different criteria may yield different partitions. Some software tools (e.g., KliqueFinder, Frank, [reference]) provide goodness of fit tests that are helpful in evaluating the quality of the analysis they provide.

The findings of both studies had implications for the design of organizational environments that support teachers learning. For Penuel et al, SNA demonstrated that the internal structure of the school communities they studied helped explain differences in access to resources and expertise. For Spillane and Kim, the SNA identified, for their schools, differences in the relative centrality and role of different school leaders:

> Our analysis suggests that, although the school principal is not a central actor in the instructional advice and information network in a majority of these schools, formally designated school leaders as a group do occupy central positions. Examining formal school leaders' network and subgroup membership as well as their distribution across subgroups in schools, our account suggests that full-time and especially part-time formal school leaders play important roles in how elementary schools organize for instruction, brokering relations among staff and as representatives of the formal organization in subgroups. (p. 73)

Generalization in SNA may take various forms. In their study of just two schools (two cases), Penuel, et al. (2009, p. 126) did not directly address the generalizability of their conclusions, but asserted that "As we argue from these cases, a network perspective on teacher community offers a powerful lens for analyzing how each school's practices for facilitating the flow of expertise and resources across teachers and classrooms led to these different [changes in achievement over time]." Thus, as with theoretical generalization from ethnographic case studies, some SNA researchers may look to generalization "based on the development of a theory that can be extended to other cases or refined in light of them" (Eisenhart, 2009, p. 59). Spillane and Kim (2012, p. 77) state directly that, "We make no attempt to generalize beyond the schools in our study." From their thoughtful discussion of the roles of full-time versus part-time instructional leaders, however, it appears that they are merely leaving it to the reader to consider the applicability of their theoretical insights and empirical findings in other situations. In their final paragraph, they cautiously assert that, "if our analysis is roughly right, albeit

based on only 30 schools in one urban school district, educational policy makers and school principals should pay careful attention to who occupies formally designated leadership positions in their schools" (Spillane & Kim, 2012, pp. 96-97). Statistical methods and procedures have been developed for formal hypothesis testing within SNA, and statistical generalization within SNA is an area of intense current interest. As one recent example, McFarland, Moody, Diehl, Smith, and Thomas (2014) use these methods to test hypotheses concerning the importance of various contextual factors on adolescent social structures in high schools. The authors generated a large number of random graphs according to models representing different large-scale contexts, then summarized the properties of these graphs to demonstrate that the modeled contextual factors did in fact show the hypothesized effects on the small-scale properties of the simulated social networks.

Complex Adaptive Systems Research (CAS): Agent Based Modeling (ABM)

Work within the tradition of complex adaptive systems (CAS) research is relatively new to the field of education research and not widely practiced. While histories of this diverse tradition vary in terms of the theoretical precursors on which its historians draw, it is generally represented as first being developed within mathematics, computer science, and the natural sciences, like physics and biology, and later within more socially oriented disciplines and fields, including economics, linguistics, organizational theory, developmental psychology, and education (Byrne and Callaghan, 2014; Larsen-Freeman and Cameron, 2008). It has become a supra- or transdisciplinary approach that can and has been applied to many different sorts of systems ranging in scale from the human brain and immune systems to the global economy and the world-wide web (Mitchell, 2009). Miller and Page (2007) characterize this transdisciplinarity as the sort of "emergence" of system-level phenomena, absent a knowing designer, that CAS research is well suited to describe. Interdisciplinary collaborations, like the seminal Santa Fe Institute (www.santafe.edu) or the Center for the Study of Complex Systems at the University of Michigan (www.lsa.umich.edu/cscs) bring scholars together across disciplinary boundaries to explore synergies and commonalities across studies of different kinds of complex adaptive systems. Most if not all of the social phenomena with

which we are concerned in this chapter can be productively framed in terms of complex adaptive systems.

In this section, we focus on one approach to the study of complex adaptive systems: computational modeling and, more specifically, agent-based modeling (ABM), where a key goal is to use computer simulations to understand how system-level phenomena can evolve in light of interactions among individual agents or groups of agents. Here, our primary theorists will be Miller and Page (Miller and Page 2004, 2007; Page , 2010) and Axelrod and colleagues (Axelrod and Cohen, 2002; Axelrod and Tesfatsion, 2006), who provide accessible introductions to computational modeling with complex adaptive systems. Our empirical example will be Maroulis, et al. (2010a), who model the system level effects of a school district's shift in policy from neighborhood-based school assignment to school choice by household. In a subsequent section on methodological pluralism, we'll follow the lead of education (Davis and Sumara, 2005; Morrison, 2002; Osberg and Biesta, 2010) and social (Byrne and Callaghan, 2014; Larsen-Freeman and Cameron, 2007; Morçöl, 2012) researchers who draw more broadly on "complexity theory" as a metaphor, way of thinking, or frame of reference to analyze complex adaptive systems to support multi-methodological research.

Aims and Conceptions of Social Phenomena

A complex system is one that is composed of multiple heterogeneous agents who interact with one another over space and time. Agents—in our case mostly humans or groupings of humans (like families, schools, and government agencies)--are entities that can react to their environments and to one another. A complex system becomes "adaptive" when the "interaction of the agents is contingent on past experience and…[the] agents continually adapt to that experience (Axelrod & Testfatsion, 2006, p. 1649). Complex adaptive systems (CASs) can produce "emergent properties"—system level behavior—that "cannot be deduced simply by aggregating the properties of the agents" (Axelrod & Tesfatsion, 2006, p. 1649). Understanding how emergent properties are produced requires tracing interactions among agents across space and time. Common examples of CAS cited in pedagogical texts involve bird flocks, insect colonies, traffic jams, and standing ovations, each of which can "self-organize" to exhibit patterns of behavior at the group level without an organizer or designer. Agent Based Modeling

(ABM) is one tool for understanding how system level patterns emerge from interactions among agents over time.

Teachers and their students are embedded in a range of CAS, including classrooms, schools. Students are embedded in CAS including families, neighborhoods, and peer groups (face to face and on line). Teachers are embedded in professional organizations, and learning communities. Both are influenced by still others, including local and state school boards and education agencies, textbook publishers, preparation programs, state legislatures, research traditions, and the world wide web. An instructive CAS example used by Miller and Page (2004, 2007) to introduce students to CAS research is the "Standing Ovation Problem" (2004) where the goal is to model whether a standing ovation (system level phenomenon) will emerge from the actions and interactions of individual members of the audience following a lecture. Because it captures an evolving CAS in a constrained space and time, it is useful for illustrating concepts. We will refer to it during the introduction, but then turn to the Maroulis et al. (2010a) school choice policy example in the sections on design and reporting of ABM.

Understanding CAS involving humans is further complicated because of the nature of human agency. Human agents have differing goals, strategies for pursuing them, and criteria for success that can change over time. They "predict and react to the actions and predictions of other agents"; they "continually make choices, either by direct cognition or reliance on stored (but not immutable) heuristics, about their actions" (Miller & Page, 2007 p. 10). Adaptation occurs as agents observe others and receive feedback on their actions. Variations can also develop through error, through migration (when agents encounter new agents), and through combining of strategies (Axelrod and Cohen, 2002, p. 2). Further, "The various connections inherent in social systems exacerbate these actions as agents become closely coupled to one another" (Miller & Page, 2007, p. 10). This sort of "social contagion" can "drive multiple social processes" (Miller & Page, 2007 p. 14).

The standing ovation problem illustrates a number of these features of CAS with human agents. Miller and Page (2004) acknowledge that if the decision to stand is simply a personal choice based on the perceived worth of the lecture, the problem is trivial. However, they suggest, "A seated audience member surrounded by people standing

might be enticed to stand, even if he hated the performance. This behavior mimicry could be strategic (the agent wants to send the right signal to the lecturer), informational (maybe the lecture was better than he thought), or conformal (he stands so as not to feel awkward)…. These peer effects…set the stage … for interesting dynamic behavior" (pp. 1- 2). Location and timing shape audience behavior as well. For instance: "the agents seated in front row can communicate their information best", but "have the worst information about what others are doing" (p. 4). To model this system of interacting agents over time and space, Miller and Page suggest researchers consider how sophisticated the agents are, how they influence one another, how information spreads, in what order events occur, at what time scale events occur, and so on. Miller and Page (2004) note additional issues that researchers might consider such as the architecture of hall (does the presence of a balcony matter?) and the policy implications. Suppose one could place shills in the audience; if one wanted to influence the likelihood of a standing ovation, where should the shills be placed?

With CAS, then, causality can flow in multiple directions, from agent to agent as they interact, from interacting agents to phenomena at higher levels of the system, from pre-existing social and physical structures and emerging phenomena back to the interacting agents. These causal influences occur on different time scales. Further, the behavior of complex adaptive systems is often non-linear. Small shifts in the interactions among agents can cascade into large effects (e.g., a traffic jam), or a system can maintain stability (equilibrium) across substantial disruptions in interactions among agents (e.g., students and teachers can come and go and yet a school system can maintain stability across these shifts). These features of CAS are often cited in critiques of more widely used and familiar approaches to social science, like experimental and quasi-experimental design where the goal is to arrive at generalizations describing individuals in a population. Rather than focusing on averages and commonalities, CAS researchers attend to the particularities and potentialities of individual systems—variation that in other methodological traditions might be relegated to an error term (Page, 2010). This sort of complexity makes difficult both prediction and the design of interventions to improve systems. "Self organized systems can produce cooperative, robust outcomes, but they can also spiral into chaos. We need to understand how to encourage the former

and guard against the latter." (Page 2010, p. 3). "The challenge and the opportunity for social science are to build in the robustness necessary to limit the damage of large events and to harness complexity to produce better social outcomes." (Page, 2010, p. 3).

With CAS research, the goal is to find out "when and why productive systems emerge and how they can persist" (Miller & Page, 2007, p. 7)). Computational modeling allows us to see "changes in outcomes…as natural progressions in a dynamic process" (Page, 2010, p. 3). It lets us see how social outcomes can emerge from the bottom up—how systems can "self-organize"[32]. While this goal can and has been accomplished through ethnographic and discourse analytic work (e.g., Agar, 2004; Larsen-Freeman & Cameron, 2008), computational modeling, including ABM, is an approach widely used by CAS researchers that allows efficient exploration of possible system level outcomes under varying conditions to support understanding, planning and decision-making. Beyond the understanding and explanation of system behavior**,** study of complex adaptive systems has been used to generate learning games, to design and anticipate the effect of new policies, and to manage organizations, including educational organizations. The study of CAS provides a broader range of possibilities for managing a system and leveraging change. There are multiple ways to produce emergent functionalities including, for instance, judicious placement of agents, enhancing diversity, considering combinations of interventions that individually may make a small difference but collectively result in substantial improvements.

While understanding how standing ovations emerge is not a particularly consequential social problem (except perhaps for lecturers and performers), Miller and Page (2007) note how it serves as a metaphor for a variety of social situations "in which agents make binary decisions and interact spatially"—including whether to send children to public or private school, violate the law, riot, search for jobs, retire, vote, experiment with drugs and so on. Our school choice example reflects one such situation.

The purpose of the study in Maroulis et al. (2010a) is to evaluate the impact of a within-district policy initiative: transitioning from a neighborhood-based system of assigning pupils to schools to a system based on students' (families') choice among

[32] Some complexity theorists use self-organization and emergence interchangeably, some define the terms differently. See Morçöl (2012) for an explicit discussion of this issue. For our purposes in this chapter, we will not attempt to distinguish them.

public schools in the district. The agents are students and schools; the district level outcomes considered are average achievement, the treatment effect of choice, and school survival. The questions focus on how (micro) interactions among students' households and schools (with differing attributes) result in these (macro) district level outcomes. The authors used data from the Chicago Public schools to set initial conditions of the model. They explore the impact of different participation rates (exercising the option to choose) and school capacity constraints on the treatment effect (average achievement of choosers versus non choosers). The authors also modeled how different assumptions about agent choice strategies—based on schools' overall achievement, value-added, or proximity—led to different system level outcomes.

Design and Enactment

As Axelrod & Tesfatsion (2006) describe it, "ABM begins with assumptions about agents and their interactions and then uses computer simulation to generate 'histories' that reveal the dynamic consequences of these assumptions" (p. 1650). Miller and Page (2007) liken models to maps. Good maps have just barely enough details to allow us to address the questions we care about, but leave out unnecessary details. ABMs typically consist of agents with varying characteristics, an environment (often geographic sites and links among agents), and rules for the behavior of agents, their environment, or their interactions (Maroulis et al, 2010a). In the standing ovation example, the agents are audience members with varying preferences for the lecture and access to information about others' behavior; the environment is the lecture hall (with its presumed structural features mediating the flow of information); and the rules specify when an agent stands based on a combination of available information and individual preference. In the Maroulis et al. example, the agents are students' households and schools; the environment is a district with different neighborhoods containing schools; and the rules describe the behavior of students and schools (elaborated below).

The hypothesized scenario underlying the simulations in Maroulis et al. (2010a) involves a district containing multiple high schools with incoming 9th grade students and graduating 12th grade students leaving the number of students in the district equal from year to year. The students are simulated to differ in test-score-based academic achievement and geographic location. Schools are located in neighborhoods simulated to

differ in average academic achievement and socioeconomic status. At the beginning of a simulation, a catchment model is assumed to be in place such that all students are assigned to schools based on geographic proximity. Across successive iterations (school years), some proportion of incoming 9th graders is designated to choose which school to attend. Depending on the model, and the simulation within the model, students' choices are simulated based on some weighting of school proximity and average school achievement; and acceptance by the chosen school is based (in most simulations) on the school's capacity. After each iteration, schools' student enrollment and average achievement are updated, as are individual students' achievement (based on value added of school attended). The collective outcome variables considered (emergent phenomena) are the district's average achievement, school closings due to low enrollment, and the treatment effect of school choice (i.e., the difference in average student achievement between choosers and non-choosers).

The authors try out three models[33], differing in complexity of factors considered and the extent to which the initializations and updates (of students and schools) are informed by actual achievement, capacity, and geographic data from the Chicago Public Schools CPS. We focus here on the second model, where the authors' representation is most elaborated. In this model, the authors "ensured that the initialization of the model parameters, such as student achievement and school value-added, corresponded with a set of reasonable, real world values, in this case, data on students and schools from Chicago Public Schools open enrollment program during the years of 2000 – 2003" (p. 11).

For each iteration of the second model, incoming 9th graders (n=5000) each year are simulated based on characteristics of cases randomly sampled from the 17,131 incoming high school student data from the CPS in 2002-03. The data include home census block, eighth grade achievement test score, concentration of poverty in census block, and assigned school under the catchment model. Similarly, schools are assigned attributes based on CPS data including geographic location, total students enrolled, mean achievement of all students, value added estimated by the author (once for each school),

---

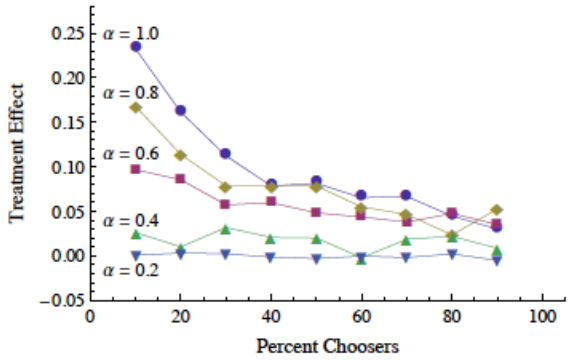
[33] Maroulis et al. (2010a, p. 6) report: "all models in this paper were created using the NetLogo Programmable Modeling Environment (Wilensky , 1999). Net logo comes with an interface to the Mathematica scientific computing environment (Wolfram Research Inc., 2007) which was used to collect and analyze batches of simulations run under different parameter settings."

and maximum enrollment capacity.[34] Some parameters are allowed to vary across successive simulations of the model: for Model 2 these include the percent of incoming students designated as choosers and the criteria students use in making their choices. The rules specify that schools close if they drop below a certain proportion of their capacity. Once initialized under a catchment model, the simulated agents follow the same set of rules for making choices across the 20 iterations (simulating 20 years) of each model. A third model, to which we'll refer briefly below, complicates the situation by allowing new "charter" schools to enter the district.

Reporting and Evaluation

Results are typically presented in tables or figures for each collective outcome variable within the Model, representing the multiple simulations (for different levels of the tunable variables) and iterations (of each simulation). The primary results for Model 2 are presented in two figures, one focusing on the outcome variable of average district achievement by percent choosers for different weightings of choice criteria and one focusing on the difference in achievement for choosers and non-choosers, again by percent choosers for different weightings of choice criteria. The figure depicting the treatment effect for difference in achievement between choosers and non-choosers is duplicated below.

Figure 7: Treatment Effect vs. Percent Choosers



Maroulis et al, 2010a, p. 17.

[34] "Schools were otherwise initialized by running the model for five time periods, under the conditions that students were forced to attend their assigned school" (Maroulis et al., 2010a, p 13).

Each point within the graph represents a different simulation—based on 20 iterations for that simulation. The tunable variables were percent choosers (ranging from 10% to 90%) and ratio reflecting weighting of students' choice criteria [Alpha (α)]. Alpha (α) represents a ratio of school choice based on highest average achievement to choice based on proximity. So, for instance, α =1.0 indicates that choices for that simulation were based solely on highest academic achievement whereas α=0.2 indicates that choices were made primarily on the basis of proximity. The graph represents the outcome for 45 different simulations, each reflecting a different intersection of α and percent choosers. The treatment effect is the standardized difference between choosers and non-choosers on the district's achievement test. Not immediately apparent in the figure is that the rules of the simulation close schools whose enrollment falls below a designated proportion of their building capacity. The authors present a similar figure where the outcome variable (vertical axis) is average district achievement.

The effect of distribution of outcomes for overall district achievement (not shown here) behaved as anticipated, with overall district achievement generally rising when there were more choosers and the choosers emphasized average achievement over proximity in choosing schools. However, the outcome for the treatment effect (displayed above) was counter-intuitive: "as participation rates rise, the magnitude of the treatment effect falls, because capacity constraints increasingly limit the amount of choosers who are able to attend the highest value-added schools" (p. 24). A third model (Model 3) complicated the above scenario further by allowing new schools to enter the district (much like charter schools do); they were initialized with value added scores comparable to those of the highest performing schools, although their average achievement was initially unknown since they were new to the district and initialized at the median of all schools in the district. Model 3 shows that with a high proportion of choosers (80%), "after a certain point, the more the population of students emphasizes achievement in their decision making process, the lower the mean achievement of the district" (p. 21). Additional analyses of the distribution of outcomes suggest that high emphasis on school quality "constrains the number of new schools that can survive," since students select schools with known higher achievement values (p. 23). Thus, Model 3 "shows that too much emphasis on academic achievement at an individual-level can constrain district-

level improvement by limiting the number of new schools that survive." (p. 24). From a policy perspective, Maroulis et al. (2010a) note, the simulations can "set expectations for typically measured outcomes in school choice" (p. 24). They suggest, for instance, the importance of taking capacity constraints into account and that "diverse preferences among households may be more important in getting a school choice model to 'work' than a singular emphasis on academic achievement" (p. 25). They suggest as well that future models could be used to design school choice programs: "For instance, by calibrating the model with more detailed information about the distribution of household decision-making rules (e.g., religious preferences), one could identify locations for new schools that would most increase district-level achievement" (2010b, p. 39). Models might also allow schools to change their value-added parameter as a result of within-school improvement in teaching practice.

Validation strategies include both internal and external validation of ABMs (Boero & Squazonni, 2005). Internal validation focuses on the internal corrections of the computer implementation, the alignment of models across different platforms, and the replication of models across different tunable parameters. External validation includes taking empirical data into account in both model construction and model validation. "Empirical data are needed both to build sound micro specifications of the model and to validate the macro results of simulation" (Boero & Squazonni, 2005, p. 2). This entails the integration of ABM research with a number of other methodologies. Ultimately, the validity of a particular model is "tied to its ability to capture the behavior of the real world" (Miller & Page, 2007, p. 39). A website maintained by Tesfatsion (http://www2.econ.iastate.edu/tesfatsi/acecode.htm) provides up to date advice on modeling tools.

While this Maroulis et al. study is unrelated to the preferred focus of our examples on data use by teachers or about teachers and teaching, a second study by the same senior author comes closer.[35] Maroulis (2009) addresses the question of "why it is so difficult for reform programs to take hold inside a school" (p. 3). He combined simulation and fieldwork at a large high school engaged in a major change effort. The

[35] The model used in this second study was considerably more complicated to explain, which is why we chose not to use it as our key example.

ABM revealed "how stasis can emerge even in an organization comprised of well-informed agents striving for improvement" (p. 3). The study included attention to the schools "social network structure, the strength of social influence within subgroups, and environmental turbulence" (p. 16). The study suggests that "incremental improvement at the micro-level of an organization can eliminate the work activities needed to support the early steps towards more radical, macro-level change" (p. 16). Other ABM studies we found relevant to our focus include one addressing questions about the impact of same or different teacher-student race/ethnicity on academic performance (Montes, 2012) and another on the assignment of peer tutors to students seeking help in on-line learning environments (Westera, 2007). As Maroulis et al. (2010b) note, "Complex system methods can help bridge these aggregate outcomes to underlying mechanisms at work in the system, as well as discover new and unanticipated systemic consequences" (p. 38). We will return to the study of CAS in a subsequent section where we consider its role in supporting multi-methodological research.

Critical Race Theory (CRT)

Critical Race Theory (CRT) emerged initially in the field of law, in response to the relative neglect of "the specific racialized nature of the law and its impact on persons of color" by the Critical Legal Studies movement of the early 1980s (Lynn & Parker, 2006, p. 259). Beginning in the 1990s, it was applied to the analysis and investigation of problems in education. Our key theoretical resources for this methodological tradition are a review article by Lynn and Parker (2006), examining a decade of critical race studies in education, and the *Handbook of Critical Race Theory in Education* (Lynn & Dixson, 2013). We also draw upon a major paper by Ladson -Billings and Tate (1995), as well as Ladson-Billings 's (2006) presidential address to the American Educational Research Association and her chapter in Lynn and Dixson's (2013) *Handbook*. To illustrate CRT, we have chosen an article by Howard (2008), which "[used] critical race theory as a paradigmatic lens to examine the schooling experiences of African American males in PreK-12 schools" (Howard, 2008, p. 954) as well as a CRT analysis of education policy by Gillborn (2013), also included in Lynn and Dixson's *Handbook*.

As explained by Lynn & Parker (2006, pp. 259-260), beginning roughly in the early 1980s, legal scholars argued for a recognition that "racism has been a normal daily

fact of life in society … [so] ingrained in political and legal structures as to be almost unrecognizable." Ladson-Billings and Tate (1995) brought CRT perspectives to bear in education. They and other scholars showed how CRT could be applied in examining educational issues including curriculum, instruction, assessment, school funding, and desegregation (Ladson-Billings, 1998, pp. 18-21). There was, by this time, a tradition of multicultural studies in education, but according to Lynn and Parker (2006, p. 265), "there were serious debates about whether or not multiculturalism or other culture-centered frameworks paid enough attention to issues of inequality and the centrality of race." They suggest that educational scholars using CRT "may have been more heavily influenced by ethnic studies paradigms such as Afrocentric education … [although also] borrowing partly from diverse traditions in education such as critical pedagogy, Black Studies, Chicano Studies, Black feminist and Chicana feminist thought, as well as multiculturalism and multicultural education" (Lynn & Parker, 2006, p. 265-266).

Lynn and Parker (2006) describe how CRT evolved from an early focus on "the material manifestations of racism as a way to argue for social justice and … attention [to] the inclusion of the subjugated voices of racially marginalized peoples" to the application of the same ideas in addressing "issues of gender, ethnicity, language, culture, sexuality and other key markers of difference" (p. 262). As more recent work has borrowed from other traditions, there has been increasing attention to the intersections of race with these other dimensions. More specialized discourses have evolved, including LatCrit, Asian American Critical Race Theory, and critical race feminism, among others.

Aims and Conceptions of Social Phenomena

Lynn and Parker (2006, p. 261) quote Matsuda, Lawrence, Delgado and Crenshaw (1993), who set forth six unifying themes defining the CRT movement:

1. CRT recognizes that racism is endemic to American life.
2. CRT expresses skepticism toward dominant legal claims of neutrality, objectivity, colorblindness, and meritocracy.
3. CRT challenges ahistoricism and insists on a contextual/historical analysis of the law ... Critical race theorists ... adopt a stance that presumes that racism has contributed to all contemporary manifestations of group advantage and disadvantage.

4. CRT insists on recognition of the experiential knowledge of people of color and our communities of origin in analyzing law and society.
5. CRT is interdisciplinary.
6. CRT works toward the end of eliminating racial oppression as part of the broader goal of ending all forms of oppression. (Matsuda, et al., p. 6)

In arguing for CRT studies in education, Ladson-Billings and Tate (1995, p. 48) set forth three central propositions:

1. Race continues to be a significant factor in determining inequity in the United States.
2. U S. society is based on property rights.
3. The intersection of race and property creates an analytic tool through which we can understand social (and, consequently, school) inequity.

With regard to the first of these propositions, Ladson-Billings and Tate (1995) first allude to the substantial statistical evidence of racial inequality in the United States, then go on to discuss the problematical nature of racial categories. Even though "the world of biology has found the concept of race virtually useless," (p. 49), "thinking of race strictly as an ideological construct denies the reality of a racialized society and its impact on 'raced' people in their everyday lives" (p. 48).

In her more recent chapter, Ladson-Billings (2013) expands upon this proposition:

> The first tenet of CRT is the notion that racism is not some random, isolated act of individuals behaving badly. Rather, to a CRT scholar racism is the normal order of things in US society. This is the thing that distinguishes CRT scholars from others who investigate race. Some focus on specific instances of racism or might admit to institutional racism. However, few outside of CRT would declare that racism is normal. Most argue that racism resides in individual (and sometimes corporate) beliefs and behaviors.... (p. 37)

Ladson-Billings goes on to quote Delgado and Stefancic (2001, p. 7) as stating that CRT scholars believe racism "is the usual way society does business, the common, everyday experience of most people of color in this country." To be sure, CRT scholars substantiate these claims. In her AERA presidential address, Ladson-Billings (2006) marshals statistical evidence documenting, for example, disparities in per-pupil

expenditures in predominantly white versus predominantly minority school districts, disparities in salaries and wages between black and white workers in the U.S. labor force, racial disparities in voter registration rates prior to and following passage of the Voting Rights Act, and racial gaps in achievement test scores, advanced course taking, and high school graduation rates.

Concerning Ladson-Billings and Tate's (1995) second proposition, that U.S. society is based on property rights, they explain that despite the conflation of the ideologies of democracy and capitalism, they are in fact separable. They observe that "traditional civil rights approaches to solving inequality have depended on the 'rightness' of democracy while ignoring the structural inequality of capitalism" (p. 52). In other words, the ideology of individual rights has served to mask the fact that legal and institutional structures function to protect property more so than individuals. Consequently, they argue that concepts of democracy and capitalism must be disentangled. They cite educational implications including school funding, access to rich curriculum, and other sources of disparity in opportunity to learn.

With regard to their third proposition, concerning an analytic tool centered on the intersection of race and property, Ladson-Billings and Tate (1995) challenge claims of neutrality, color-blindness, objectivity, and meritocracy. They call for research approaches that recognize the racialized nature of sociopolitical power structures, which situate persons of different races in distinct ways, and that acknowledge the centrality of this racialized reality in the lived experiences of people of color:

> A theme of 'naming one's own reality' or 'voice' is entrenched in the work of critical race theorists. Many critical race theorists argue that the form and substance of scholarship are closely connected. These scholars use parables, chronicles, stories, counterstories, poetry, fiction, and revisionist histories to illustrate the false necessity and irony of much of current civil rights doctrine. (pp. 56-57)

This theme matters because "political and moral analysis is situational" and "social reality is constructed by the formulation and exchange of stories about individual situations" (p. 57). In addition, the naming-one's-own-reality theme can support the psychic preservation of marginalized groups, as a counter to demoralizing self-

condemnation. Finally, "naming one's own reality with stories can affect the oppressor" (p. 57). The dominant group constructs reality and rationalizes inequality with its own stories (stock explanations). Thus, counterstories can catalyze cognitive conflict and consequent self-examination.

From its inception, CRT has embodied a concern for social action. The sixth of the guiding principles set forth by Matsuda, et al. (1993) held that "CRT works toward the end of eliminating racial oppression as part of the broader goal of ending all forms of oppression." Lynn and Parker (2006, p. 270) argue for viewing "CRT and education scholarship … as a form of academic scholarship as well as a form of activism … [challenging] CRT in education scholars to expand their notions regarding what it means to conduct research that has relevance to and for the lives of the disenfranchised."

Design and Enactment

Lynn and Parker (2006, pp. 272-273) cite Solórzano and Yosso's (2002) definition of critical race methodology in education research:

> We define critical race methodology as a theoretically grounded approach to research that (a) foregrounds race and racism in all aspects of the research process; … (b) challenges the traditional research paradigms, texts, and theories used to explain the experiences of students of color; (c) offers a liberatory or transformative solution to racial, gender, and class subordination; and (d) focuses on the racialized, gendered, and classed experiences of students of color. Furthermore, it views these experiences as sources of strength and (e) uses the interdisciplinary knowledge base of ethnic studies, women's studies, sociology, history, humanities, and the law to better understand the experiences of students of color. (Solórzano & Yosso, 2002, p. 24)

The reporting and theorizing of counterstories is salient in the CRT tradition. These counterstories often take the form of first-person accounts of microaggressions[36] suffered by people of color, although Ladson-Billings (2013) emphasizes that "the point of storytelling is not to vent or rant or be an exhibitionist regarding one's own racial

---

[36] *Microaggressions* are defined by Sue, Capodilupo, Torino, Bucceri, Holder, Nadal, and Esquilin (2007, p. 271) as "brief and commonplace daily verbal, behavioral, or environmental indignities, whether intentional or unintentional, that communicate hostile, derogatory, or negative racial slights and insults toward people of color."

struggle" (p.42). There must be a larger point to be made. Counterstories are used to challenge received truths and "to illustrate and underscore broad legal principles regarding race and racial/social justice" (p. 42). In summary, "the chronicle or counter-story is about racial justice principles, not personal afront" (p. 43). In addition to counterstories, CRT scholars marshal empirical data to document racial inequities, and employ a set of analytical tools to analyze both law and social policy.

In their review, Lynn and Parker (2006, footnote 1, pp. 283-284) note that "CRT has moved into other areas of education research, from qualitative study and the emphasis on narrative and counter-story, to its impact in analyzing and interpreting the effects of educational policies and practices in K-12 and higher education settings." They discuss in particular applications of CRT in sociology. For example, Mitchell (2001) used an experimental design to investigate the influence of race on teachers' decisions about students' suitability for an advanced algebra course. In this chapter, we draw upon Howard (2008) primarily to illustrate the use of counterstories, and upon Gillborn (2013) primarily to illustrate one CRT approach to policy analysis, acknowledging that these two articles are merely illustrative of a broader range of methodological approaches used in CRT.

The use of counterstories is well illustrated in Howard's (2008) account of the disenfranchisement of African American males in PreK-12 schools. Howard situates himself as "an African American male scholar who has endured the challenges that scores of young Black men face from pre-school through post-secondary education and beyond" and also "as a father of three young Black males," which make the issues he addresses of "not only professional interest" but "real, personal, and tangible." He uses CRT as "a theoretical framework that does not have to 'make the case' or 'explain beyond a reasonable doubt' how and why race matters in the schooling experiences of African American males, but that accepts race and racism as integral parts of life in the United States" (p. 959).

Howard (2008, p. 963) sets forth as "one of the basic tenets of CRT … the normalcy and permanence of racism." He describes CRT within education as "an evolving methodological, conceptual, and theoretical construct that attempts to disrupt race and racism in education" and explains that:

> Critical race theorists in education anchor their interrogation of racism in four primary ways: (1) by theorizing about race along with other forms of subordination and the intersectionality of racism, classism, sexism, and other forms of oppression in school curriculum; (2) by challenging dominant ideologies that call for objectivity and neutrality in educational research (CRT posits that notions of neutrality typically serve to camouflage the interests and ideology of dominant groups in the U.S. …); (3) by offering counterstorytelling as a liberatory and credible methodological tool in examining racial oppression …; and (4) by incorporating transdisciplinary knowledge from women's studies and ethnic studies to better understand various manifestations of discrimination. (pp. 963-964)

As part of another study, "two hundred African American middle and high school males were surveyed about their schooling experiences, and potential roles that race may play in them" (Howard, 2008, p. 969). Over the course of a year, Howard conducted in-depth interviews with a subsample of these students, and from that subsample, ten middle- or high-school students, representing five different metropolitan schools, served as participants. Five of the ten were from urban schools in primarily low-income areas, with mostly African American and Latino students, and five were from racially mixed schools in suburban communities, with mostly White, middle-class students.

After setting forth his theoretical framework, Howard (2008) recounts a number of stories related by his participants, vividly documenting "their keen awareness of negative stereotypes about African American men" (Howard, 2008, p. 969), which all of the young men reported making explicit attempts not to reinforce. He includes brief quotations from one participant recounting the surprise of teachers and students at his appearing on the honor role, of another (who is a school leader with good grades) being told by a teacher, "You're not like the rest of them," and from another attributing to teachers the attitude, "Why should we [teachers] care when they [Black males] don't?" despite the fact that "A lot of us [Black males] are doing homework every day, studying, working hard, so that we can improve ourselves." There are also accounts of more punitive treatment for Black versus White students committing the same minor school infractions and of suspicion immediately falling upon "all the Black boys" when

"something goes wrong at school." Another participant observes that when "White or Asian kids … [challenge] stuff in the book" they are praised as "critical thinkers," but when he does the same he is accused of being "disrespectful" (Howard, 2008, pp. 970-971).

After presenting and discussing these and other stories documenting racial stereotypes, Howard (2008) turns to accounts of racial microaggressions, presented in a similar format. To better represent the character of this writing, Howard's report of this example from one of his research participants is quoted in full:

> We had an assembly at school for all the seniors who were going to college. They had the students come up on stage and said what college they were going to. For whatever reason, they forgot to call me up during the assembly, so I asked Mr. Matthews (the assistant principal) after the assembly why I didn't get called up, and he said that they were only calling up the kids who were going to 'good colleges,' and they didn't think that Morehouse was a really good college. That was like a slap in the face to me. Morehouse is a good college. I'm one of the first kids to go to college in my family, and he says that, it is not a good school. How does that make me feel? (Howard, 2008, pp. 973-974)

Several stories with similar themes are related, including an account of an advanced placement teacher's disbelief when an African American male student showed up in her class.

Howard's participants affirmed the value for themselves of articulating these experiences. He reports that, "One of the enamoring aspects of this work has been the young men's surprise about someone wanting to hear about their experiences" (Howard, 2008, p. 975). He then goes on to quote one of his high school participants:

> I'm kinda shocked that you are asking us about this kinda stuff, because we never get asked about racism and stuff. Because you just assume that nobody's gonna believe you when you tell them about it. But this stuff is real....

Our second example illustrates additional aspects of the CRT perspective. Gillborn (2013) offers a CRT analysis of education policy, focusing on legislation in Arizona aimed at curtailing ethnic studies programs, especially Mexican American/Raza Studies in Tucson and elsewhere. He illustrates how CRT counters the dominant

narrative of rational, objective, color-blind policy formation, challenging the "taken-for-granted traditional view of policy as an incremental process moving toward greater justice and inclusion" (p. 130). Consistent with CRT's insistence on a contextual/historical analysis of the law, Gillborn situates the 2010 Arizona legislation in its historical context, beginning decades before with a grassroots community movement to establish a Mexican American Studies Department within the Tucson Unified School District. He documents the success of that program, citing statistical evidence of its success as shown by test scores, high school graduation rates, and college matriculation. He goes on to describe an ensuing campaign against multicultural education in the state, passage of the legislation itself, and its aftermath. Through a careful analysis of both legislative language and elected officials' public statements, he shows how a "color-blind" perspective, although ostensibly neutral, in effect places "the fears and interests of White people … at the forefront of policy" (p. 132) and uses seemingly neutral language to mask racial references:

> … in their official pronouncements, reference to 'resentment toward a race or class of people' has been widely interpreted as an explicit attempt to protect White people … from accusations of bias and race discrimination. (Gillborn, 2013, p. 132)

Consistent with Ladson-Billings and Tate's (1995) reference to an analytic tool based on the intersection of race and property, Gillborn's analysis illustrates how CRT identifies and challenges a narrative implicit in much education policy:

> The wording of the Arizona statutes … enforce a *neo-liberal* world view as the only permissible basis for action. Neo-liberalism is a conservative perspective that stresses the importance of individual self-interest and free market operations as the basis for the most efficient and just form of society … The supremacy of an individualistic and 'color-blind' perspective is guaranteed by law where advocating 'ethnic solidarity' is prohibited. (p. 132)

Gillborn (2013) then goes on to further theorize his CRT policy analysis by explaining concepts of interest convergence, contradiction-closing cases, and interest divergence. Without attempting to do justice to his discussion, interest convergence refers to the idea that the interests of minorities in moving toward racial equality will

only be accommodated to the extent that they converge with the interests of powerful members of the dominant group. Contradiction-closing cases refer to policy shifts that appear to remedy an obvious injustice, thereby removing an apparent contradiction between that injustice and an "official rhetoric of equity and fairness" (p. 135), but which are quietly cut back or only haltingly implemented afterwards. Interest divergence refers to the common case where racial interests are assumed to diverge, which can support the notion that the majority will somehow benefit from further marginalization of minority groups.

Reporting and Evaluation

Howard points out that notions of meritocracy, democracy, and equality accepted unquestioningly by many citizens are in fact "ideas and concepts, not realized ways of life" (p. 976). He quotes Sleeter and Delgado Bernal (2003, p. 249) who state, "At issue is the question of what counts as truth and who gets to decide." Howard then makes it clear that "Although CRT situates race at the center of its analysis, it recognizes the fluidity and multifaceted nature of identity for all individuals, and does not attempt to create monolithic constructions or experiences of any group" (Howard, 2008, p. 976).

Reflecting CRT's concern with action and engagement, Howard (2008) concludes his paper with a section discussing the CRT challenge in education. He calls for additional research and research funding to find interventions that can help this group and repeats calls for further CRT analysis of curriculum, instruction, and assessment. Within these areas, he calls out standardized testing and school discipline in particular.

Gillborn (2013, p. 138) explains that "Critical race scholars argue for a fundamentally different interpretation of the role of educational policy … as largely serving to *manage* race inequality at sustainable levels while maintaining and even enhancing White dominance of the system." His analysis shows the importance of "taking a historically contextualized perspective … [and] of follow-up in the aftermath of legislative change" (p. 134).

In concluding their review of ten years of Critical Race Studies in Education, Lynn and Parker (2006) articulate similar criteria for success in CRT research, calling for "Critical Race Praxis" to defend against erosion of civil rights protections, expand current legal recognitions of racism and racial injury, document and remedy racial inequities in

opportunity to learn, and in other ways challenge disparate allocation of resources. They affirm that:

> Critical race scholars in education are concerned with how schools act as racial stratifiers. They are committed to conducting both qualitative and quantitative research that exposes racist beliefs, practices and structures in schools and the broader society. However, they are also committed to using their critiques as a basis to create more equitable and humane discourses, practices and structures that will enable racially marginalized students and other school people to overcome and ultimately destroy existing obstacles to their freedom. (Lynn & Parker, 2006, p. 282)

Participatory Action Research (PAR)

Participatory Action Research (PAR) is a complex research tradition. According to Jordan (2009, p. 16), "PAR is itself a blend of a broad range of research approaches and epistemologies that include participatory research, action research, feminist praxis, critical ethnography, [indigenous][37] research methodologies, transformative education, critical and eco-pedagogies, and popular and community education." It is distinguished not by the use of any specific kind of data collection or analysis, but rather by its goals and by the relationships it strives to establish between the researcher and the community where the research is carried out. It has in common with other community-based research traditions a thoroughgoing commitment to "doing research *with* and not simply *on* members of a community" (Root, 2007, p. 565). More specifically, PAR seeks to educate, to empower, and to directly benefit marginalized groups. Kapoor and Jordan (2009, p. 6) go so far as to suggest that "the only way to ensure that a PAR process is initiated and sustained is for academic researchers to continually work at embedding all aspects of participatory research in a living praxis, where participants learn to take control and academic researchers become willing hostages to their concerns."

Our two key theoretical resources for this methodological tradition are a chapter by Root (2007) on "Community-Based Research" and a book by Kapoor and Jordan

[37] Jordon refers to "aboriginal research methodologies" but the term "indigenous" is more common today in most contexts.

(2009) focusing on PAR and social change in international contexts.[38] We have chosen to illustrate PAR with an article by Wyman, Marlow, Andrew, Miller, Nicholai, and Rearden (2010) describing a project undertaken with the Yup'ik villages served by the Lower Kuskokwim School District (LKSD) in southwestern Alaska. That project's goals were to better understand and to work against the decline in young people's learning and use of their Indigenous language, Central Alaskan Yup'ik, in the face of accountability pressures and associated state and federal education policies, especially under NCLB. The first two authors, Leisy Wyman and Patrick Marlow, list their affiliations as the University of Arizona and the University of Alaska, respectively. The remaining four authors list their affiliations as the Lower Kuskokwim School District. As explained in the article, "All research team members represented insider-outsiders in the villages served by LKSD." Rearden, Andrew, and Nicholai were "longstanding bilingual educators [who] had developed Yup'ik language programs and led … summer bilingual institutes," Miller, "the head of academic instruction in LKSD, was a former LKSD village principal," and Wyman, now a university researcher, was a former LKSD teacher who had done previous research on language shift in a Yup'ik village. Marlow "had collaborated … to develop university Indigenous language teacher training programs supporting Yup'ik language revitalization" (Wyman, et al., 2010,p. 706).

Aims and Conceptions of Social Phenomena

PAR is grounded in a deep respect for the wisdom and practices of the communities where research is undertaken. Historically, it was often, though not always, "allied with radical or liberatory political movements" (Jordan, 2009, p. 16). Citing the work of Paulo Freire and Orlando Fals-Borda, Jordan (2009, pp. 16-17) describes a scenario whereby "social research was to be transformed from an abstract, detached, disinterested, and objective science conducted by outside experts into an emancipatory process centered on what Freire called conscientisation, where the poor were to become agents of social and political transformation aimed at creating just, peaceful, and democratic societies." Similarly, Root (2007, p. 571) observes that "community-based research usually begins with a problem to be solved rather than a theory or hypothesis to

[38] Another helpful introductory resource describing a broadly adaptable approach to PAR is provided by the American Bar Association's Rule of Law Initiative (ABA ROLI, 2014).

be tested. Thus, the research rests on a judgment that a particular condition or practice is problematic. Since the word 'problematic' implies undesirable, community-based research does not claim or aspire to be value-neutral."

These principles are illustrated in the paper by the research team. They are concerned with the phenomenon of Indigenous language endangerment, evidenced by "early and uneven signs of a language shift to English" (Wyman, et al., 2010, p. 701). They briefly review a long history of heritage language instruction in Alaska, marked by "[attempts] to eradicate Indigenous languages and ways of knowing by punishing students for speaking their languages" in missionary and government-sponsored schools (Wyman, et al., 2010, p. 703). They note that bilingual programs like those "supported by LKSD for decades" were "federally envisioned as transitional bilingual programs to promote English language development and mainstream education goals" (Wyman, et al., 2010, p. 704). Moving up to the present day, they note that "NCLB does not outlaw bilingual education, yet under NCLB funding structures and accountability systems, schools and school districts are under strong pressure to adopt English language instruction and standardized curricula" (Wyman, et al., 2010, p. 705). Summarizing the state of affairs leading up to their study, they quote an article from the *Journal of American Indian Education* reporting that "in 2006 LKSD administrators noted that Yup'ik was under threat from 'attempts to impose English-only laws (at the state level), English testing requirements in the third grade in Yup'ik instructional environments, and professionals from outside the area identifying the Yup'ik language as the 'flaw' preventing students from succeeding in the western schooling system" (Wyman, et al., 2010, pp. 705-706). Thus, the social phenomenon of concern is conceived in terms of the Yup'ik language and Indigenous ways of knowing being threatened by educational policies pushed by state and federal governments, abetted by the pressures of high-stakes testing and accountability under NCLB.

Given this framing, the research team sought to better understand changing patterns of Yup'ik language use among young speakers, accountability pressures, and local efforts to support Yup'ik language use. Research questions were developed over a period of four months. The authors are careful to state, "In accordance with Indigenous ethical research standards (Smith 1999), the team relied on Yup'ik team members'

judgment to ensure that research goals, questions, and methods were respectful and responsive to the needs of participants and communities represented" (Wyman, et al., 2010, p. 707).

Their research questions were as follows:

(1) What is the current state of Yup'ik language use in villages served by LKSD?

(2) How are schools in LKSD performing on state measures of achievement?

(3) What, if any, relationship exists between school-level achievement on state measures and bilingual program histories for villages in LKSD?

(4) How, if at all, are new state accountability measures affecting local decisionmaking around bilingual programming in communities where Yup'ik is spoken?

(5) What resources exist for bilingual development in sites where Yup'ik is still spoken by children and in sites where language shift is already taking place? (p. 706)

Design and Enactment

The activities undertaken in PAR and the methods employed are determined in collaboration between researchers and community members. Root (2007, p. 566) lists several assumptions of other research traditions, often unexamined, which are opposed by community-based research approaches, including PAR. Two of these are, "Since the expert has training in the social sciences and her subjects do not, she should control how the study is conducted and they should not" and "An expert is better able to interpret or explain facts about her subjects or their community than they are." In opposition to these assumptions, a central and distinctive feature of PAR design and enactment is the degree of collaboration called for in the research planning process itself. A PAR study is often undertaken to address a problem identified in the local setting. Thus, participants' understandings and perceptions of the situation are critically important. Participant engagement in the research may also "increase the members' ability to identify problems and develop plans to solve them" (Root, 2007, p. 566).

Authentic collaboration in research planning implies that other aspects of the research process must be reexamined. University requirements for informed consent in research with human subjects, for example, may call for research hypotheses and study

procedures to be spelled out in advance, but if questions and methods are in fact worked out collaboratively, then that is not possible. Informed consent procedures center on protection of and securing agreement from individuals, but as Root (2007, p. 570) explains, "the subjects in community-based research are often entire groups or communities." This raises questions of benefits and harm to the community, not just individual members, and also forces researchers to confront the question of who is able to consent on behalf of the community.

The research team begin their description of research methodology by explaining their own "insider-outsider" status and the collaborative process of developing research questions. Having developed their list of research questions, the research team "invited Yup'ik educators with at least five years of experience in local village schools to become research participants." Design of the research was collaborative, and guided by ethical considerations. For example, they "decided early on … that understanding processes of policy-making, language shift, and language planning from Yup'ik educators' points of view was more important than publicly labeling the language-shift/maintenance status of specific villages" (p. 707). Thus, all surveys and discussions of village sites were anonymous, and participants decided "whether and what types of data should inform district efforts and what could be shared with broader audiences." Concerned with "the ethical dimensions of reproducing and/or countering specific 'rhetorics of endangerment' … [and the risk of] inadvertently [erasing] community strengths, local meanings of bilingualism, and specifics of community struggles" (p. 707), the research team relied on a carefully crafted mix of open-ended and more focused survey and interview questions, as well as less conventional data collection strategies. For example, "in one session, … participants [were asked] to 'tell the story' of the last time their community made a five-year plan for their school language program" (p. 709). The team "debated the benefits and drawbacks of analyzing NCLB-related progress data" (p. 707), confronting the dilemma that "standardized tests … commonly render the knowledge of minority communities invisible, yet generate numbers which are 'awarded blind trust from the public' as well as educators" (p. 708) and in the end "decided to analyze state-determined school progress scores in light of historical bilingual education efforts." This effort included "[analyzing] some of the unintended, as well as the intended consequences of

NCLB-related measures in village sites … by asking participants to reflect on changes and continuities in their language programs between 2002 and 2007, the first five years of NCLB implementation in Alaska, on surveys and in focus groups" (p. 708).

Root (2007) also highlights the importance of respecting participants' preferred categories and not relying solely on those of the researcher. In this spirit, as the research team prepared questions and research tools, they "paid particular attention to local patterns of diglossia, terms for language use, and discourse norms such as a reluctance to speak for others … in order to create bilingual spaces in which participants could share observations and translanguage[39] … as part of the research process" (p. 708). Participants could respond in whatever language was most comfortable, and breakout discussion groups "were led or co-led by a fluent Yup'ik-speaking researcher."

Reporting and Evaluation

University-based studies addressed to professional audiences are often "written in the language of experts and not available to members of the community who are the source of the study's data." In contrast, "Community-based studies allow the subjects to choose the language in which the findings are written and the audience to whom they are to be addressed" (Root, 2007, p. 569). As a glance at the reference lists of our key theorists suggests, some reports of PAR studies are also addressed to professional audiences, appearing in professional journals. The publication by Wyman, et al. (2010) in *The International Journal of Bilingual Education and Bilingualism* is but one example. However, PAR research findings may be presented in pieces written in different voices for different audiences, as determined by the research participants' choices. Findings must often be presented in different ways if they are to be taken up in local settings. Care must be taken to frame findings with an eye toward the ways they might be appropriated by parties with other agendas, which may be inconsistent with those of the PAR researchers themselves. In our illustrative study, the research team document dissemination of their findings via "full group sessions [in which] participants shared their experiences" (p. 709). In particular, participants apparently shared their strategies for working with communities, for "[using] English test scores to demonstrate the

[39] Translanguage is related to the natural movement between languages referred to as codeswitching, but pertains more specifically to the deliberate use of shifts between languages for pedagogical purposes, recognizing that exact translation of ideas from one language to another may not be possible.

effectiveness of heritage language instruction" (p. 714), and for "strategic classroom language planning" (e.g., using seating arrangements to assure that English speakers were separated among their Yup'ik-speaking peers). As noted, the article also refers to some types of data used to "inform district efforts" (p. 707), not to be shared with broader audiences. Wyman (personal communication, October 28, 2014) reports that,

> "In addition to disseminating findings in the sessions [described in the article], we wrote up reports and executive summaries of the research for the school district office personnel and district school board members, and made presentations to district school board members, (who themselves represented multiple villages). Our dissemination and publication process also involved sending drafts of reports and articles to participants for their scrutiny and feedback, so they could approve the use of their data in context before we publicly shared products based on the research either for in-district use or publication. At times, this meant that we worked over our exact wording to ensure that our representation of the research in articles would make sense both to outside academics, varying local stakeholders in LKSD, and Alaskan policymakers. While the various levels of dissemination, feedback, discussion and revision took time, my colleague Patrick Marlow and I saw this as an important process that allowed us to avoid potential misunderstandings, and ensure maximal buy-in, uptake and use among participants, district stakeholders, and policymakers in a district that was already engaged in an intense debate over bilingual education. We asked participants to help us choose the language in which the findings were written, and to identify audiences that we needed to keep in mind as we wrote up and disseminated our research ... We also had a fairly extensive additional process that we used to disseminate draft findings in multiple written formats for review and feedback, which helped us address the needs and concerns of community audiences, even as we sought ways to write for our own academic audiences, as well."

Root (2007, p. 574) suggests that community-based researchers might choose to use quantitative methods in communicating with some audiences, and indeed, Wyman, et al. (2010) do present descriptive statistics and a bar graph showing percentages of young

students in various village sites who "spoke Yup'ik well for their age" (p. 711). However, the authors devote more space to summaries of their participants' descriptions of "how, as they struggled with the challenges of early language shift and the implementational and ideological fallout of NCLB, they were working to convince communities that bilingual education approaches could meet educational accountability demands, as well as long-term language maintenance goals" (Wyman, et al., 2011, p. 714). These efforts included documenting students' success on third-grade tests immediately after they transitioned out of a Yup'ik early literacy program, a teacher's showing parents videos of her own early literacy instruction in Yup'ik at a community meeting, and using a group activity at a community meeting to convey how quickly language shift could cut children off from the wisdom of Yup'ik monolingual elders.

Much of the "Findings and discussion" section describes participants' reports of promising practices they have used to encourage Yup'ik language development and maintenance or to assuage parental concerns over the potential impact of heritage language programs on high-stakes accountability test scores. One teacher, for example, used classroom seating arrangements to separate English-speaking students among Yup'ik-speaking peers, and reported that by mid-year, use of Yup'ik had increased. Recommendations are embedded throughout the discussion, for example: "Data … indicated that educators in LKSD urgently need to re-examine language programs and pedagogies designed for students who share relatively even sets of language skills" in the face of increasing heterogeneity in young children's language use both within and across research sites (p. 710).

Theorists emphasize the potential of PAR as a resource for social change. As Root (2007, p. 574) states in his conclusion, PAR studies "are often … more about changing than describing the situation in which [participants] find themselves." He explains that "a goal [of PAR] is to transfer a skill from the expert to her subjects or allow her subjects to employ their skills in the serve of their interests. As a result, community-based studies are often more about educating members of a community than predicting or explaining their behavior …" (p. 574). Similarly, Kapoor and Jordon (2009) include nine chapters on "Case Studies of PAR, Education, and Social Change" in international contexts. Among others, these rich examples include a study of arts-based

PAR with street-involved youth in Canada (Conrad & Kendal, 2009), a study of rural women and land tenure in Tanzania (Mhina , 2009), and a study of science education in Zimbabwe (Shizha , 2009). At the same time, some theorists (e.g., Hale , 2006) caution against the assumption that activism or advocacy in research entails an abandoning of intellectual rigor or pursuit of knowledge valued in the academy. Acknowledging the challenges, tensions, and contradictions inherent in activist research, Hale (2006, p. 105) argues that "within an academic setting, the case for activist research can only be made in rigorous academic terms. … [It] must be justified by the claim that it yields privileged scholarly understanding. It must generate new knowledge and theoretical innovation …. Outside academia, however, activist research will be judged in starkly different terms: what is its potential contribution to the political struggle underway? At the end of the day, activist scholars must embrace two quite distinct sets of objectives and forms of accountability, and they must negotiate the often considerable tensions between them."

In their concluding paragraph, Wyman, et al. (2011, p. 717) recapitulate their own PAR goals and their own evaluation of their attainment:

> Here we have demonstrated how researchers, administrators, and educators can open spaces for reflexive, collaborative research, thereby gaining insight into the influence of policies and schooling on local and regional linguistic ecologies. Highlighted, as well, are ways that educators can generate and share local language planning strategies when they have sustained opportunities for working through the complexities of bilingualism and educational policy-making. Overall, our findings underscore the urgent need for, and also the promise within, spaces for locally directed language investigation and language planning in the current era of standardized educational accountability systems.

Analytic Summary

In the opening of this section on MTs, we foregrounded the categories that we used to compare methodological traditions: aims and social phenomena, enactment, evaluation and dissemination. In closing, we foreground the differences illuminated by these contrasts among traditions. Consistent with theoretical sampling, we developed these categories through a dialectical process of exploring alternative MTs, thinking about the dimensions highlighted by contrasts among them, considering additional MTs

to flesh out these dimensions, and so on. We have sought to illustrate these dimensions with a purposive sampling of MTs representing a range of stances or choices with respect to the categories encompassed by each dimension; however, it would be a mistake to assume that these categories could not also be addressed by other MTs. It would also be a mistake to suppose that any particular MT could be located within just a single category with respect to any of these implied dimensions. Any attempt to do so would be both controversial and misleading with respect to the complex and dynamic nature of MTs. With these caveats, we invite the reader to reflect on the MTs we have described, as well as the illustrative studies we have chosen. Our message is that significant, high-quality programs of research show great variation with respect to each of these categories or dimensions:

- **numbers and depth of cases**, from in-depth study of single cases, to studies that compare small numbers of cases, to studies that aggregate large numbers of cases;
- **emic and etic perspectives** in other words, attending to understandings and intentions or research participants in context (emic) and to the theoretical frames and perspectives that researchers can bring to the context from outside (etic);
- **nature and aspects of phenomena studied**, whether status at one point in time or process over time; whether relatively decontextualized or studied in relation to other case-specific circumstances and to antecedents and consequents; whether formally constituted or informal and emergent;
- **time scales & relationships among them**, from more micro to more macro phenomena to relationships among phenomena operating at different time scales;
- **approach to design and analysis**, from more inductive, to more deductive, to more iterative and dialectical;
- **conceptions of cause or explanation**, from focus on a single cause to multiple interacting and recursive causes on different time scales; from designed interventions or other external influences to emergent phenomena arising in self-organizing systems; and from those that operate beneath the awareness of actors involved to those based on intention and action-orienting critical reflection;

- **conceptions of change or progress**, across time scales and levels of the organization or system, and across distance from the phenomenon studied, from more proximal to more distal;
- **loci for generation of research problems or questions**, from locally generated to theory or literature based to those that evolve in contexts of interest as a study unfolds;
- **relationship between researcher and researched**, including roles and authorities of those studied at different points in the research process;
- **standpoints of researchers**, from explicit to implicit, from espousing an ideal of neutrality to espousing an explicit value orientation, from participatory to remote.

Looking across the categories highlights the affordances and the limitations of any MT. They speak to the limitations of any single methodological tradition. They also provide a partial map of potential alternatives—or better complementarities--that might be considered in developing a program of research.

3. APPROACHES TO METHODOLOGICAL PLURALISM (AMPs)

By "approaches to methodological pluralism" (AMPs), we mean explicit theories about how one might work across methodologies or methodological traditions (MTs) in building a multi-study program of research. AMPs support researchers in conceptualizing, designing, enacting and evaluating programs of research that draw on multiple methodologies or MTs. They provide overarching frameworks in which particular studies can be located. While AMPs can be relevant to single studies, our focus is on their use in conceptualizing multi-study programs of research. Unlike the previous section on MTs where there are direct methodological implications, the AMPs operate for us at a meta-methodological level, providing broader frameworks within which different MTs can be put to work.

Attention to mixed or multiple methods is hardly new. Histories of social research (Bryman, 2008; Alastalo, 2008) cite examples of research drawing on multiple methods (e.g., surveys and interviews) dating back at least to the 1930s.[40] This was not seen as remarkable at the time, but rather as the way things were done. It is probably safe to say that all the MTs described in the previous section have drawn on methods from other MTs, and the examples we provide illustrate this. In Generalized Causal Inference (GCI), Slavin et al. (2012) used as evidence narratives from administrator "walk throughs" to document treatment implementation and interpret findings for their experimental study. In ethnographic research (EthR), Hubbard et al. (2006) attended to quantifications reflecting other methodologies that were taken up emically in the contexts they studied; they also drew on previous research, for example, raising questions about changes in students' standardized test scores. Sometimes the use of alternative methodologies was reflected in the MT theorists' validity guidelines as with Agent Based Modeling (ABM) (where, e.g., comparisons between simulated and actual phenomena were advised) and GCI (where external validation relied, in part, on qualitative comparisons); and sometimes the alternative methodologies complemented the focal methodology, by providing social and historical context for a segment of discourse analyzed, or by illuminating mechanisms or processes underlying relationships or

[40] Hunter and Brewer (2003) offer a mixed methods reading of the Hawthorne studies begun in 1924 (p. 585).

patterns observed, or by providing corroborating evidence about the typicality of a case. In these examples, a single MT was dominant, but methods from alternative MTs complemented the dominant tradition. Some MTs have benefited from the selective integration of multiple methodologies and theoretical perspectives and serve in a sense as "umbrella[41]" methodologies within which other methodologies can be located. Critical Race Theory (CRT), for instance, can be seen to draw on ethnographic research, discourse analysis, and survey research to support its arguments. Similarly Participatory Action Research (PAR) can draw on a range of methodologies in addressing problems relevant to participants. Design Based Research (DBR) can routinely be seen to incorporate ethnographic research, discourse analysis of participants' interactions, measurement of relevant constructs, comparative case study (CCS) research (as DBR is implemented in different contexts or with different learners), and even context specific experiments to address questions about different design choices. Cobb (2007) uses the concept of *bricolage* (adapted from Levi-Strauss) to describe alternative traditions "as sources of ideas that we can appropriate and modify" (p. 29) in developing a new approach from the combination.

The AMPs illustrated in this section operate at a more general level and provide conceptual resources that can encompass multiple MTs in a program of research. They reflect explicit attempts to theorize and guide a multi-methodological research practice, initiated in response to some perceived need within the field. In situating themselves within and against the field of social research to which they are contributing, each AMP also portrays a somewhat different map of the broader field, foregrounding different boundaries, different lacunae to be addressed, and different visions of an enhanced future.

We focus on three quite distinct AMPs that provide conceptual resources (including analytical language) for methodological pluralism. Each has a somewhat different emphasis: the "Mixed Methods" (MM) approaches tends to foreground ways of working across methods or methodologies, most often crossing boundaries named simply as qualitative and quantitative; approaches under the rubric of "multi-, inter-, or trans-disciplinary" research (MD, ID, TD, respectively or *MIT*-D to encompass them all) tend

[41] Thanks to the 2014 Complementary Methodologies class at the University of Michigan School of Education for suggesting this term.

to foreground ways of working across disciplines, including but not limited to their methodologies; and approaches under the rubric of "complexity theory" or "complex systems theory" (CST) foreground systems of various sorts and the ways in which their parts dynamically interrelate, the study of which entails multiple methods and methodologies. Each of these AMPs has a substantial history, with its genesis located in different seminal documents, conferences, or circumstances, and with handbooks or encyclopedias, textbooks, and one or more journals devoted to them. While there is some cross citation between proponents of these AMPs, it is clear from the historical narratives and citation patterns that these conversations remain quite distinct. While they are not, for the most part, in conflict—each can be largely framed in terms of the other—they do foreground different conceptual resources for methodological pluralism.

It is also important to note that these three AMPSs are far from the only explicit discussions that theorize methodological pluralism. MM and MIT-D appear, based on our scanning of multiple general handbooks and textbooks on social and education research methodologies, to be most frequently represented. CST, while more recently put to the question methodological pluralism in social research, especially education research, seems to us to provide a distinct and generative heuristic for engaging in methodological pluralism, with a growing number of texts, and so we have included it here as well. Beyond these three AMPs, if one were to look within particular fields of research, like organizational studies (Buchanan & Bryman, 2009) or learning sciences (Sawyer, 2006, 2014) or reflexive sociology (Bourdieu &Wacquant, 1992), one would find rich theorizations of methodological pluralism situated within the substantive theories that inform those fields. We also note that many researchers and teams of researchers engage in programs of research that draw productively on multiple methodologies. Further, while many researchers use terms like "mixed methods" or "interdisciplinarity" in describing their work, not all uses of these terms are situated within the AMPs described here.

In the subsections that follow for each AMP, we situate the AMP briefly in historical context, highlight the conceptual framework/analytic language that supports the development of multi-methodological programs of research, acknowledge key differences in perspective to which our theorists point, and offer examples of research

programs by researchers working within each AMP consistent with our theme of research use and engagement. As with the MTs, we have selected prominent theorists for each AMP, from whose perspectives we illustrate the AMP, rather than attempting to provide essentializing characterizations. Further, while we acknowledge that the AMPs can be seen to inform single studies as well as programs of research, our focus is on the relevance of the AMPs to programs of research. We close with a brief analytic summary, which contrasts the affordances of each AMP and considers the ways in which elements of the analytical language might be used, complementarily, to support the development of multi-methodological research programs.

Mixed Methods (MM) Research

In education research, "Mixed Methods" (MM) Research is clearly the best known AMP, routinely foregrounded in handbooks and textbooks representing the field as multi-methodological. There are multiple textbooks (Bergman, 2008; Creswell, 2014; Greene, 2007; Teddlie & Tashakkori, 2009), Handbooks (Tashakkori & Teddlie, 2003, 2010), and a *Journal of Mixed Methods Research (JMMR)* devoted to the discussion about "Mixed Methods" in social research.[42]

The MM discourse portrayed here is often described as beginning in response to the so-called "paradigm wars" (Gage, 1989) of the 1990s (Bryman, 2008; Denscombe, 2008) where the debate focused on whether qualitative and quantitative "paradigms" were sufficiently coherent ("commensurable" or "compatible") to justify mixing and, if not, which provided the more appropriate framework for social research (see Howe, 1988). However, earlier theorizations of work across methods or methodologies were published and cited by MM theorists as precursors to their work. MM theorists often cite Campbell and Fiske's 1959 multi-trait multi-method matrix—a method of construct validation that compares methods for measuring the same and different constructs—as an early theorization of mixed methods research. Other texts appropriated into the history of mixed methods research include Webb, Campbell, Schwartz, and Sechrest (1966), which

[42] There have been variations in the label by which this AMP is characterized—for instance, mixed methodology (Tashakkori & Teddlie, 1998), multimethod (Hunter and Brewer, 2003), multiple methodology (Smith , 2006)—and sometimes mixed methods research has been used as a subcategory of a broader concept. Tashakkori & Teddlie, 2003 referred to "mixed methods (plural) designs as a cover term for mixed method and mixed model research" (p. 11), but it appears that "mixed methods" is widely acknowledged as the covering term. We note as well that the terms "methods" and "methodology" can be used in different ways and readers should consider for each scholar what the terms mean.

called for comparison of unobstrusive measures with surveys and questionnaires in social science research; Sieber (1973), which discussed the integration of survey and field research methods; Denzin (1978), which called for comparisons across researchers and theories as well as methods and data sources; and Cook and Reichardt (1979), which explored ways to combine qualitative and quantitative methods in evaluation research (see, e.g., Denscombe, 2008, for an overview). These texts emphasized a confirmatory approach to mixed methods research often called "triangulation." Subsequent discussions of triangulation in the MM conversation have elaborated (or sometimes renamed) the concept to embrace and seek to understand the *differences* comparison illuminates. And, in fact, MM theory has expanded to incorporate many purposes for mixing methods or methodologies as outlined below.

One set of theorists in the MM conversation—arguably the modal set— portrays mixed methods alongside "qualitative" and "quantitative", as "*the* third…" paradigm, methodological movement, approach, community, or tradition, *explicitly* (Denscombe, 2007; Tashakkori and Teddlie, 2003; Johnson and Onwuegbuzie, 2004), or *implicitly* in the way their texts are organized (Creswell, 2014; Johnson & Christensen, 2008). Other theorists soften or complicate the boundaries implied in this trichotomy (e.g., Greene, 2007; Smith, 2006); and some resist them altogether (Bergman, 2008; Biesta , 2010). Below we illustrate some of the issues underlying these different perspectives. We should note that the word "paradigm" is used in multiple ways in the mixed methods literature. Sometimes it is applied to each element of the trichotomy—qualitative, quantitative, mixed (Johnson and Onweugbuzie, 2004; Tashakkori and Teddlie, 1998; 2009)—and sometimes to the differing philosophical perspectives that might be used to frame mixed methods research (e.g., Greene, 2007; see Morgan, 2007, for an overview of the ways the term "paradigm" is used).[43]

In the inaugural issue of *JMMR,* Tashakkori and Creswell (2007) define MM research as follows: "research in which the investigator collects and analyzes data, integrates the findings, and draws inferences using both qualitative and quantitative

[43] Kuhn 's *The Structure of Scientific Revolutions* (1962) is often cited as the source of the term, "paradigm," although that citation does not clarify it's meaning. As Morgan (2007) notes, Kuhn used the term in multiple ways, something he acknowledged and tried to clarify in his 1970 postscript, and "it is all too easy for social scientists to talk about "paradigms" and mean entirely different things" (p. 50).

approaches or methods in a single study or program of inquiry" (p. 4). The notion that mixed methods research involves qualitative and quantitative methods, in a single study or program of inquiry, across stages of the research process appears typical of definitions by researchers who represent MM as a "third" approach to research. These theorists argue, further, that it is the research question or purpose that should guide the choice of methods.

In pedagogical texts, qualitative and quantitative approaches have often been represented somewhat monolithically as involving particular combinations of philosophical perspectives (or paradigms) and methods.[44] Here, for instance, are Creswell's (2014) definitions of qualitative and quantitative approaches to research:

> A quantitative approach is one in which the investigator primarily uses postpositivist claims for developing knowledge (i.e., cause and effect thinking, reduction to specific variables and hypotheses and questions, use of measurement and observation, and the test of theories), employs strategies of inquiry such as experiments and surveys, and collects data on predetermined instruments that yield statistical data.
>
> Alternatively, a *qualitative* approach is one in which the inquirer often makes knowledge claims based primarily on constructivist perspectives (i.e., the multiple meanings of individual experiences, meanings socially and historically constructed. with an intent of developing a theory or pattern) or advocacy/participatory perspectives (i.e., political, issue-oriented, collaborative, or change oriented) or both. It also uses strategies of inquiry such as narratives, phenomenologies, ethnographies, grounded theory studies, or case studies. The researcher collects open-ended, emerging data with the primary intent of developing themes from the data. (Creswell, 2014, p. 18).[45]

---

[44] Many also acknowledge diversity underlying these terms: For instance, in the preface to their most recent textbook, Teddlie and Tashakkori (2009) say: "We apologize for using the terms *qualitative* and *quantitative* so many times in this book, especially because we advocate that there is not dichotomy but rather a continuum between the terms. We use these terms in many discussions in this book, as proxies for a variety of diverse and complex concepts, constructs, techniques, political/personal ideologies and lenses, and even marketing tools….These terms are perhaps necessary now for pedagogical reasons, but *mixed methods research will have taken a quantum leap forward when they no longer permeate our writings*" (p. viii).

[45] Note that some MM theorists cite advocacy/participatory perspectives as an alternative paradigm for the MM tradition (e.g., Mertens, 2003; Mertens et al., 2010).

Some of the theorists we've cited here also appear to seek a unified vision for the field of mixed methods research, with some suggesting that this vision could or should provide guidance for social research more generally. For instance, in their introduction to the 2010 *Handbook of Mixed Methods Research,* Teddlie and Tashakkori charge chapter authors to "set the stage for a unified field of mixed methods with clear definitions of methodological concepts and innovations that would provide guidance for the next generation of scholars" (2010, p. xiii). This stance is also reflected in arguments for a single paradigm to underlie mixed methods research and in the proliferation of typologies of mixed methods designs.[46] With that goal in mind, Teddlie and Tashakkori nevertheless acknowledge important disagreements among the authors of the handbook about "whether the field is ready to come to consensus on important issues….[and] on the possibility of universal social science principles that are devoid of paradigm considerations, as opposed to the belief that any study must adhere to one (or multiple) paradigm perspective(s)" (p. xiii).

Theorists seeking a unified paradigm for MM research often present typologies—and typologies of typologies—that classify and illustrate mixed methods studies along a number of dimensions. Greene, Caracelli, and Graham (1989) is frequently cited as the progenitor of such typologizing practices. Working from a purposive sample of 57 mixed methods studies between 1980 and 1988, they categorized them in terms of purposes for mixing methods (see quotation from Greene, 2007 below) and in terms of various design elements. Design features on which they categorized the studies included the extent to which qualitative and quantitative methods were: used to investigate the same or different phenomena, framed in terms of the same or different paradigms, equal or unequal in status in the study, implemented interactively or independently and sequentially or simultaneously, and implemented in a single study or a set of studies.

Examples of typologies can be found in any number of texts including Tashakkori and Teddlie (2003, 2009, 2010), Creswell (2014; Creswell at el, 2003), Johnson and Onwuegbuzie (2004); and Morse, 2003. The qualitative/quantititative dichotomy is

[46] We considered the idea of treating MM as a "methodological tradition" with Tashakkori and Teddlie as our key theorists, and, indeed it is represented as a methodology to be learned in many texts. However, it is also clearly, a sustained and multi-perspectival conversation about ways of working across methodologies, so we have chosen to treat it as an AMP.

deeply embedded in the representation of many of these taxonomies. For instance, various combinations of the symbols "QUAL" "qual", "QUAN", "quan", "+", and "→" (Morse, 2003) have often been used to identify generic designs, where "→" indicates sequential use of qualitative and quantitative components, "+" indicates simultaneous use, and upper versus lower case indicates relative emphasis (Morse, 2003). The terms "quantitizing" and "qualitizing" signal transformations of data from qualitative to quantitative (e.g., by coding) and from quantitative to qualitative (e.g., by narrating) to facilitate comparison and integration (Teddlie and Tashakkori, 2009).

Bryman (2006) raises the concern that "the discussion of the integration of quantitative and qualitative research has increasingly been taken over by a formalized approach which is especially apparent in the discussion and proliferation of typologies of integration" (p. 98), especially in North America. He worries that the typologies have become too refined—to the point of "indigestion" (2008), that they imply "forward commitment to type of design" (2006, p. 99) and that they tend to emphasize "logically possible types of integration, rather than being built up out of examples" (2006, p. 98).

Greene's (2007) definition of mixed methods research does not rely to the same extent on the qualitative/quantitative dichotomy. As she describes it, "the core meaning of mixed methods social inquiry is to invite multiple mental models into the same inquiry space for purposes of respectful conversation, dialogue, and learning one from the other, toward a collective generation of better understanding of the phenomena being studied" (Greene, 2007, p. xii). Mental models, for Greene, are unique to each individual, and reflect a confluence of substantive theory, disciplinary perspectives, philosophy of science, methodological traditions (e.g., experimentalism, case study inquiry, survey research, secondary data analysis, and participatory inquiry), education and training, contextual factors, political factors, and personal values (p. 12). She frames her 2007 book as offering a counterpoint to two trends in the mixed methods conversation: (a) the tendency to overemphasize design alternatives in the characterization/typologies of mixed methods, without giving adequate attention to other aspects on which designs might differ, and (b) the goal of developing "an" alternative paradigm for mixed methods (e.g., pragmatism, or transformative research, or critical realism) without exploring the value of dialogue across paradigms within a program of research. She uses the term,

paradigm, to refer to "an integrated set of assumptions about the nature of the social world, about the character of the knowledge we can have about the social world, and about what is important to know" (p. 14). She argues for the richness of mixing on multiple dimensions, including multiple "paradigms". Her preferred approach to mixing paradigms is "dialectic": "Important paradigm differences should be respectfully and intentionally used together to engage meaningfully with difference, and through the tensions created by juxtaposing different paradigms, to achieve dialectical discovery of enhanced, reframed, or new understandings" (p. 69). The purposes she lists for mixing methods (building on Greene et al., 1989) dialectically include the following:

> TRIANGULATION seeks convergence, corroboration, correspondence of results from the different methods.
>
> COMPLEMENTARITY seeks elaboration, enhancement, illustration, clarification of the results from one method with the results from the other method.
>
> DEVELOPMENT seeks to use the results from one method to help develop or inform the other method, where development is broadly construed to include sampling and implementation, as well as measurement decisions.
>
> INITIATION seeks the discovery of paradox and contradiction, new perspectives of frameworks, the recasting of questions or results from one method with questions or results from the other method.
>
> EXPANSION seeks to extend the breadth and range of inquiry by using different methods for different inquiry components. (p. 259).

A mixed methods study by Jang, McDougall, Pollon, Herbert, and Russell (2008), entitled "Integrative Mixed Methods Data Analytic Strategies in Research on School Success in Challenging Circumstances" illustrates the analytic language described above. While the example draws on language from theorists seeking a unified approach to MM, we also see (and point to) purposes that illustrate many of Greene's categories. Jang et al. used a "concurrent mixed methods research" design "to investigate school improvement processes in [20] elementary schools facing various challenging circumstances" (pp. 221-222). Their data collection consisted of a fixed choice survey administered to all principals and teachers in these schools and interviews with principals

and randomly selected teachers from each school as well as focus groups with students and parents from each school (which they refer to as quantitative and qualitative data respectively). The primary focus of the article we cite is in how they analyzed these data. They first engaged in separate "parallel" analyses using factor analysis to empirically identify factors (clusters of items) in the survey and an inductive analysis to develop themes from the interviews and focus groups. They then compared the themes noting areas of overlap and areas where the interviews and focus group data illuminated themes not reflected in the factors from the survey data. They noted, for instance, that the themes of professional learning and data-based decision making processes, among others, were similarly reflected in the interview and survey data and that the survey data "were enriched by contextually rich accounts of the themes from the interviews" (p. 233) [illustrating both triangulation and complementarity in Greene's (2007) terms]. To make the comparisons more explicit, they created narrative descriptions of the factors arising from the factor analysis (qualitizing the data) and displayed them opposite the themes arising from the interview analysis. They noted that other themes, such as parent involvement and attention to children's non-cognitive development, only appeared, initially, in the interview and focus group data. They then reviewed the survey items to match them to themes that had emerged from the inductive analyses, finding items related to eight of the 11 themes. They created a new set of variables from the survey data based on these themes, presenting descriptive statistics with reliability data for each of these themes, acknowledging that some did not have sufficient items associated with them to achieve adequate reliability [illustrating purposes of development and possibly expansion, in Greene's 2007 terms]. They examined correlations among the themes based on the re-analized survey data and were surprised to see that the community outreach theme was negatively correlated with the rest of the themes. They used the interview and focus group data to seek an explanation for this correlation. They noted "the nature of the integration was iterative, moving back and forth between the qualitative and quantitative strands of data and allowing for the recognition of emergent themes and new insights" (p. 241). They also noted the ways in which a sequential design would have improved the study, for instance, using the interview and focus group data to help with construction of the survey [development again].

Some theorists in the MM discourse have raised serious concerns about over-reliance on the trichotomous qualitative-quantitative-mixed language (Bergman, 2008; Biesta, 2010). Biesta, for instance, worries that:

> The terms [qualitative and quantitative] tend to stand for a whole cluster of aspects of research, such as methods, designs, methodologies, epistemological and ontological assumptions, and so on. To use the words quantitative and qualitative to refer to such clusters is not only imprecise but also unhelpful….The notions of qualitative research and qualitative [sic] research tend to obscure those aspects that really matter in the discussion and can even create quasi-problems and oppositions, for example, when researchers who use number and researchers who use text assume that they have nothing to share, even if their research is actually informed by similar assumptions about the nature of social reality or driven by similar assumptions … about knowledge creation." (p. 98).

Similar to Greene, he proposes careful attention to the aspects of research that are being mixed.

Bergman (2008) speaks explicitly to the relationship between mixed or multiple methods research and mono-method research, challenging any notion the MM research would dominate mono-method research:

> "Mono method research will not be replaced by mixed methods research but will, on the one hand, make more powerful and more specific contributions to research in general, and, on the other hand, mono method research will be more fruitfully and more clearly integrateable within a mixed methods research design. Indeed, while mono method research may well exist within mixed methods research, mixed methods research cannot exist without mono method research. It is therefore in the interest of mixed methods researchers to learn from and to work closely with mono method researchers." (p. 3)

Like Biesta, Bergman challenges the qualitative/quantitative distinction in his edited *Advances in Mixed Methods Research,* with multiple examples of the ways in which various, far more nuanced mono-methods work together in addressing social problems. Coburn et al., 2012 (cited in the CCS subsection above), illustrates the productive use of multiple MTs consistent with this advice. They use a qualitative approach to social

network analysis to observe shifts in teachers networks, and in a subsequent study, they use qualitative comparative analysis (QCA) to help in identifying configurations of factors that contributed to successful maintenance of reform after supports were withdrawn.

Multidisciplinary, Interdisciplinary, and Transdisciplinary (MIT-D) Research

As with MM research, Abbott (2001) notes that "interdisciplinarity is old news" (p. 131): "it emerged contemporaneously with, not after, the disciplines" (p. 132). Explicit (meta-disciplinary) discussions of interdisciplinarity have been traced back to the formation of the Social Sciences Research Council (SSRC) in 1923. According to Klein (2007), the SSRC "aimed to accelerate the tendency toward breaking down boundaries by cross fertilizing ideas and joining methods and techniques. It brought together representatives of anthropology, sociology, political science, economics, psychology, statistics and history, with the aim of producing purposive and empirical social problem-oriented applied research" (p. 35). While our emphasis is on working across methodologies, the literature on interdisciplinarity offers multiple conceptual tools for methodological pluralism, both by analogy—think multi-, inter-, and trans-'methodological' research—and by the way in which methodology is treated in discussions of interdisciplinarity.

The literature on interdisciplinarity is huge and spans many fields of work. Theory and research frequently emphasize the social dimensions of research, including the organization of institutions of higher learning, strategies for working across disciplinary communities, relationships between academic disciplines and various sectors of society (e.g., Frodeman , Klein, and Mitcham, 2010; Lattuca, 2001; NRC, 2004b; Repko, 2012;). For this section, we introduce conceptual aspects of the literature we see as most relevant to the design, enactment, and use of social research involving multiple methodologies; in the conclusion we will return to the more social dimensions. There are multiple Handbooks, edited volumes, and textbooks that variously theorize, guide and provide examples of interdisciplinary research (Bammer , 2013; Frodeman et al., 2010; Hirsch Haddorn et al., 2008; Repko, 2012).

The concepts foregrounded in current discussions—multidisciplinary, interdisciplinary, and transdisciplinary-- can be traced to a 1970 international conference

on problems of interdisciplinary research and teaching organized by the Organization for Economic Cooperation and Development (OECD).[47] While different authors' definitions of these three terms differ somewhat, the distinctions among them typically entail different degrees of integration and interaction among academic disciplines and sometimes also between the academic disciplines and other sectors of society (Klein, 2007; Pohl, 2011; Repko, 2012). Interdisciplinarity is sometimes used as an umbrella term, as we have in the opening sentence of this section, and sometimes as a more specific term, representing levels of disciplinary integration that fall between multi- and trans-disciplinary research. We will use it in both ways.

"***Multidisciplinary***" is the term that signals the least degree of integration; it was defined in the OECD report as involving "juxtaposition of disciplines".[48] Following Klein (2007), *"Multidisciplinary* approaches juxtapose separate disciplinary perspectives, adding breadth of knowledge, information and methods. Individuals and groups work independently or sequentially…. They retain their separate perspectives, and disciplinary elements remain intact" (p. 37). Klein (2007, 2010) cites as examples syntheses of published research, conference programs, projects that involve borrowing from one discipline to contextualize another (e.g., to provide historical background), or books that present different views of the same topic or problem in serial order," but without any explicit integration (Klein, 2010, p.17). From this perspective, the "methodological traditions" section of this chapter, with its illustration of different approaches to the study of data use, is multi-disciplinary, or better, multi-methodological in nature.

The OECD Definition of "***Interdisciplinary***" emphasized interaction but was, as Klein notes, quite broad. They characterized an interdisciplinary "group" as "organized into a common effort on a common problem with continuous intercommunication among the participants from the different disciplines" (p. 26). Most current definitions of interdisciplinary research emphasize integration as well as interaction. The National Research Council's Committee on Facilitating Interdisciplinary Research (2004b) defined interdisciplinary research as

---

[47] A fourth term foregrounded in the 1970 OECD conference, pluridisciplinary, is much less used. It was used to describe "juxtapositions of disciplines assumed to be more or less related" (OECD, 1972, p 23), like mathematics and physics, whereas multidisciplinary referred to any juxtaposition of disciplines.
[48] OECD (1972) authors defined "discipline" as "a specific body of teachable knowledge with its own background of education, training, procedures, methods and content areas" (p. 25).

> "a mode of research by teams or individuals that integrates information, data, techniques, tools, perspectives, concepts, and/or theories from two or more disciplines or bodies of specialized knowledge to advance fundamental understanding or to solve problems whose solutions are beyond the scope of a single discipline or area of research practice. (p. 2)"

The NRC report cited four drivers of interdisciplinarity: "the inherent complexity of nature and society, the desire to explore problems and questions that are not confined to a single discipline, the need to solve societal problems, and the power of new technologies." (p. 2). They emphasized goals of fundamental understanding as well as problem solving. And, indeed, one can see interdisciplinarity operating at the level of a field—like learning sciences or organizational studies (Sawyer, 2014; Easterby-Smith & Lyles, 2011)—and at the level of specific problems—like our problem of data use to inform professional practice—and sometimes at the level of a particular problem within a particular context.

As with mixed methods research, many current discussions of interdisciplinary research introduce taxonomies describing nuanced approaches and levels of integration of research (see, e.g., Klein, 2010, for an overview). The typologies frequently grow empirically out of classifications of actual studies, research agendas, long-term projects, and curricula. They tend to foreground issues of what is integrated (e.g., concepts, methodology, or theory), who is involved (a single researcher learning to speak a second disciplinary language or an interdisciplinary team), the extent of integration (from bridge building to restructuring) and how different or compatible the integrated disciplines are. Here we'll highlight one set of categories that foreground key dimensions prominent in multiple discussions about methodology and epistemology: whether the integration is instrumental or critical. Later, in the section on transdiscipinarity, we'll highlight a second distinction between whether the source of the problem is endogenous or exogenous to the university—in other words, whether or not it "generated by 'real' problems of the community" (Klein, 2007, p. 40).

Klein (2007) distinguishes "instrumental" or "strategic" versus "critical" or "reflexive" forms of interdisciplinary research as follows:

> Strategic forms integrate disciplinary, professional and/or interdisciplinary approaches without regard for questions of epistemology or institutional structure. In contrast, 'critical' and 'reflexive' approaches interrogate the existing structure of knowledge and education with the aim of transforming them, raising questions of value and purpose that are silent in strategic forms." (Klein, 2007, p. 39).

While this distinction is sometimes seen to reflect a "faultline", Klein (2007) notes that the difference is not absolute. Many enactments of interdisciplinary research involve aspects of both.

We use a recent report by a committee of the National Research Council (2012) on "Using Science as Evidence in Public Policy" to illustrate one type of interdisciplinary work arguably reflecting both instrumental and reflexive approaches. The committee's goal was to "develop a framework for further research that can improve the use of social science knowledge in policy making" (p. 12) based in part in a review of relevant literature. First they noted that while "science should be at the policy table…, it shares that table with an array of nonscientific reasons for making a policy choice: personal and political beliefs and values…, lessons from experience, trial and error learning, and reasoning by analogy, [as well as] …political pressures (p. 3) The framework was intended to illuminate "what happens during policy arguments, with a specific focus on whether, why, and how science is used as evidence in public policy.' (p. 63). While they began with a review of the knowledge utilization literature, they considered as well the affordances of a range of disciplines and methodologies of social science that could inform their framework. They imagined a program of research that drew simultaneously on studies of policy argumentation, psychology of decision making, and systems perspectives, each involving a range of research discourses. They focused their presentation primarily on the sorts of questions or issues that could or should be addressed to understand the phenomenon of use of social science knowledge in policy making. Under policy argumentation and decision making, they cited a range of disciplines and fields of research, including cognitive, situative, and social psychology, behavioral economics, organizational sociology and organizational learning to address issues like the following:

> how mental models, belief systems, organizational rules, societal norms, and other

> factors influence the behavior of decision makers. …how people learn, when they optimize and when they satisfice; why they organize themselves, form institutions, communicate, establish norms, and develop routines; how they assess risks; and how they make decisions, individually and collectively. …. the cognitive operations and biases that policy makers and scientists bring to their work and the context-specific situations, practices, logics (ways of reasoning and understanding), and cultural assumptions of the settings in which they operate. (p. 4)

Under systems perspectives, pointed to a range of fields that address questions about:

> the many effects a policy may produce and the ways in which a planned social intervention interacts with other existing interventions and institutional practices; [and] …systems effects on individual actors and the system as a whole, including emergent, indirect, and delayed effects, as well as unintended and unpredictable consequences from the interactivity of a system's elements. (p. 5)

Given their inter-*disciplinary* orientation, the main section of the report privileges the research questions that different disciplines and fields of research can address, with occasional brief references to methodologies. As the above research questions suggest, however, the methodologies include many of the same ones we've named in the MT section.

Theories of ***transdisciplinarity*** take interdisciplinarity a step farther in at least one of two ways. In fact, the term seems to have taken on two somewhat distinct meanings, depending in part on the scope of phenomena addressed in the integration, in part on whether its source is endogenous or exogenous to the university, and in part on the role in which stakeholders outside the university play in the research process.

The initial OECD (1972) characterization of transdisciplinarity focused on "establishing a common system of axioms for a set of disciplines" (p. 26). From this perspective, the goal is to build "a comprehensive general view, a theoretical synthesis, or an integrative framework" across disciplines (Klein, 2007). An example often cited is that of complexity theory (elaborated below) which provides a framework for studying complex adaptive systems across multiple types of systems and disciplinary contexts. As Klein notes, while this reflects "the contemporary version of the ancient quest for

systematic integration of knowledge, [it is] not in the name of a single totalizing theory but new paradigms that recognize complexity and difference" (p. 40).

More recently, a different characterization of transdisciplinarity has taken priority, one that both privileges problems generated in the life world and that involves a range of stakeholders impacted by the problem in the framing, design and enactment of relevant research. Transdisciplinary research is most needed, it's argued, "when the concrete nature of problems is disputed, and when there is a great deal at stake for those concerned by problems and involved in dealing with them" (Pohl and Hirsch Hadorn, 2008, p. 431). At the most general level, these have included problems of poverty, sickness, crime, and environmental degradation, although they are often investigated in specific contexts where consequential decisions about policy and practice must be made.

Following Wiesmann and colleagues (2008), transdisciplinarity "acknowledges that knowledge…exists and is produced in society in fields other than science" (p. 435) and therefore focuses on the links between different sciences and other parts of society…[including] deliberation about facts, practices and value (p. 435). Transdisciplinary research "connects problem identification and structuring, searching for solutions, and bring results to fruition in a recursive research and negotiation process" (p. 436). Discussion of transdisciplinary research often focuses on these distinct aspects of the research process and what sort of integration/interaction is involved at each stage.

As Pohl and Hirsch Hadorn (2008) note, transdisciplinary research "takes into consideration a large array of potential disciplinary paradigms and life-world perspectives, and it explicitly narrows down its focus to a few of them in the phase of identifying and structuring problems" (p. 37). The goals of transdisciplinary research are "(a) to grasp the complexity … of problems, (b) to take into account the diversity … of life-world and scientific perceptions of problems, (c) to link abstract and case-specific knowledge, and (d) develop knowledge and practices that promote what is perceived to be the common good " (Pohl and Hirsch Hadorn, 2008, pp. 431-432).

While the language of transdisciplinarity is only just gaining a foothold in education research (e.g., a search of *AERJ* on 1/29/14 yielded only three references to "transdisciplinary"), there are multiple synergies between transdisciplinary and participatory action research, a point that a number of transdisciplinary theorists

acknowledge. [49] Thus the study by Wyman et al. (2010) we used to illustrate PAR also serves as an illustration of the conception of transdisciplinary research that draws on knowledge from scientific disciplines and other sectors of society.

Complex Systems Theory (CST) as an Approach to Methodological Pluralism

In an earlier section, we focused on a methodology that originated within the field of complex adaptive systems research—computational modeling, especially agent based modeling (ABM)—which provided tools for addressing questions about how collective behavior of a social[50] system can "emerge" from the interactions among its agents. As we noted, a complex system is one that is composed of multiple heterogeneous agents who interact with one another over space and time. …A complex system becomes "adaptive" when the "interaction of the agents is contingent on past experience and…[the] agents continually adapt to that experience (Axelrod & Testfatsion, 2006, p. 1649). We noted further that while agents were often humans or groupings of humans, they could also include non-human interacting agents from neurons to the many conceptual and material tools that agents use to constellations of stars. From this perspective, all the phenomena in which we are interested in this chapter both constitute and are embedded in complex adaptive systems.

In this section, we follow the lead of social and educational researchers who use complexity theory or complex systems theory (CST) conceptually--as a frame of reference, heuristic, metaphor, or way of thinking--for analyzing complex adaptive social systems (Byrne & Callaghan, 2014; Larsen-Freeman & Cameron, 2008; Morçöl, 2013).[51] Byrne and Callaghan (2014), following Morin (2008), frame this distinction as

---

[49] Connections are often made between transdisciplinary research and "Mode 2" knowledge, a term originated by Gibbons (1994 in Gibbons & Nowotny, 2001) and elaborated by Nowotny, Scott, and Gibbons (2001) (see Klein, 2007, for a comparison). Gibbons and Nowotny (2001) argue for : "…new links between traditionally segmented producers and users of knowledge. Contextualization of research around the interests of stakeholders fosters a more 'socially robust' knowledge that transgresses disciplinary and institutional boundaries'" (p. 67) We are grateful to Lisa Lattuca for drawing our attention to this connection.

[50] It's important to note that complexity theorists use the word "social" somewhat differently than we have used it in characterizing social research. While we, following Camic et al. (2011), focus on research involving humans' interaction with one another and with elements of their environment, the sort of complex adaptive systems that complexity theorists label "social" can include everything from the interactions among neurons in human brain cells, to bird flocks and insect swarms, to markets in the global economy, to galaxies.

[51] While Miller and Page (2007) caution us about the distinction between "*amorphous metaphors* and cutting edge research that requires a technical background" (p. 6), they note how "the tools and ideas

between "restricted" complexity, of which ABM is an example, and a more "general" complexity in social research. In general, complexity theory helps us imagine how different methodologies might work together to enhance our understanding of social systems, including those in which teachers and students participate, where consequential gaps in our understanding might exist, what (combinations of) levers might enable change, and with what consequences. In essence, CST supports us in mapping the systems potentially relevant to the social phenomenon of interest and then bounding those aspects that are (arguably) crucial in addressing our research questions.

Complexity theory can and has been used to guide the design of comprehensive programs of research, to organize pre-existing research to support integrative understandings, and to guide management (teaching and leading) and decision-making in educational systems. The central question here is "can the process(es) I am examining be understood and interpreted better if I think about them as complex adaptive systems and does thinking about them in that way help me to understand their nature and to inform managers and decision makers how to act effectively in relation to them?" (Byrne and Callaghan, 2014, p. 232). The goal is not to replace existing theories—indeed many call for an interdisciplinary approach to research within a complexity frame (e.g., Page, 2010; Byrne & Callaghan, 2014; Cochran-Smith et al., 2014)—but rather to "work at a supra-disciplinary, more abstract level offering a general set of patterns, outcomes, and descriptions that can be applied to many different sorts of systems" (Larsen-Freeman & Cameron, 2008, pp. 15-16), considering the full range of disciplinary theories and methodologies necessary to support sound practice.

Complexity theory highlights the dynamic and interconnected nature of the "web" of systems relevant to education. It draws our attention to evolving interactions (a) among agents—individual and collective—and other elements within a focal system and (b) between a focal system and other systems with which it interacts and co-adapts. The complex adaptive systems operate at "various levels of granularity, from macro-level to micro-level; for example, from ecosystem to sub atomic particles" (Larsen-Freeman & Cameron, 2008, p. 240). The focus is on continual change and co-adaptation. Even the

---

emerging from complexity theory…should allow us to build much better theories about the world *when they are carefully integrated with existing techniques*" (p.6, italics ours).

apparent stabilities in a system—the entities like students, teachers, classrooms, and schools—can be seen to emerge "from the dynamics of the system" (Larsen Freemen & Cameron, 2008, p. 9); "they exist only through the fluxes that feed them [interacting students, teachers, and leaders], and they disappear or become moribund in the absence of such fluxes (Larsen-Freeman & Cameron, 2008, p. 2).

In conceptualizing a program of research around a given phenomenon (which may itself be a complex system or an element of such a system), one goal is to bound those aspects of the relevant system(s) necessary to understand how the phenomenon functions and how the focal system(s) of which it is a part might be enhanced. Another goal (often enacted iteratively with the first) is to analyze the focal system, naming its elements, the levels and time scales on which they interact, the other systems with which the focal system interacts, and the relationships (links, information flows) within and between them. Potentially relevant systems range from individual minds up through sociopolitical contexts, and from minute-to-minute time scales through lifetimes of teaching and learning (Larsen-Freeman & Cameron, 2008, p. 198). Consider for instance the instructive questions Lemke and Sabelli (2008) raise about students' learning that cut across levels and time scales:

> How do brief actions by teachers and students add up to coherent activities over periods of minutes and hours, days and months? How do curriculum change processes that occur over periods of years exchange information with classroom activities that occur over periods of minutes? How do learning events in a laboratory or at a computer workstation and those in classrooms and hallways and cafeterias add up to a coherent longer-term process of educational development, or perhaps the development of facility with a particular concept? How do networks of social interaction with peers in the classroom, in the wider neighborhood community, and in virtual online communities contribute to long-term processes of identity development and formation of lasting attitudes and values, which affect decisions and actions on very short timescales? How do the changing priorities, populations and problems of a local community influence the larger educational system's agendas and programs? (p. 115).

Of course, one cannot consider the whole web of inter-related systems at once, but where one draws the line around a focal area needs to be carefully considered and justified in light of these interconnections. "Before we can recognize meaningful parts we must define the functional whole of which they are the constituents. We will then recognize quite different ways of breaking up the organism depending on what we are trying to explain." (Lewontin, in Larsen-Freeman & Cameron, 2008, p. 235). Third and fourth goals include deciding how to productively model the system (where multiple methodologies come into play) and how to intervene to change the system where a range of potential sets of change levers can be considered. These sets of change levers are often referred to as "control parameters".

As we noted in the section on ABMs, "cause operates in any and all directions" (Byrne and Callaghan, 2014, p. 190) from pre-existing processes and structures to agents, from agents to agents linked in various ways across time and space, from agents to emerging phenomena at higher levels of the system ("emergence"), from emerging phenomena back to agents, and to and from other social systems that co-adapt with the focal system. The emergent phenomena can be ephemeral, guiding only immediate actions and interactions, or they can become stable and continue to exercise causal power in new circumstances (Sawyer, 2005, 2007). "We are always dealing "with multiple interacting causes and…causal sets" (Byrne & Callaghan, 2014, pp. 189-190). These interacting causes are further "mediated by the meanings and values assigned by the actors, individually and collectively" (Lemke & Sabelli, 2008, p. 116). Thus, we need to attend simultaneously to the individual and the group or organization as each influences the other, and we need to attend to different levels and time scales. To focus on a single causal variable--like a curriculum, or a required use of value added scores, or a program to support teachers' inquiry into their students' learning--is insufficient. As Lemke and Sabelli (2008) note, "any focal pedagogical 'innovation' introduced into a tightly constrained school system is in fact a series of embedded innovations at levels above and below the focal intervention, and strategies for all levels have to be considered coherently" (p. 116). One must consider the web of relationships of which it is a part to understand "how, when, where, and why" it enhances or fails to enhance teaching and learning in any given context.

From this perspective, causal questions focus on the state of the system—"the character the system has at any given time point in its trajectory" (Byrne & Callaghan, 2014, p. 190)—and the multiple interacting causal variables and sets of causal variables or "control parameters" from which that state emerged. By "control parameter…we mean the causal set …which includes elements, which may themselves be complex, which in interaction with each other determine the state of the system" (Byrne & Callaghan, 2014, p. 190). The key question is "what can we do to move any human social system… toward a desired state?" (Byrne and Callaghan, 2014, p. 197). Change can often result from a shift in one element of a control parameter as illustrated in the example from Larsen-Freeman and Cameron (2008), where a small shift in a teacher's questioning routine changes the nature of her classroom interaction. Put in other terms, Mason asks:

> If education is about fostering the emergence of learning, of creativity, of imaginative and critical perspectives, then educators would be fair in asking of complexity theory how we might set about establishing, or at least contributing to the establishment of, the conditions, insofar as it is possible to influence those conditions, for emergence to occur. (Mason , 2008, p. 43)

A useful notion is that of an "***attractor state***," referring to any of one or more equilibrium (stable) states toward which the system tends—which may or may not be productive in terms of the goals of those asking the questions. The question for the change agent is how to move the system from one attractor state toward a new and more desirable attractor state. Relevant control parameters or levers for change can be explored at any level of the system and involve any of its components and interactions or relationships among them. For managing the system, Lemke and Sabelli note "Active adaptive management is a 'process of testing alternative hypotheses through management action, learning from experience, and making appropriate change to policy and management practice'"(Lemke & Sabelli, 2008, p. 113 [internal quote not cited]).

By way of example, Larsen-Freeman and Cameron (2008) sketch the range of complex adaptive systems that are implicated in a Tuesday afternoon lesson in a second language classroom. What can be seen at this given point in time are, in fact, points on the evolving trajectories of multiple interacting systems. Their example then continues to represent an "attractor state" of the classroom system inconsistent with the goals of the

curriculum and a small shift in a part of a control parameter that leads to a more productive attractor state. Their example shows how a complexity-informed analysis could be used to improve the learning opportunities in the classroom.

- The language class or lesson event that takes place that Tuesday afternoon can be seen as the trajectory of a sociocognitive complex system. The elements and agents of the system include the learners, the teacher, the books, the language items being used, the physical environment. Whole class talk and group talk are just two modes of interaction from the many types of relations between the learners as agents that occur during the classes....
- When four or five learners come together as a group to carry out a task, the groups acts as a coupled system with its own dynamics emerging from the adaptations of the individual system. ….Collective variables [i.e., variables characterizing emergent properties of a complex system considered at a higher level of aggregation] will describe the action of the group, while the task requirements act as a control parameter that can shift the nature of that action....
- If we go to a lower level or scale from the lesson, the action of each person in the class can be seen as the trajectory of a complex system over a state space. Each engagement in action provides language exemplars and language use that may stabilize attractors in the state space as learning emerges across timescales from use…. Complexity thinking can be extended to constructs usually seen as individual factors but in need of connecting to other timescales and levels of human organization, such as personality, motivation…, style of learning or teaching, ability, intelligence, background, etc. Each of these might be understood as arising from self-organizing systems and as dynamic. (Larson-Freeman and Cameron, 2007, pp. 201-202).

In their characterization, they point further to lower and higher levels of interaction, including those going on inside bodies (e.g., the brain), as well as larger systems like the school (with all its interacting agents and elements), the curriculum, the sociopolitical contexts, and so on. They note as well that artifacts produced in one system yet taken up in another (like the curriculum) are not fixed, but rather emerge from the interactions of teachers and learners. "In applying the complexity lens to an aspect of classroom action,

we will need to select, out of all that is connected and interacting, particular systems to focus on. Other aspects of systems become the dynamic environment on which these focal system operate, but are still connected to and able to influence them" (p. 203). So, even when we've on settled particular systems, we need to be open to other potential influences to and from other systems.

Larsen-Freeman and Cameron then used a complexity frame of reference to observe unfolding interactions among a teacher and students in the language learning classroom. They were trying to understand why students were routinely providing only short answers to questions in spite of a curriculum and a teacher intending to encourage extended turns of talk. They cited an interaction between a teacher and student where the student did not respond to the teacher's initial open-ended question and the teacher asked a series of short answer questions to scaffold the student's response; they noted an increase over time in interactions like this; and discovered that students were choosing to wait, following an open ended-question, for the scaffolding that routinely followed. Thus these minimal responses could be seen as an "attractor state" in the CAS of the classroom, and an unproductive one in light of the goals of the classroom.

While teachers spoke of students as the locus of the problem: "'they only give short answers', the problem was observed to concern both teachers and students in their adaptation to each other" (p. 215). The goal of the intervention was to "[perturb] the system out of its stabilized attractor state" (p. 215). The researchers proposed a collective variable to trace shifts in this interaction system "derived by comparing the actual language used by the learner with the expected language as set up by the teacher's utterances" (p. 208), scaling the difference between the cognitive demands of the teachers' initial utterance and those of the students' subsequent utterance. They explored a set of strategies, "control parameters" encompassing teachers' awareness, teachers' explaining to the students what was expected, wait time, and enhancing students' content knowledge so they had more vocabulary on which to draw. And they observed a shift to a new attractor state where the students' utterances were more in line with the teachers' initial invitations.

A number of authors have provided systematic guidance for analyzing a focal system and its relation to other systems to support the articulation of a multi-

methodological program of research within a CST frame, Larsen-Freeman and Cameron (2008) offer the following guidelines for analyzing complex social systems, guidelines that we have re-ordered and categorized to support readers in using them:

Steps to Guide Complexity Thought Modeling of Complex Systems (Adapted from Larsen-Freeman and Cameron, 2008)[52]
[[FOCAL SYSTEM OVERVIEW]]
Identify the different components of the system, including agents, processes, and subsystems.
For each component, identify the timescales and levels of social and human organization on which it operates.
Describe the relations between and among components.
Describe the state space landscape of the system *["A state space is a collection of all possible states of a system; each point in the landscape represents a state of the system" (p. 46).]* • Where are the attractor states in the state space (i.e. stabilities in the changing system)? *["An attractor is a region of a system's state space in which the system tends to move" (p. 50).]….* • How stable are the attractor states?
Identify regions of the state space that are most used by the system, and those which are seldom visited. (i.e. What does the system do out of all it could possibly do?)
Describe what happens around attractors. (i.e. What kind of variability is there around stabilities?)
[[FOCAL SYSTEM DYNAMICS]]
Describe the dynamics of the system: • how do the components change overtime? • how do the relations among components change over time?
Describe the kinds of change that can be observed in the system: steady change or discontinuous leaps from one state, or mode of action, to another….
Describe the trajectory of the system in its state space. (i.e. What are common patterns of activity?)
Identify possible emergence and/or self-organization across timescales and/or levels of human organization.
[[INTERACTIONS WITH OTHER SYSTEMS]]
Identify the contextual factors that are working as part of the system.
Describe how the system and context adapt to each other.
Identify processes of co-adaptation with other systems.
[[STUDYING AND PROMPTING CHANGE]]

[52] We have partially reordered the table, collapsed some items, grouped items into categories, and provided names for the categories in centered header rows with double brackets (which Larsen-Freeman [personal communication, 11/20/14] indicated represented their model). Material in italics and single brackets is taken (quoted or paraphrased) from elsewhere in their text to gloss the more technical terms they use.

Identify candidate control parameters, i.e. the motors of change that seem to lead to phase shifts *[when the system "changes to a new and radically different mode"]*.
Identify candidate collective variables that can be used to describe the system, before and after phase shifts. *[A collective variable "brings together elements of the system that work together…can only be made for the collectivity…and does not apply to individual elements" (p. 61). It "offers ways of describing the relations among elements in a complex system; its trace over time reveals points of emergence and self-organization. [e.g., length of traffic jam, applies to collectivity and not individual elements] (p. 61)]]*

(Adapted from Larsen-Freeman & Cameron, 2008, pp. 70-71)

Lemke and Sabelli (2008) offer a complementary set of guidelines for analyzing systems, foregrounding, in particular, relationships and information flows among ***levels***. With respect to given level of organization, they ask: "What next higher level of organization determines constraints on the dynamics at the focal level?" and "What units of analysis at the next level below interact to constitute units (or processes or patterns) at the focal level?" (p. 115)

Almost all of the methodologies we reviewed in the previous section—ethnographic research, discourse analysis, Small-N research, survey research, design based research, social network analysis, and participatory action research have been utilized within a complexity frame of reference. Ironically, the one methodology which is routinely challenged by complexity theorists—and the contrast used to illuminate the important contributions of complexity theory to our understanding of causality—is experimental and quasi experimental designs for generalized causal inference. This same methodology is the one most frequently valorized in education as the "gold standard" for causal research. Both judgments, negative and positive, are related to GCI's static focus on a single causal variable and relegation of unexplained variance to error. Morçöl, (2013) however sees a role for GCI in the study of complex systems, as do we, so long as it is used in conjunction with other methodologies. One framework that has been used for organizing different methodologies focuses on time scales and the relations among them and the extent to which they take status or process into account. Morçöl, for instance, in the context of public policy, focuses on different methodologies relevant to macro phenomena, micro phenomena, and the relationships between them (which accounts both for emergence, where one traces how system level phenomena are produced from interactions among agents and other components, or vice versa where

system level phenomena enable and constrain those interactions). A complementary framework focuses on the extent to which the methodology takes status or process into account. Others have distinguished the goals of on-going management from research intended to contribute to more general knowledge (Cochran-Smith et al., 2014).

Macro level phenomena, with slower time scales and scopes that encompass multiple systems, might be illuminated by experimental and quasi-experimental design or survey research, but also by case studies (e.g., of national education systems) and historical narratives (methods not foregrounded here). Micro level phenomena, often apparent in moment-to-moment or day to day interactions in particular contexts, might be illuminated by discourse analysis and ethnographic research, but also by, for instance, laboratory experiments, cognitive labs, and design experiments. Relations between levels might be illuminated by social network analysis, small-N analysis, ethnographic research (again), discourse analysis (again), agent based modeling and actor network theory (Latour, 2005; Fenwick and Edwards, 2010). Such relationships include the way in which institutional logics (like the press for evidence based practice) and related policies are taken up in local contexts and how local practice gets amplified, aggregated, and reified to produce institutional logics that in turn impact multiple local contexts (all instances of "emergence"). Methodologies like design research, ethnography, or participatory action research might focus on multiple levels of the system. Of course what counts as micro and macro varies by the focal system under consideration. Representations of the role of different methodologies within a complexity frame of reference can be found in Larsen-Freeman and Cameron (2008), Byrne and Callaghan (2014), and Morçöl (2012). Thus, many of the issues addressed in complexity theory are addressed in other research discourses in other terms. "As is so often the case in thinking about complexity, something already being done can be framed in complexity terms" (Byrne & Callaghan, 2014, p. 223).

Analytic Summary

The resources we've reviewed in this section offer (a) a range of purposes for methodological pluralism, (b) structures and strategies for engaging it—including consideration of which communities of researchers and other stakeholders participate, when and how they participate, and what level of integration across perspectives is

entailed, and (c) how the system(s) of which the phenomenon of interest is a part might be analyzed or "mapped" in developing a program of research.

With respect to purposes, drawing most heavily on the MM discussion, especially Greene (2007), we highlight

- development, such that earlier studies inform the design of later studies;
- triangulation, for corroboration or convergence, but also for illuminating differences for further study;
- complementarity or configuration, to illuminate different aspects of the phenomena of interest and the systems within and through which they function;
- challenge or confrontation, to illuminate taken for granted perspectives and allow them to evolve.

With respect to structures and strategies for engaging pluralism, drawing most heavily on discussions of MIT-D (especially Klein, 2007, 2010), we highlight differing:

- research communities that reflect different methodologies, theories, philosophies etc.;
- communities of stakeholders who work with(in) the phenomenon of interest; and
- publics who are impacted by the work;
- timing and degree of dialogue and integration; and
- authorities to shape the research.

Finally, with respect to strategies for mapping systems implicated by a phenomenon of interest, we highlight the following interacting components and systems as reminders of the far more elaborated guidelines provided in the CST section (by Larsen-Freeman & Camerson, 2008):

- identification of a focal system and analysis of its components, including people (agents, participants); conceptual and physical objects; subsystems; levels and types of organization or structure, both formally designated and emergent; and activities and processes;
- other related systems that might be implicated, their relationships with the focal system, the ways in which people and objects cross boundaries between systems;

- the evolving interactions and relationships among these components/systems and the different time scales on which they occur, from moment-to-moment micro-level interactions among people to slower macro-level interactions among organizations and agencies;
- the ways in which interactions at one level of a system enable and constrain interactions at other levels, including how group or system level features emerge from interactions among elements (e.g., routines, building architectures) and how these emergent phenomena in turn shape subsequent interactions; and
- how the focal and related systems change over time.

In our fourth section, "Heuristics for Engaging Methodological Pluralism", we will consider the implications of these AMPs for designing multi-methodological programs of research, following our discussion of how different conceptions of generalization support multi-methodological research

4. CONCEPTIONS OF GENERALIZATION (CoGs)

Learning from diverse methodological traditions (MTs) requires attention to how knowledge produced in one set of circumstances—particular cases, contexts, time frames, studies, theories, MTs—can be combined with knowledge produced in other circumstances to enhance understanding and action. These are questions of how knowledge generalizes, including how knowledge can accumulate across studies and support practice in local contexts. In our review of different MTs, we noted how our key theorists conceptualized the issue of generalization within that tradition. Here we move to a different level of abstraction, and consider questions of how knowledge generalizes across studies, with particular attention to conceptions of generalization that can encompass different MTs. The scholarship we review in this section thus complements that reviewed in the approaches to methodological pluralism (AMP) section. While the AMP section provides resources for developing and pursuing research agendas that encompass multiple methodological traditions, this section provides resources for considering how generalizations can be constructed across the particulars of different studies, both to contribute to general knowledge and to inform practice in particular contexts.

We use the term "general knowledge" as a rough and ready placeholder for knowledge that is available for use in multiple contexts. Within this chapter, we refer primarily to the kind of general knowledge to which social research contributes. By "particular" or "local," contexts, we do not simply mean classrooms, schools, and districts; we also mean, for instance, legislatures, offices in national and international agencies, professional development programs, research discourse communities, and collaborations among them. While the reach of some decisions is broader than others, decision making always occurs in particular circumstances in interaction with particular people and material resources. Thus, by practice, we encompass the differing practices of all the professionals who engage with research in these contexts.

Generalization has often been conceptualized as relying on a sample to population logic: if cases are randomly sampled from a clearly defined population, then it is argued the study's conclusion can be generalized to the population with a known degree of precision From this perspective, the significance of small N and single case studies

employing purposeful sampling must be argued in other terms. However, it's important to note that all studies, including those involving random sampling are, nevertheless, situated in time and place, and framed in particular ways (e.g., Campbell and Stanley, 1963; Gomm et al., 2000a; Ragin and Becker, 1992; Yin, 2014). From this perspective, all studies can be considered cases or case studies (Ragin, 1992). In order to develop knowledge across multiple studies, a different conception of generalization is needed, one that relies on *theory* to carry the generalization. The theory-based conceptions of generalization we present in this section are, we argue, fundamental to practice and progress of social research.

In this section, then, we first present a range of logics that rely on theory to construct and warrant generalizations across cases or studies. These conceptions of generalization explore different relationships between cases and theory, some privileging the particulars of individual cases and building theory inductively, some privileging hypotheses reflected in a priori theory and using cases to test and elaborate the theory. These logics can be used prospectively, as a study is conceptualized with an eye toward its contribution to general knowledge, and retrospectively as multiple studies are analyzed and synthesized. We then turn directly to the burgeoning methodological literature on conducting research syntheses or systematic reviews of research, which draws on and extends the logics of generalization from the previous section. While we address some conceptions of generalization relevant to particular MTs, we emphasize conceptions of generalization that can work across multiple MTs. Finally, we turn from these more product-oriented conceptions of generalization to the ways in which knowledge or "knowing" might be constructed and put to work to inform policy, practice, or subsequent research. Here we draw on theories of professional and organizational learning, considering the status of general knowledge in local contexts, the relevance of local knowledge, and the kinds of structures and cultures that support evidence-informed practice in context.

Readers wanting a more comprehensive comparison of the logics of generalization, including those that involve sample to population logics and case to case transfer, can turn to the work of scholars who have compared and critiqued approaches to generalization (Byrne, 2009; Eisenhart, 2009; Ercikan and Roth, 2009, 2014; Gobo,

2008; Gomm, Hammersley & Foster, 2000b; Ragin, 1992, 2009; Tsoukas, 2009). Here we emphasize those theory-based logics that support generalizations *across cases or studies*, both within and across MTs.

Theory-Based Logics of Generalization

The logics of generalization we review here all depend on theory of one sort or another to carry the generalization, and they work iteratively between cases and theory. Different types of theory entail different logics for relating the particular (case or cases) and the general (theory)—some privileging similarities among particulars and theories that articulate what they have in common, some privileging differences among cases and theories that illuminate these nuances. Because of the central role of "theory" in all of these approaches, we begin with a brief overview of differing conceptions of theory.

Differing Conception Of Theory

Given the variations with which the term is used, some authors refuse any general definition of theory and list, instead, a range of types (Abend, 2008).[53] Others do attempt more general definitions but quickly move beyond them to characterize the differences the term subsumes. Here are two such definitions within which the range of logics below can arguably be located:

> A theory states relationships between abstract concepts and may aim for either explanation or understanding (Thornberg & Charmaz, 2012 in Charmaz, 2014, p. 228).
>
> A theory, in both everyday and scientific use, is normally used to denote a model or set of concepts and propositions that pertains to some actual phenomena; a theory can provide understanding of these phenomena or form the basis for action with respect to them. (Maxwell & Mittapali, 2008).

One distinction that's routinely drawn, and is reflected in the definitions above, is between explanatory theory and interpretive theory. Explanatory theory tends to focus on "why" questions and to seek causes of outcomes, events, or states of a system; these include the sort of general causal questions about particular variables or interventions

[53] Abend (2008), for instance, illustrates multiple ways in which the term has been used in sociology-- for instance, theory focused on propositions and relationships among them, on explanations (including but not limited to causal explanations), on interpretations of meaning, on perspectives from which one views the world, and on normative and political accounts.

addressed by GCI, but also questions that privilege explanation in terms of intention, in terms of macro social forces beyond the purview of individual actors, and in terms of complex interactions within and between systems. While explanatory theory may be the modal understanding of theory (Maxwell and Mittapali, 2008) it is not the only approach to theory. Non-explanatory or interpretive theory typically foregrounds understanding, both of researcher and researched. As Charmaz (2014) describes it:

> Interpretive theories aim to understand meanings and actions and how people construct them. Thus these theories bring in the subjectivity of the actor and may recognize the subjectivity of the researcher. ….
>
> What do people assume is real? How do they construct and act on their views of reality? Knowledge and theories are situated and located in particular positions, perspectives, and experiences. (p. 231).

Non-explanatory theories can also provide frames or lenses through which researchers can view the social phenomena (Alasuutari, 1996), including normative and political accounts (Abend, 2008). They can also refer to maps, models, or typologies of social phenomena. As Schatzki (2001) describes it, non-explanatory theories can be:

> typologies of social phenomena; models of social affairs; accounts of what social things (e.g., practices, institutions) are; conceptual frameworks developed expressly for depicting sociality; and descriptions of social life-so long as they are couched in general, abstract terms. (p. 13).

Our work in this chapter is arguably an act of non-explanatory theory building in the ways described here. Both explanatory and interpretive conceptions of theory are relevant to empirical studies and reflected in the logics of generalization we present below.

Another distinction routinely drawn that, arguably, cuts across explanatory and interpretive categories, can be framed in terms of the level of "abstraction" entailed in the theory. Levels of abstraction refer, initially, to the distance between phenomena being studied and the concepts/categories in which they are represented or, in other words, to the comparative scope of the concepts/categories. Levels move from concepts closely tied to the ideographic details of a particular case (which necessarily involve some level of abstraction in the very act of describing), to comparative concepts that refer to similarities and differences across cases, to typologies that group cases or features of

phenomena studied within domain, to mega- or trandisciplinary concepts that work across domains (Byrne, 2009; Geertz, 1973; Goeken and Borner, 2012, Mjøset, 2009; Tsoukas, 2009). Theories, which explore relationships among concepts, have also been categorized in terms of level of abstraction. Categories commonly used are: grand theory, middle-range theory, and case-specific or "grounded" theory[54]. Case specific theories stay very close to the empirical materials at hand, perhaps moving up only one level of abstraction (e.g., from a transcript of interaction among teachers, to a characterization of the interaction as, for instance, one where teachers are offering alternative interpretations of students' work, a category that could then be used to note other similar instances). Middle-range theories, a concept developed by Merton (1967),[55] can be applied to different contexts but "operate in a limited domain…[representing] attempts to understand and explain a limited aspect of social life" (Bryman, 2008, p. 22), like our domain of research use and engagement by teachers. Grand theories focus on a higher level of abstraction and provide frameworks or lenses from which one can view the world. While these are not the distinctions we use to organize this section, we will reference them as needed.

The categories we use below reflect differences in the way theory and cases are related. We begin with more deductive forms of theorizing (testing theory, privileging the general), including replication, systematic variation, and analytical generalization, where the theory is typically specified a priori and cases are used to test and elaborate the theory. Then we move to more inductive forms of theorizing (developing theory, privileging the particular), including analytical induction, grounded theory, analytical refinement, and essentialist theory where theory is built from the case(s) at hand. A third role for theory, which spans these distinctions in different forms, involves process tracing, typologizing, and configuring of the mechanisms that produce outcomes or states of the system. All of these perspectives involve comparison and categorization of some

---

[54] Sometimes theorists use the word "grounded" to refer to the tradition of Grounded Theory as evolved from the work of Glaser and Strauss (1967) and sometimes they use the word "grounded" more generally.
[55] According to Merton, "theories of the middle range are theories that lie between the minor but necessary working hypotheses that evolve in abundance during day-to-day research and the all-inclusive systematic efforts to develop a unified theory that will explain all the observed uniformities of social behavior, social organization and social change." (Merton, 1968 in Hedstrom & Ylikoski 2010, p. 61)

sort and all involve movement between the empirical and some level(s) of abstraction. They entail primarily middle range and grounded theories although they can be informed by more abstract or grand theory and can, in the long run, contribute to such theory.

More Deductive Forms Of Theorizing Emphasizing Theory Testing

We begin with a cluster of logics where theory plays a primary role and cases are used to challenge or test the theory, which evolves accordingly. The purpose is most frequently the generalization of causal effects on particular outcomes.

One logic of generalization has been closely associated with experimental and quasi-experimental design (our GCI): *replication and systematic variation*. Branch and Pennypacker (2013) distinguish these terms.

> *Direct replication*…refers to repeating an experiment as exactly as possible. If the results are the same or similar enough, the initial effect is said to be replicated. Direct replication, therefore, is mainly used to assess the reliability of a research finding, but as we show, there is a sense in which it also provides information about generality. *Systematic replication [or variation]* is the designation for a repetition for the experiment with something altered to see whether the effect can be observed in changed circumstances. If the results are replicated, then the generality of the finding is extended to the new circumstances. Many varieties of systematic replication exist, and it is the strategy most relevant to examining the generality of research findings. (p. 160)

As they note, "differences nearly always exist", even with direct replication. Failure to replicate pushes researchers to search for the variable(s) responsible for the differences in outcome and to identify the boundaries or limiting conditions beyond which the theory does not hold. The scientific community must judge which differences warrant further investigation. Scientific generality, they argue "is not characterized in terms of breadth of applicability, but instead in terms of identification of factors that are required for a phenomenon to occur" (p. 167).

The importance of theory to generalization or "external validity" can be traced back to the early texts on experimental design (our GCI). Campbell and Stanley (1963, p. 17) argue, for instance, that "generalization is never fully justified logically…. [it] always turns out to involve extrapolation into a realm not represented in one's sample." For

instance, with a pre-test post-test control group design we described in the GCI section, one has demonstrated the effect:

> only for pretested groups of a specific age, intelligence, socioeconomic status, geographical region, historical moment, orientation of the stars, orientation in the magnetic field, barometric pressure, gamma radiation level, etc. (p. 17)

Of course, only some of these limiting conditions are likely to matter. The goal is to distinguish those that do from those that do not:

> by guessing at laws and checking out some of these generalizations in other equally specific but different conditions. In the course of the history of a science we learn about the "justification" of generalizing by the cumulation of our experience in generalizing. (p. 17)

Shadish, Cook, and Campbell (2002) extend this logic of "external validity", calling on researchers to "develop and test explanatory theories about the pattern of effects, causes, and meditational processes that are essential to the transfer of a causal relationship" (p. 25). They also call for exploring surface similarity between the study and the target of generalization, to attempt to rule out things that do not change the generalization, and to clarify key discriminations that limit the generalization (p. 43). They note as well that any generalization involves higher order [i.e., more abstract] concepts or constructs, including how we label persons, settings, treatments, and outcomes in any given experiment, "except for the trivial case in which the correct construct that characterizes the population is already known" (p. 348) (e.g., third grade teachers).

> One way this can be done is with programs of research on the same topic by the *same* researcher (or in the same lab). Such programs allow a more directed investigation of variables that bear on generalized causal inference. A researcher can systematically vary important variables from study to study, slowly homing in on a more refined understanding of the targets of generalization, the variables that might bound or limit generalization, and the mediators that might explain it. (Shadish et al., 2002, p. 418).

They acknowledge the supplementary value of other methodologies, including "on-site intense observations" for elucidating "the more temporal, spatial, and micromediational processes that explain an effect" which helps "identify hypotheses about possible causal

contingencies to be tested in the next round of studies" (p. 391). They also acknowledge the value of literature reviews, especially quantitative literature reviews that aggregate treatment effects (meta-analyses), a topic we will address in subsequent section.

In GCI reports, one can find relevant examples in literature reviews that show how studies build on past research, in descriptions of factors mediating cause and effect relationships built into designs, and in discussion sections where factors potentially limiting the generalization are suggested for further consideration. Greenleaf et al. (2012), for instance, investigated the impact of "professional development integrating academic literacy and biology instruction on science teachers' instructional practices and students' achievement in science and literacy" (p. 647). Within the design itself, they considered multiple outcome measures for each of the general outcomes and "differences in impacts for the following subgroups: (a) English learners and English-proficient students (test scores only), (b) females and males, (c) racial-ethnic groups, and (d) and low- and high-performing students" (p. 692) thereby examining the generalizability across these features. In their discussion they noted that "both treatment and control teachers who participated and were retained in the study exhibited high levels of experience and engagement in other professional development activities targeting science instruction, suggesting that the results generalize to classrooms with experienced science teachers" (p. 704). In their recommendations for further research, they speculated about the extent to which the findings might generalize (and be enhanced) if more teachers in a school had been permitted to participate so as to form a professional community and if the professional development had extended over a longer period of time.

With his concept of "***analytic generalization***" Yin (2010, 2014) likens the problems faced by case study researchers to those faced by researchers engaged in experimental and quasi-experimental design.

> Case studies, like experiments, are generalizable to theoretical propositions and not to populations or universes. In this sense, the case study, like the experiment, does not represent a "sample," and in doing case study research, your goal will be to expand and generalize theories (analytic generalizations) and not to extrapolate probabilities (statistical generalizations). (2014, p. 21)

He characterizes analytic generalization as involving a two-step process.

> The first involves a conceptual claim whereby investigators show how their case study findings bear upon a particular theory, theoretical construct, or theoretical (not just actual) sequence of events. The second involves applying the same theory to implicate other, similar situations where analogous events also might occur. (2010, p. 21)

The relevant theory entails one or more hypotheses that, he argued, should be stated at the outset of the case study, much as with experimental design. These hypotheses are typically based in existing literature and are proposed "at a conceptual level higher than that of the specific case." (p. 66). Again, as with GCI, he highlights the importance of considering rival hypotheses in the case at hand and showing how subsequent cases support or challenge the developing theory, thus locating the limits of the appropriate generalization.

> "Each case must be carefully selected so that it either (a) predicts similar results (a literal replication) or (b) predicts contrasting results but for anticipatable reasons (a theoretical replication)…. If the cases are in some way contradictory, the initial propositions must be revised and retested with another set of cases. (2014, p. 57)

The result of any given case study may be "(a) corroborating, modifying, rejecting, or otherwise advancing theoretical concepts that you referenced in designing your case study or (b) new concepts that arose upon the completion of your case study" (2014, p. 41). These form working hypotheses (Cronbach, 1975) that can be used to reinterpret the results of existing studies or design new studies. He notes that the generalizability of findings from a single case increases as similar results are found with other cases, whether those that already exist in the literature or are subsequently completed (2010).

A study by Halverson, Grigg, Pritchett and Thomas (2007) entitled "The New Instructional Leadership: Creating Data-Driven Instructional Systems in Schools", illustrates many of the features in analytic generalization. The researchers were interested in understanding "how local school leaders build data-driven instructional systems to systematically improve student learning" (p. 159). Citing Yin, among others, they note "The study's method lies midway between hypothesis testing and grounded theory" (p.

165). They began by sketching a theory of data driven instructional systems (DDIS) based on previous literature. They represented DDIS as a system containing six organizational functions: Data Acquisition, Data Reflection, Program Alignment, Program Design, Formative Feedback, and Test Preparation. They proceeded to purposively select four schools for study "with strong records of improving student achievement scores and reputations for using data effectively" (p. 165). They then engaged in a year-long data collection, including interviews with leaders and teachers, observation of meetings where data were used, and relevant artifacts. They used their initial theory to code these data and compare findings across schools. They used their DDIS framework to compare the practices in different schools. The data they collected show that each school had different practices in place and implemented the state's DDIS in different ways. They elaborated each of the organizational functions inductively to develop a more refined category system that was then used for comprehensively coding and comparing the data collected. Data acquisition, for instance, included processes of collection, storage, and reporting, as well as reference to the agent of acquisition, whether internal or external to the school. Comparisons across schools within category illuminated substantial variations. So, for instance, continuing with the theme of data acquisition, they noted many different types of data used at the local school level and a range of low- and high-tech data storage practices, as well as mismatches between district and school level practices. Thus their initial theory, which enabled systematic comparisons, became substantially more elaborated in terms of a more refined coding scheme that illuminated particular variations across schools within category. Their representation of their evolved theory reflected this productive tension between the particular and the general:

> We do not want to suggest that a prescriptive re-creation of these schools' DDISs will provide a universal template for the new instructional leadership. …Rather, we used the DDIS functions to draw attention to the organic interaction between artifacts and actors that seems to constitute school capacity for data-driven decision making. Capable leadership and well-designed artifacts were necessary to create DDIS functions: Artifacts without leaders are mere structures; leaders without artifacts are confined to interpersonal interaction. The organic

> development of each school's DDIS required that leaders and teachers used professional judgment to determine which artifacts to alter, which to import, and which to simply leave alone. (Halverson et al., 2007, p. 188).

More Inductive Forms Of Theorizing Emphasizing Theory Building.

One of the critiques of replication and systematic variation as strategies for supporting generalization is that they tend to treat cases (in a sample) as "sites for observing and measuring variables" (Byrne, 2009, p. 1); and it is the variables from which theory is developed and tested; the specification or particularization of the case is lost except for its position at the intersection of particular variables. However, treating cases as unique instances each "to be understood in their own terms", or as enabling at best one-to-one comparisons by readers from the target context, does little to enable community learning from multiple comparisons. "How are we to preserve the epistemic significance of the particular without losing sight of the general?" (Tsoukas, 2009, p. 286). Mjøset (2009) suggests "specification and generalization are not opposites. Specification is only possible through more general knowledge" (p. 53) because it requires comparative reference to other cases at least at the concept level. "As specifications are made in this way, the results also feedback into more general knowledge: denser and broader typologies, concepts and models of contextualized social interaction patterns" (p. 53). And so, "the use of comparison is the main alternative to subsumption under as general as possible concepts and theories in the standard conception" (p. 53).

In this subsection, we focus on logics of generalization via theory that privilege the cases at hand over a priori theory in the design and enactment of research, drawing comparisons as appropriate, in the reporting of research. Two of the oldest and most widely cited approaches to generalization-via-theory are analytic induction (Becker, 1998; Hammersley et al., 2000; Znaniecki, 1934) and grounded theory (Glaser and Strauss, 1967; Charmaz, 2014). Both are still widely cited and (purportedly) employed[56] today. We referenced both in the EthR MT section in the context of within case

[56] We have come across many references to these methods in the methodology sections of articles sometimes with little or no evidence of *how* the methods were used to generate the conclusions reported.

generalization, and here we turn to their value for theory building and testing across cases.

Following Hammersley (2004) analytic induction:

> portrays inquiry as an iterative process that goes through the following stages: initial definition of the phenomenon to be explained; examination of a few cases; formulation of a hypothesis; study of further cases to test this hypothesis; and, if the evidence negates the hypothesis, either redefinition of the phenomenon to exclude the negative case or reformulation of the hypothesis so that it can explain it. This is followed by study of further cases, and if this further evidence supports the hypothesis, inquiry can be concluded. However, if any negative case arises in the future, inquiry will have to be resumed: Once again, either the hypothesis must be reformulated or the phenomenon redefined, and further cases must be studied, and so on. (Hammersley, 2004, pp. 16-17)

He suggests that analytic induction is less frequently invoked today than other approaches, although he points readers to a use by Howard Becker (1998) who argues that analytic induction "best captures the logic of social scientific investigation" (p. 17).

Grounded theory (following Charmaz, 2006, 2014) "refers to a set of systematic inductive methods…aimed toward theory development (2006, p. 440). It uses data "to construct abstract analytic categories through an iterative process" (2014, p. 15) of data collection and analysis. "Early analytic work then directs…subsequent data collection…to answer specific analytic questions and to fill gaps in the emerging analysis ["theoretical sampling"]….[which] consist[s] of abstract concepts and their interrleationships" (2006, p. 442). Researchers often proceed through two distinct phases of coding: open coding (often line by line) to generate ideas and then focused coding to allow systematic reanalysis. A process called "constant comparison" by the originators of grounded theory entails comparing data with data, data with theoretical categories, and category with category as categories are formed and relationships among them explored. Via "theoretical sampling", researchers engage data collection "to illuminate, extend, or refine theoretical categories" (2006, p. 442) until a point of "saturation" where no new ideas are emerging.

> For a grounded theory to have explanatory power, its theoretical categories should be abstract, explicit, and integrated with other categories. Thus, grounded theorists complete the following tasks: locating the context(s) in which the category is relevant; defining each category; delineating its fundamental properties; specifying the conditions under which the category exists or changes; designating where, when, and how the category is related to other categories; and identifying the consequences of these relationships. (2006, p. 444)

Charmaz notes further that theory development may include comparisons among multiple studies. "Subsequently these comparisons can result in more abstract—and, paradoxically, general—theories", but "the generality arises from scruitinizing numerous particulars" (2014, p. 322).

Both analytic induction and grounded theory have as their goal the development of theory that subsumes cases, albeit a theory that is revised until it comprehensively addresses the cases at hand, including narrowing the relevant cases and thereby limiting the generalization. A subtly different logic foregrounds the particular and privileges ongoing elaboration and "*analytical refinement*" in theory development such that the epistemological significance of the particular is always maintained. Tsoukas (2009) draws on late Wittgenstein to argue that concepts function differently. We learn them not by following strict rules but rather "in the context of social practices in which we have been practically shown to make certain distinctions, use certain tools utter certain words under certain circumstances (pp. 287-288). [The examples of concepts he gives include diagnosis, professionalism, and service.] The particulars a concept subsumes thus reflect family resemblances: they "have no one thing in common which makes us see the same word for all, but …they are related to one another in many different ways" (p. 287). There are commonalities among them, but not always the same commonalities. So the usage is "partly bounded, partly open ended. Through usage, we necessarily *revise* concepts—'extend' their reach and redraw their blurred boundaries" (p. 288). "The particular is not subsumed into the general; it rather further specifies the general" (p. 288) and this, he argues, makes accumulation of knowledge possible. That means that each new case contributes dialogically to the theory the researcher brings to bear. As we

noted in the ethnography section, Eisenhart (2009) draws on Geertz to make a similar point.

Ercikan and Roth (2014; Roth, 2009) describe an approach to generalization they call *essentialist* generalization. Drawing on sociologist Bourdieu (1992), they describe essentialist generalization as "the result of systematic interrogation of the particular case by constituting it as a 'particular instance of the possible'…in order to extract general or invariant properties" (Bourdieu, 1992, p. 233, in Ericikan and Roth, 2014, p. 10). "In this approach, every case is taken as expressing the underlying law or laws" (p. 10). They point out that it is "crucial...not to universalize the particular case but to reveal the invariant properties that hide themselves under the appearance of the singularity of each case", which "derive from the fact that there is a *common history* underlying each case" (p. 11). They offer the example of "queues" which can be found at supermarket checkout counters, freeway on ramps, or bus stops, suggesting that the underlying structure of the queuers' work is shared in common across these particular contexts. So a researcher can posit a generalization based on one case that can be tested by examining any other concrete case. They note, however, that any generalization, like queue, will manifest itself differently in different cases, such that the contextual particular must always be taken into account.

A study by Horn (2007) illustrates an approach to cross case analysis and theory development that more closely resembles the sort of inductive theorizing and analytic refinement that privileges particulars than the Halverson et al. example above. The study entitled "Fast Kids, Slow Kids, Lazy Kids: Framing the Mismatch Problem in Teachers' Conversations", inquires into the conversational category systems teachers use to frame the 'mismatch' between students' perceived abilities and the intended school curriculum. The literature review situated the study in relevant research and pointed to the sorts of questions and methodological resources that might be engaged, but did not result in any specific theory that would be tested or used to analyze these data.

While this study was part of a larger project at two high schools in the same state, the article focused on the dialogue in two mathematics teachers' learning communities where teachers were working "together to change their mathematics classes to address inequalities in students' educational opportunities" (p. 71). Horn's analysis began with a

relatively open-ended question: "What are the conceptual resources high school mathematics teachers bring to their encounters with equity-geared reforms?" (p. 46). She looked for instances in the dialogue where encounters with reform were made visible, which she came to label "episodes of pedagogical reasoning" (EPRs) where "teachers exhibit their understanding of an issue in their practice" (p. 46). She worked inductively to develop topical categories for coding EPRs, for example, student engagement, student failure, assessment, and alignment of students and curriculum. For this article she focused on one of these, "alignment of students and curriculum" and compared "two thematically related EPRs to illustrate the differences in the reasoning resources and practices of the teachers" (p. 47) at each school. She first presents each EPR in substantial detail, quoting long segments of dialogue followed by interpretive commentary for each segment. To facilitate the comparison between EPRs across teacher communities that follows, the interpretive commentary following each segment is organized by categories of talk and intersections among the categories--kinds of kids, mathematics curriculum, teaching practices--although within sections her interpretation stays very close to the preceding text.

These comparisons led her to sketch a tentative theory of what she had observed across the cases, which she frames as a contribution to what might be subsequently learned from additional studies. Here are some illustrations of that theory:

> Equity-geared reform sparked conversations about the mismatch problem [between students and the curriculum]…. Both conversations imply a source for the mismatch problem, although the roots are found in different places…. At South High, … the students ['who won't be able to handle the logic' (p. 62)] appear to be at the root of the mismatch problem…..
>
> The East High teachers' model of students places more responsibility on the teachers for addressing potential "mismatches" between curriculum and kids. The onus is not on the students to "mature" or "get turned on"; instead, the teachers must work actively in their classrooms to find group-worthy curriculum, address status issues, and set learning agendas for individual—even "fast"— students. (pp. 71-73)

Horn comments as well about the role of senior teachers in either challenging teachers' conceptions of the practices they could enact to support students' learning or ratifying

statements of the problem as rooted in students capabilities that resulted in tracked curricular pathways. She then connects the comparisons to practices known from the larger project to be ongoing in each school. Horn argues "this study contributes to the understanding of how teacher conceptions of students are negotiated and reified in context, specifically through interactions with colleagues and experiences with school reform" (p. 37). She notes her study suggests that teachers need "a more adequate technical language to represent the complexity of student learning" (p. 75) incorporated into their models of practice, but that future work will need to examine her theory in other teachers' communities in other settings.

Mechanism-based Explanations and Configurations

A third set of logics for generalization-via-theory foregrounds the *mechanisms* through which outcomes are produced in context in complex social systems. While this approach shares multiple features in common with both of the logics described above, it is focused, in particular, on mechanism-based explanations in response to causal questions. It is relevant both to interventions designed to enact change as well as to retrospective accounts of change in social systems (Mjøset, 2009). These explanations address questions not just about *what* works to change complex systems, but about *how* (through what mechanisms in what order), where and when, with whom and under what conditions, and whether the same workings can be anticipated "elsewhere" and "elsewhen" (Byrne, 2009, 2013, p. 291; Pawson, 2006, 2013). Most of the major theorists conceptualize cases as complex systems, drawing on many of the features of complexity theory we described above (Byrne, 2009, 2013; Goeken & Borner, 2012; Hedstrom & Swedberg, 1998; Hedstrom & Ylinski, 2010; Pawson, 2006, 2013).[57]

Theorists working within this approach argue that variable based explanations of cause—those that rely on (mediated) relationships between independent and dependent variables across cases—are insufficient to enable action-orienting causal conclusions or generalizations: "at best these might suggest where to look for candidates for causal mechanisms" (Sayer, 2000, in Pawson, 2006, p. 22). In the case of interventions, the

[57] Penuel (personal communication) notes the similarities between the mechanism-based explanations we cite here and well enacted DBR.

cause will seldom be the intervention alone. "What matters is how the intervention works in relation to all existing components of the system and to other systems and their sub-systems that intersect with the system of interest" (Byrne, 2013, p.219). While many who work within the GCI tradition acknowledge the importance of addressing "how" and "why" questions with descriptive studies from other methodological traditions, what mechanism-based logics of generalization add to our methodological toolkit are strategies for generalizing these mechanism-based causal conclusions across cases.

Thus, research supporting this approach is initially case based, illuminating *configurations* of mechanisms within cases, followed by systematic comparisons across cases. We focus first on how mechanisms are conceptualized and configured within case. As Hedstrom & Swedberg (1998) note, "Explanations of most concrete social events or states require resort to several elementary mechanisms; one is not enough. Sometimes these mechanisms counteract one another, and sometimes they work together" (Hedstrom & Swedberg, 1998, p. 21). The same outcome may be generated in more than one way, so any theory needs to account not just for complex causes, but also for multiple complex causes. Further, similar mechanisms may lead to different outcomes as they interact with different features of the context. Central to these contextual features are the understandings, choices, actions and interactions of the people involved. "Human agency informed by meaning is absolutely part of the causal nexus" (Byrne, 2013, p. 221). These are shaped by their past histories and relationships within and across contexts as well as the social structures and institutions within which they live and work (Bryne, 2013, p. 221).

Further complicating the picture, mechanisms occur at different levels. Hedstrom and Swedberg (1998) argue "one should always try to establish how macro-level events or conditions affect the individual (Step 1), how the individual assimilates the impact of these macro-level events (Step 2), and how a number of individuals, through their actions and interactions, generate macro-level outcomes (Step 3)" in a macro-micro-macro configuration (pp. 21-22, citing Coleman). Pawson suggests attending to the following levels in tracing mechanisms: the individual capacities of the key actors, the interpersonal relationships supporting the intervention, the culture of the institutional setting, and the wider infrastructural system. Thus explanations focus on "complex

contingent causation" (Byrne, 2013) and entail "context, mechanism, outcome configurations (CMOCs)" (Pawson, 2006) within complex systems. "Confirurations can be understood as assemblages…. [as] complexes, not variables" (Byrne, 2013, p. 223). And there can be multiple CMOCs within a case, as long-range outcomes often entail multiple complex mechanisms associated with intermediate outcomes.

Systematic comparisons of configurations or mechanism-based explanations across cases lead to theories that illuminate commonalities as well as variations. Thus "cases can be thought of as ensembles … and those with the same causal configuration can be thought of as near neighbours. This term is extremely valuable since … when asking what works, when and where, we need to find answers that work *for cases of particular kinds*. (Byrne, 2013, p. 226, italics ours). The goal is not to justify a particular course of action across contexts but rather to illuminate the features that local decision-makers might take into account in considering how best to promote change in their own contexts.

Those engaging mechanism-based generalizations ground their work in a range of more specific theoretical perspectives that can be informed by complexity theory, differences among which are beyond the scope of this chapter. They include realist evaluation (Pawson, 2006, 2013), analytic sociology (Hedstrom and Ylinski, 2010), and contextualist research (Mjøset, 2009). The techniques on which they draw to generalize can include combinations of methods from traditions we described earlier, for example process-tracing and qualitative comparative analysis from CCS and grounded theory and analytic induction from ethnographic research.

While readers can find features relevant to this logic in the examples by Halverson and Curry—multiple mechanisms, situated in different contexts, with varying outcomes, and carefully nuanced theories to reflect them—these researchers do not theorize their work in mechanism-based terms. The most methodologically explicit examples we've found come from outside of education in international development (including generalizations built from participatory action research studies) and in practices in business and industry (see, for instance, the special issue of *Evaluation* in which Byrne, 2013, appears; Brannen, 2011; Goeken and Borner, 2012). Readers may

find particularly useful their frequent step-by-step illustrations of the ways mechanism-based generalizations-via-theory are built.

Literature Reviews/Research Syntheses

In this section, we focus on methodologies for literature reviews that synthesize learnings from "primary" empirical research studies (including research reports and sometimes the data to which the reports refer). These are typically called "research syntheses" or systematic reviews. Of course, literature reviews can have other foci beyond empirical studies: they can focus, for instance, on philosophical or theoretical perspectives, on methodologies (our chapter being a case in point), or on "transdisciplinary" concepts that may be manifest in multiple fields (e.g., boundaries, emergence), with or without attention to empirical studies. While the guidance we review may be useful to such purposes, our emphasis is on reviews of primary empirical research.

In an age where proliferation of research makes it impossible to "know" all the literature relevant to a topic or problem, reviews of existing literature are becoming increasingly important. They enable the development of theory and conclusions that draw on multiple studies, thus providing a stronger basis for generalization. Many genres of literature reviews enable integration of studies utilizing multiple methodologies. Some are intended to contribute to general knowledge about a topic; others are intended to support decision-making by policy makers and other practitioners; still others play a more critical role, using the comparisons to highlight taken-for-granted perspectives and practices, often tracing their genesis and effects; and many reviews make multiple such contributions.

There is a burgeoning body of methodological literature that provides guidelines for research syntheses reflecting a wide variety of named review genres that can, by our reading, be grouped into broader logics. For the purposes of this chapter, we illustrate a range of those broader logics. The references we cite provide far more extensive reviews of these review genres. Because many of the logics of generalization we have already described underlie research syntheses as well, this section will be relatively brief.

We begin by acknowledging review genres specific to primary studies from particular research tradition(s)—meta-analysis and meta-ethnography. While neither

alone support multi-methodological syntheses, both have been used as components of broader synthesis projects. Then we turn to review genres that have addressed the problem of working across primary studies reflecting multiple research traditions. Finally, we turn to review genres that privilege critique and innovation from differing theoretical perspectives and from the perspectives of the stakeholders involved. For each of these subsections we provide a concrete example of a review arguably relevant to data use.

Syntheses Specific to Particular Research Traditions

Meta-analysis. We begin with an approach to research synthesis that is firmly grounded in the GCI tradition. Here the focus is on the effectiveness of an intervention or type of intervention. Hence these syntheses are said to focus on questions about "what works". They have been championed by two sustained international "collaborations" of researchers--the Cochrane Collaboration, founded in 1993, which focuses on interventions relevant to health care, and the Campbell Collaboration, founded in 2000, which focuses on interventions in education, crime and justice, social welfare and international development--and by the What Works Clearinghouse (WWC), established by the US Education Sciences Reform Act of 2002 as part of the Institute of Education Sciences, which focuses on *programs, products, practices,* and *policies* in education. These prominent organizations each have extensive and explicit guidelines for conducting systematic reviews on their websites where reviews meeting these guidelines are registered, endorsed, and disseminated [http://www.campbellcollaboration.org; http://www.cochrane.org; http://ies.ed.gov/ncee/wwc]. While this genre of review is often called "meta-analysis" as we have here, some prefer to limit the use of the term to the analytic methods themselves, as Glass (1976) had initially proposed (Cooper, 2010). Additional guidance for conducting systematic reviews in education relevant to this tradition can be found in Cooper (2010) and Cooper, Hedges, and Valentine (2009).

We turn primarily to the Campbell Collaboration (C2) (2014) and their *Policies and Guidelines* to provide a brief overview of research syntheses in this tradition. The C2 focuses on "systematic reviews of research evidence on the effectiveness of social programs, policies, and practices" (p. x). Interventions addressed "range from narrow to broad", including specific named programs, generic types of programs or practices, and a

range of programs for a problem or population (pp. 7-8). Perhaps the key distinguishing feature is the *exclusion* from primary synthesis of research designs that *do not* provide quantitative estimates of the effects of interventions relative to comparative (counterfactual) conditions as in experimental and (some) quasi-experimental designs:

> "The critical feature of the research methods in this regard is the ability of the basic design to yield an unbiased estimate of the effects on the target outcomes relative to a defined counterfactual condition, that is, the *internal validity* of the research design (Shadish, Cook, & Campbell, 2002)." (Campbell Collaboration, 2014, p. 9).

C2 reviews include (a) randomized controlled trials (RCTs) wherever available; (b) non-RCTs that have "sufficient claim to internal validity" (p. 10) (e.g., with "evidence of equivalence at baseline on key variables…and/or statistical control of such characteristics" (p. 10)); and (c) if a convincing case can be made (such as no stronger evidence being available), designs such as "time-series, single case studies, or multivariate observational and econometric studies with statistical controls" (p. 10) as long as an effect estimate relevant to a counterfactual condition can be obtained and the potential for bias is made explicit. Unacceptable designs include "simple before-after intervention studies without comparison groups or appropriate counterfactual conditions [and] studies in which no counterfactual conditions are observed" (p, 11). They note that while "qualitative studies" cannot be used as the primary basis for conclusions about intervention effects, they need not be excluded entirely, as they can provide supplementary evidence about such topics as implementation, generalizability, and cost (p. 35). Effect estimates from each study are expressed in a common metric, called the "effect size" (described in our discussion of the GCI MT), which enables direct comparison of findings across studies conducted using different outcome measures. To enact the synthesis, the effect sizes are then statistically weighted and aggregated across studies and reported with an error estimate such as a confidence interval, often with attention to moderator variables that mediate or condition the effect (as illustrated below). Other key features of meta-analyses include the *a priori* specification of inclusion and exclusion criteria based on the purpose and research question, an exhaustive search for all the studies that meet those criteria, and codes for data extraction that can be reliably

applied by multiple researchers to all included studies. The methodological details should be sufficient to enable replication. C2 reviews have additional requirements including the systematic search for unpublished reports (to avoid publication bias), a search that is international in scope, a protocol (plan) that is peer reviewed before the synthesis is undertaken, inclusion and coding decisions by at least two reviewers whose results can be compared, peer review of the completed report, and an expectation that the synthesis will be updated every five years.

A systematic review by Kingston and Nash (2011) (not filed with C2 or the WWC) illustrates this research synthesis methodology relevant to our theme. Their review updated a now-classic review by Black and Wiliam (1988) on the effects of formative assessments on students' achievement. Their research questions asked first about "the average effect size of formative assessment on educational achievement," (p. 29) and then whether this average effect size is moderated by grade, content area, and specific formative assessment practice (p. 29). They selected studies for inclusion based on five criteria: (1) description of the intervention as formative or assessment for learning, (2) participants must be from an academic K-12 setting, (3) use of a control or comparison group design, (4) inclusion of appropriate statistics to calculate an effect size, and (5) publication date of 1988 [the date of the earlier review] or later. Formative assessment was broadly defined to encompass assessment where "information was gathered and used with the intent of assisting in the learning and teaching process" (p. 29). Although they located over 300 articles on formative assessment within their time frame, only 13 studies met their criteria. They then coded each article and extracted the information they needed to address their research questions. They coded for sample descriptors, research design, types of formative assessment, dependent measure descriptors, and effect size, considering inter-coder reliability. Types of formative assessment practice for which they coded were "professional development", "curriculum embedded assessment systems", "use of a computer based formative assessment system," "use of student feedback", and "other". In addition to the overall average effect size for formative assessment, effect sizes were reported in tables by content area, grade level, and type of formative assessment practice. While they reported weighted average effect size of .20 (indicating an average differences of .20 standard deviations between a

treatment and comparison group), they noted "wide variation in the type and impact" of formative assessment. They noted further that moderator analysis suggested that "formative assessment might be more effective in English language arts than in mathematics or science, with estimated effect sizes of .32, .17, and .09, respectively" and that implementation based on professional development and the use of computer-based formative assessments, appeared more effective than other approaches, "yielding mean effect size of .30 and .28, respectively" (p. 28). They speculated about the possible reasons (e.g., differences in tasks across content area), drawing on previous research for additional support. Not surprisingly, they called for research into the factors that impacted the effect of formative assessment. [Readers will find a critique of the methodology by Briggs et al.. (2012) and Kingston and Nash's (2012) response additionally instructive.]

A number of critiques have been raised about the overemphasis on meta-analyses in research syntheses, including from those inside the Campbell and Cochrane Collaborations. In particular, there are substantial concerns about the usefulness of such studies to policy makers and practitioners charged with implementing the intervention. As Chalmers (2012), one of the founders of the Cochrane Collaboration, recently noted:

> systematic reviews are essential, but not sufficient for informing decisions in health care. Needs, resources and priorities also have to be taken into account. I have sometimes been left with the impression that these complexities have been insufficiently appreciated by those unfamiliar with the realities of clinical care. (2012, also cited in White & Waddington, 2012).

Snilstveit (2012), who works with the International Initiative for Impact Evaluation (www.3ieimpact.org), makes a strong argument for the importance of supplementing meta-analyses of what works with "sound evidence of how to make it work in different contexts and environments, and with different groups of people" (p. 390). She suggests either mining the effectiveness studies for additional information relevant to these questions ("effectiveness plus") or conducting a separate linked review with additional studies ("effectiveness plus with parallel review modules"). A protocol filed with the Campbell Collaboration (Carr-Hill, Rolleston, Pherali, & Schendel, 2014) proposes two research questions, one focusing on the effects of "school based decision making", and

hewing to the quality criteria described above for causal inferences, and one focused on the "conditions and circumstances" in which school based decision making has a positive or negative impact on outcomes. For this latter question, the approved protocol included qualitative and "mixed" methods studies, as well as process evaluations of mechanisms. The US What Works Clearing House has recently allowed publication of "practice guides" where a broader range of research can be brought to bear than with their traditional intervention reports. Practice guides provide:

> recommendations that educators can use to address specific challenges in their classrooms and schools. The recommendations are based on reviews of research as well as the expertise and professional judgments of a panel of nationally recognized experts that includes both researchers and educators (What Works Clearing House, 2014, p. 2).

However, the evidence in the practice guides must still be reported and evaluated in terms of the hierarchy of designs described above. Without at least "moderate" standards of internal validity (as described above), the evidence must be described as minimal. [See the 2009 IES practice guide on "Using Student Achievement Data to Support Instructional Decision Making" (Hamilton, Halverson, Jackson, Mandinach, Supovitz, and Wayman, 2009) as an example of this hybrid genre.]

In spite of these acknowledgements that policy makers' and practitioners' needs may not be adequately met by meta-analyses alone, these influential collaboratives have yet to offer much guidance about how to develop syntheses that draw on a broader range of sources of evidence. And yet, there is a wealth of such guidance elsewhere for working across methodologies in research syntheses.

Meta-Ethnography. Before turning to more "methodologically inclusive" (Suri and Clarke, 2009) guidelines, we consider another tradition-specific approach to research synthesis: meta-ethnography (Noblit and Hare, 1988). Noblit and Hare proposed "meta-ethnography" as a term "to characterize our approach to synthesizing understanding from ethnographic accounts" (p. 10). While they acknowledge the analogy to meta-analysis, they note that "any similarity lies only in a shared interest in synthesizing empirical studies" (p. 10). They make the case that meta-ethnography "should take the form of

reciprocal translations of studies into one another" (p. 11) consistent with the goal of privileging "emic" accounts.

> Some studies may be appropriately characterized in the terms of other studies; others may not be. In this latter case, the translation reveals how different the interpretations are. These differences become a substantive part of the synthesis. Reciprocal translations of studies into one another enable holistic accounts that… are comparative, emic, and historical. (p. 26).

Unlike meta-analysis, meta-ethnography is not widely practiced as a stand-alone approach to research syntheses in education (see Malpass et al., 2009, for an extended example in health care). However, it is widely cited and its principles of "translation" are often brought to bear as part of a broader synthesis project.

Noblit and Hare offer extended examples of reciprocal translation. With an interest in the "crisis of authority" between academic goals and social order across schools, they compare two ethnographic studies by Metz (1978) and by Collins and Noblit (1978) (both in Noblit & Hare, 1988), one involving two urban junior high schools and one involving one urban high school before and after a change in principals. While the latter study was not initially focused on the crisis in authority, they show how it can be read with that question in mind. Based on their reading (and "re-reading") of the studies, they propose four categories or *sets of* concepts used for comparison: teachers, students, classroom order, and effects. For each category, they list the specific language ("metaphors") used by the authors in describing phenomena in that category. Classroom order, for instance, was characterized in terms like "accepting opposition from no one", "own responsibility…first", for Metz's first school, and "academic", "delegating", "mediated" for Metz's second school. Noblit and Hare demonstrated that these concepts from Metz's study could be translated in terms of the concepts of "bureaucratic order" and "negotiated order" from the Collins and Noblit study. They consider as well whether the four sets of concepts were sufficient to summarize the case presented about the crisis in authority in each study.

> An adequate translation maintains the central metaphors and/or concepts of each account in their relation to other key metaphors or concepts in that account. (p. 28).

While translations reflect one level of synthesis, they "can be compared with one another to determine if there are types of translations or if some…concepts are able to encompass those of other accounts", leading to a "second level" of synthesis. They also distinguish "refutational synthesis" from "reciprocal synthesis": while they both involve translation, refutational synthesis highlights studies that contradict one another. We see examples of such translations in Horn (2007) reported above. Noblit and Hare acknowledge and endorse a complementary form of synthesis they call "line-of-argument".

> A lines-of-argument synthesis is essentially about inference: What can we say of the whole (organization, culture, etc.), based on selective studies of the parts? This is the same as basic theorizing in qualitative research and is conceptualized alternatively as clinical inference and grounded theorizing. (p. 63).

With "line-of-argument", Noblit and Hare endorse a configurational approach to synthesis (similar to the mechanism based explanations described above) and point us in the direction of more methodologically inclusive genres of research synthesis.

Syntheses Encompassing Multiple Research Traditions

In the past two decades, there has been a proliferation of texts supporting readers in developing research syntheses across differing types of primary research. While some of the texts we've found are from health-related professions (Pope, Mays, and Popay, 2007; Sandelowski and Barroso, 2007), there is also attention to multi-methodological research syntheses in the social sciences, including education, largely from outside the US (Gough, Oliver, and Thomas, 2012; Pawson, 2006; Suri, 2013). Gough and colleagues (2012) and Pawson (2006) work within collaborations that maintain websites similar in level of methodological resources to those cited above in the meta-analysis section (http://eppi.ioe.ac.uk/ and www.ramesesproject.org respectively). Like the generalizing logics we've presented above, the logics supporting methodologically inclusive syntheses range from those that privilege the general and seek commonalities, to those that privilege the particular and seek to elaborate complex understandings of differences, to those that consider configurations of mechanisms.

Most of the guidelines we describe below (like the guidelines for meta-analysis) refer to similar aspects of a process of producing a research synthesis that include:

- developing the purpose and research question;

- searching for relevant studies and deciding when to stop searching;
- developing criteria for inclusion and exclusion of studies, including criteria of quality;
- analyzing each of the included studies and synthesizing them based on the analyses;
- developing and evaluating conclusions; and
- reporting and disseminating reviews.

They recommend that researchers be explicit as possible about their methodology. Each aspect has validity issues associated with it. Unlike meta-analysis, they often treat the aspects of this process iteratively (rather than linearly) and they allow for contributions from a far broader range of methodological traditions.

The methodologically inclusive guidelines we review in this section differ in terms of how the aspects of the process outlined above are carried out, including: what role various stakeholders in addition to the researchers might play at different stages of the process; the extent to which the purpose is simply summary and integration or integration plus innovation or critique from one or more theoretical/philosophical perspectives; the extent to which the search is intended to be exhaustive or purposefully selective; and the nature of analysis and synthesis and the ways in which their validity is conceptualized.

Our focus will be primarily on analysis and synthesis. Analysis refers to coding, data extraction, summarizing, translating or otherwise interpreting each primary study. Synthesis refers to the process of bringing the studies together, based on the analyses, to create a whole that is more than the sum of the parts (Gough, Oliver, and Thomas, 2012; Noblit and Hare, 1988). Syntheses are frequently distinguished in terms of the extent to which they are aggregative or configurative, albeit most involve some of both:

> ***Aggregation*** entails the assimilation of findings considered to address the same relationship or connection between two or more aspects of a target phenomenon [as in potential cause and effect relationships]….. Pooling the findings gives an indication of where the preponderance of evidence lies. (Sandelowski, 2007, p. 323)

> ***Configuration*** entails the arrangement of thematically diverse individual findings, or sets of aggregated findings, into a coherent theoretical rendering of them. In contrast to the judgment of thematic similarity among findings required to aggregate findings, findings in configuration syntheses are conceived as thematically diverse and therefore as not amenable to pooling. Instead of confirming each other (by virtue of repetition of what are judged to be the same aspects or associations), thematically diverse findings may contradict, extend, explain, or otherwise modify each other. (Sandelowski, 2007, p. 325).

Texts by Gough et al. (2012) in social science and by Pope, Mays, and Popay (2007) in health provide relatively comprehensive overviews of research syntheses that include logics of generalization we reviewed above along with a range of additional logics for synthesis, not all of which entail causal questions. Quality criteria are typically based on the guiding principles of the research tradition in which a given study was conducted; and the sources we cite point to the multiple examples of coding sheets used to evaluate the quality of articles from a range of methodological traditions.

With respect to analysis and synthesis, these approaches all rely, in one way or another, on the development of categories and comparisons within and across categories. Thomas, Harden, and Newman (2012, a chapter in Gough et al., 2012) show how syntheses can differ "in terms of when in the process the distinguishing categories originate; whether they are determined at the outset of the review as part of its conceptual framework ('deductive'), derived from the studies themselves ('inductive'), or a combination of the two." (p. 183). They note further, whether inductive, deductive, or both, categories have much the same purpose in a synthesis: "they are the mechanisms by which studies are compared, contrasted and sub-divided; and their findings are configured." (p. 184). [58]

The mapping produced from a scoping review (see our description of scoping reviews in our introduction) is often considered a first step toward subsequent synthesis activities; scoping reviews describe and organize the studies but typically stop short of

[58] While we have emphasized more 'qualitative' approaches to the synthesis of methodologically inclusive reviews, Pope et al. describe a range of quantitative approaches that can be incorporated, including but far from limited to the sort of meta-analysis associated with effects of interventions. Thomas et al. (2014) illustrate how Qualitative Comparative Analysis (QCA) can be used within research syntheses.

presenting the findings. Thomas et al. (2012) go on to describe three additional approaches to synthesis (that include findings) that they call "thematic summary", "framework synthesis", and "thematic synthesis" (referencing logics that may not be described in the same terms by other authors). Thematic summary and framework synthesis differ primarily in terms of whether the categories are more deductively (thematic summary) or inductively (framework synthesis) derived (consistent with the more inductive and deductive logics of generalization we described above) and the level of integration provided. With "thematic summary", researchers

> organize (configure) the studies into groups according to a dimension (e.g., research method or socio-economic group) that would be salient to their intended readership. This involves a detailed assessment of the characteristics of included studies (participants, intervention and programme theory—where applicable—outcomes, etc.) in order to establish which group(s) a study belongs to, followed by a presentation of their results according to these (usually prespecified) subdivisions…. [They] consider any differences between studies in each group, whether they have similar or divergent findings, and finally the conclusions that should be drawn to answer the reviews questions. (189-190).[59]

With *framework synthesis*, any initial conceptual framework evolves during the study as reviewers become more familiar with the literature. This entails an iterative, multi-stage process beginning with open ended coding and the search for patterns, leading to an evolved conceptual framework, and systematic re-coding (analogous to grounded theory). "Each evolution of the framework will require another iteration of the coding process" (p. 192).

> Once the framework is satisfactorily coherent, and coding is complete, each element of the framework can be tackled in turn, tabulating data under key themes in order to present distilled summaries.
>
> The last stage is drawing together what can be learnt from the … summaries. This is done by creating typologies, mapping the multidimensional

[59] Pope et al. (2007) note that themes can be counted and tabulated, although they raise cautious about vote counting if used to determine the effects of interventions.

> nature and diversity of phenomena, finding associations between themes and providing explanations for the findings across the included studies. (p. 192).

Thematic synthesis incorporates explicit attention to "translation" in its staged approach to synthesis. "This activity is primarily concerned with translating the findings of studies into a common 'metric' so that it is possible to compare and contrast them" (p. 194), similar to the description of meta-ethnography above.

A *configurative*, *mechanism based* approach to synthesis is described by Snilstveit (2012) and, as we noted above, is explicitly intended to complement impact-oriented reviews. Snilstveit's approach relies on what she describes as a program theory:.

> A program theory starts with the intervention inputs and maps out all the links in the causal chain, through outputs, intermediate outcomes and final outcomes …. It lays out how the intervention is expected to lead to planned outcomes and make the assumptions underlying the intervention explicit. Mapping out the program theory can act as a 'scaffolding of the study' and assist in determining which key assumptions in the program theory needs testing and data collection. (p. 394)

Snilstveit notes that program theories have many uses in systematic reviews from planning the review, to developing analyses, to interpreting the findings. They argue that "different types of evidence are appropriate for answering different types of questions…and can be useful in reviews …related to complex social issues and interventions" (p. 394). A book by Pawson (2006) offers explicit methodological advice for conducting such reviews along with a set of well-explained examples. The review by Marsh (2014, described below) illustrates (per our reading) these sorts of configurational mechanism-based strategies.

While these characterizations are relatively broad, we note further that all of the theory-based logics of generalization we described above are as relevant to research syntheses as they are to the design and enactment of primary research studies. Taken together these strategies support both the more aggregative and configurative (Sandelowski, 2007) approaches to synthesis we cited above, allowing comparison of findings addressing similar questions and structuring/arranging of findings that reflect different elements of a theory into a coherent theoretical narrative.

Some synthesis theorists support active and on-going consideration of how stakeholders may be involved in the synthesis process (Gough et al., 2012; Suri, 2013). Gough and colleagues (2012), for instance, suggest stakeholders' engagement with researchers may be one of stakeholder control (e.g., sponsoring or funding a review), collaboration, consultation, or minimal; and researchers may work with organized stakeholders, individual stakeholders, stakeholders who initiate queries, or none at all; and this involvement may be active at any stage of the synthesis process from the development of questions through the interpretation of findings.

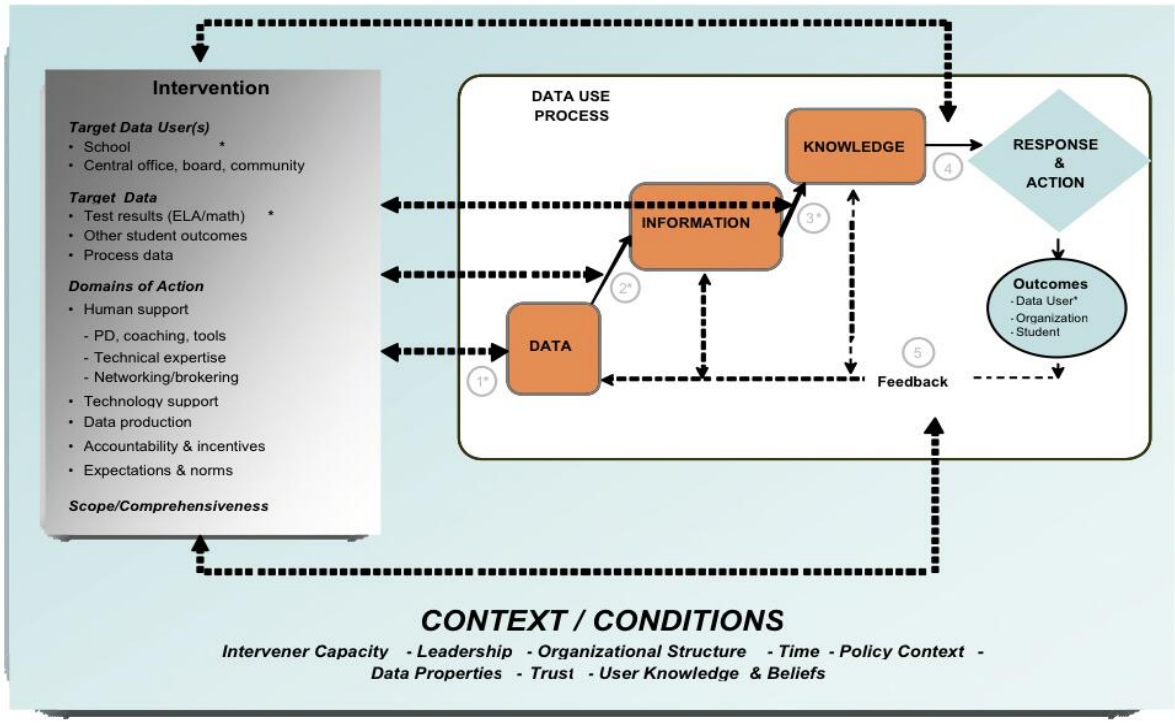
A research synthesis by Marsh (2012), "Interventions Promoting Educators' Use of Data: Research Insights and Gaps" highlights a number of the strategies addressed in this section regarding research syntheses drawing on multiple methodologies and theory based approaches to generalization. The review reflects elements of both aggregation of findings within themes and configuration of findings across themes; it reflect elements of both the *more deductive and inductive forms of theorizing* we described in the earlier section on logics of generalization (which parallel Gough et al.'s, 2012 *framework and theoretical syntheses*), and it is partially *mechanism based*, describing mechanisms at work across different phases of their data use theory. Marsh describes her objective as follows:

> This article synthesizes what we currently know about interventions to support educators' use of data—ranging from comprehensive, system level initiatives, such as reforms sponsored by districts or intermediary organizations, to more narrowly focused interventions, such as a workshop. The article summarizes what is known across studies about the design and implementation of these interventions, their effects at the individual and organizational levels, and the conditions shown to affect implementation and outcomes. (Marsh, 2012, p. 1).

Consistent with general procedures described above, Marsh conducted a systematic literature search with explicit inclusion criteria, which she narrated in the methods section of her article. She notes that "to screen for a minimum level of quality, documents come from either a peer-reviewed journal, book, or independent research institution that explicitly submits reports through a review process" (p. 5). It's also important to note that she routinely describes the quality of the available evidence as she

reports her findings.

She frames her review in terms of a "data use theory of action" that she states is "promoted by data advocates and adapted from the literature (consistent with the more deductive forms of theorizing we described above and with Gough et al.'s thematic summary). This "mapping the terrain of data support interventions" is depicted in her flowchart below.



Marsh (2012) p. 36

She uses this preliminary theory to organize her initial analysis and ultimately the presentation of her findings. The review is differentially aggregative and configurative, as appropriate to different sections. As part of this work, she engaged in a number of categorization activities, some deduced from her theory but some induced from the literature. For instance, in the "interventions" box on the left side of the flowchart, she lists the categories she induced for the different types or components of interventions, each of which reflects potential mechanisms for change. She illustrates each with more

specific examples from the individual studies to justify the meaningfulness of the categories. She engages in similar practices to develop and illustrate categories for contexts and conditions influencing data use interventions and for different types of outcomes, including effects on data users, on the organization, and on students.

In a separate section, she summarizes a set of themes or propositions regarding implementation of interventions that can be induced from the perspectives of researchers and stakeholders reflected in the literature: "these themes explore what happens when well-intentioned individuals and organizations try to enact interventions and examine the process of helping educators gather, organize interpret and action on data" (p. 11). For instance, propositions related to promising practices and features include: "making data usable and safe are important preconditions for use" (p. 12), "interventions may have greater traction when they are comprehensive and target multiple leverage points" (p. 14), and "collaboration …appears to be an important component of successfully implemented interventions" (p 15). These analyses/syntheses seem to parallel a process of analytic induction and acknowledge the degree of support and the variations underlying each proposition.

She concludes that the research base is limited in both quantity and quality and calls for greater attention "to the organizational and student-level outcomes of interventions, comparative analyses, interventions that help educators move from knowledge to action, and specific ways in which the quality of data and leadership practices shape the effectiveness of interventions" (Marsh, 2012, p. 2).

This review by Marsh was part of a series of literature reviews focused on data use that grew out interdisciplinary dialogues sponsored by the Spencer Foundation and led by Cynthia Coburn and Andrea Buschel. Each review focused on a different phenomenon relevant to data use. Taken together, these reviews sketch a multi-methodological research agenda that focuses on different aspects of the problem of data use: data-informed interactions and norms in professional learning communities (Little, 2012); teachers' information needs and the sorts of tests that might serve them (Supovitz, 2012); professional community norms and school-level routines intended to support data use (Spillane, 2012); the ways social networks support and constrain the interpretation and use of data (Daly, 2012); the sorts of supports districts provide and the ways schools

and districts interact in serving one another's information needs (Honig & Venkateswaran, 2012); how practice is shaped by "performance metrics" (like indicators of Adequate Yearly Progress required by the No Child Left Behind Act of 2001), the incentives and sanctions associated with accountability programs, or the creation of safe spaces for learning from data (Colyvas, 2012; Jennings, 2012; Marsh, 2012); how large and small scale interventions and other points of leverage intended to increase data use capacity impact professional, organizational, and student learning (Coburn & Turner, 2011; Marsh, 2012); and how politics shape data use policy (Henig, 2012). While each literature review—and the studies synthesized—focuses on a particular phenomenon relevant to data use, each also points to the ways in which the phenomenon on which they focus relates to (interacts with) other parts of the system, and thus reading across the reviews brings a productive map of the phenomenon into relief consistent complex systems theory as an approach to methodological pluralism.

Syntheses that Entail Explicit Comparison of Methodological Perspectives, Critique, and Innovation

A number of theorists highlight the potential of research syntheses for critique and innovation, especially if researchers are willing to ferret out their own preconceptions about research and practice and to seek resources outside the boundaries of their fields. Suri and Clarke (2009; Suri, 2012, 2013) call for attention to diversity in both review methodology and primary research reviewed. They highlight the implications for research syntheses of perspectives they label "interpretive", "participatory" and "critical" alongside the more conventional forms of research synthesis which they label "post-positivist" where they cite key scholars whose work can be located in those perspectives.

- An interpretive synthesis may be evaluated by its potential to do the following: stimulate healthy debate rather than achieve closure (Eisenhart, 1998); increase our understanding of a research domain; illuminate implications of contextual, theoretical, and methodological positioning of individual studies on prevalant understandings of a phenomenon; generate or expand theory; and provide a new way of understanding the phenomenon. . . .
- Participatory syntheses should "connect the local and the global" by transforming "both practitioners' theories and their practices and the theories and practices of

> others whose perspectives and practices may help to shape the conditions of life and work in particular local settings" (Kemmis & McTaggart, 2000, p. 598, emphasis in original). The value and validity of a participatory synthesis lies in providing "effective support for the stakeholders' actions, organizations, and/or communities in their processes of self-determining social change" (Greenwood & Levin, 2000, p. 94). . . .
>
> • Critically oriented synthesists should be conscious of how "reviews contain silences" and "expressions" that privilege some discourses over others, thus becoming "the bearer of truth and power effects."…. 267). Critical syntheses should be "tempered by a sense of openness" by constantly watching "for the interruptive, counter-hegemonic capacity and absent-presence of what James Scott (1990) calls 'hidden transcripts'" (Livingston, 1999, pp. 15–16). (Suri & Clarke, 2009, pp. 410-411, including their quoted citations; bullets ours.)

They call for synthesists "to actively take into consideration the varied interests of different stakeholders in the field" including their own and "to be explicit about where they are coming from and how their positioning [the methodologies they privilege] may influence the synthesis" (p. 409). "It is crucial", they argue, "to not discount alternative methodological paths without a justified reason" (p. 409). Citing King and Gordon (2006), they note: "all reviews are situated in their own frameworks and, for a review to be complete, this framework must be acknowledged and critiqued (in Suri and Clark, p. 409).

A "meta-narrative" research synthesis by Greenhalgh and colleagues (2005), cited by Suri (2013), provides one instructive example of what it can mean to take epistemological commitments into account. The topic of their review is broad, "diffusion of innovation"; and although they write from within the health related professions, they made a particular point of wanting to understand how diffusion of innovation is conceptualized and studied from a range of different research traditions. "The challenge is to expose the tensions, map the diversity and communicate the complexity of how the various different traditions contribute to an understanding of the problem as a whole" (Greenhalgh et al., p. 427). By searching broadly, they hoped not to miss studies from other sectors "which might prove the best source of original ideas for our review" (p

418). They considered their unit of analysis to be "the unfolding 'storyline' of a research tradition over time" (p. 417). Their initial literature search, intentionally exploratory and emergent, resulted in 13 different research traditions including, marketing, evidence based medicine, cultural organizational studies, organizational network studies, knowledge utilization, narrative studies, and complexity studies. They noted:

> "Researchers in different traditions had conceptualised, explained and investigated diffusion of innovations differently and had used different criteria for judging the quality of empirical work. Moreover, they told very different over-arching stories of the progress of their research." (p. 417)

From their "initial 'territory mapping' exercise", they noted a highly inconsistent approach to research design (which is somewhat ironic given the emphasis in the US educational context on RCTs):

> For example, whilst many papers from the social sciences took it as given that studies evaluating the implementation of a programme should be based on in-depth qualitative enquiry, a widely cited systematic review on the dissemination and implementation of evidence-based guidelines had dismissed as 'flawed' any primary study that was not a randomised controlled trial (RCT) with strict predefined outcome measure. (p. 418)

They cautioned members of their synthesis team not to be drawn into privileging the methodology of the research tradition they knew and dismiss others as 'methodologically flawed'. They worked to interpret and evaluate each study within "the quality criteria set by experts within that tradition" (p. 421). The synthesis began by interpreting each study in its own terms. They then developed a set of categories that would enable more focused comparison: (1) innovations, (2) adopters and adoption, (3) communication and influence, (5), the inner [organizational] context, (5) the outer [environmental] context, (6) the dissemination process and (7) the implementation process. Where contradictions were found, they tried to explain them, in part by acknowledging cross tradition differences. The following example is instructive:

> For example, empirical research within EBM [Evidence-Based Medicine] has shown that opinion leaders have little impact on the adoption of innovations…, but research from organisational sociology has produced the opposite

> conclusion—that in any particular social or professional group, opinion leaders often have profound and far reaching impact…. By nesting each primary study within its paradigm, we could systematically explore these differences in terms of how 'opinion leadership' had been differently conceptualised and explored. In this example, a key explanatory variable was the degree of agency of the opinion leader. We concluded that (on the one hand) certain people have considerable social influence on their peers, and this influence can sometimes be the making or breaking of a complex intervention programme, but that (on the other hand) individuals identified by experimenters as opinion leaders cannot be injected into a complex social situation and reliably manipulated to influence a predefined outcome. (p. 423)

A book length representation of their review can be found in Greenhalgh et al. (2005) and principles and training materials can be found at Ramesesproject.org. We should note that some theorists who share an interest in reflexive attention to epistemological perspective might question whether differences within and across tradition had been adequately represented, and whether the anticipation of single voiced explanations might undermine their goal. Nevertheless these synthesists present a viable, well theorized, and illustrated approach to embracing the complexity of learnings from different research traditions, including the perspectives and participation of key stakeholders throughout the project.

A number of literature reviews we encountered on the topic of data use took steps in the direction outlined in this section. For instance: Opfer and Peder (2011) drew on complexity theory to organize and criticize the body of literature on teacher learning; Colyvas (2012) drew on a range of literatures on performance metrics in multiple contexts beyond K-12 education to suggest ways of evaluating the practice and understanding its impact on the social contexts; Henig (2012) reviewed theories of political science "to inform our understanding of how politics affects efforts … to increase data usage" (p. 1).

Generalizing as Learning in Local Contexts

In the previous two subsections, we have focused on the development of generalizations where the goal is to make (mostly) researcher-warranted contributions to general knowledge in reported research. In this closing section, we turn briefly from these more static, product-oriented, conceptions of generalization, synthesis and accumulation of knowledge to consider more dynamic understandings of know*ing* and generaliz*ing* that foreground the evolving relationship between research and practice—between general knowledge and local decision-making and action. How can engaging with research inform the practice of teachers, administrators, policy makers, and researchers, as well as the schools and other organizations in which they work? This includes how more general knowledge reflected in research reports is put to work, but also how knowledge is produced in local contexts, and how these local forms of knowledge can contribute to more general knowledge. From this perspective, knowledge or knowing "becomes the capacity to act" (Jensen et al., 2012, p. 2).

Organizational theorists Cook and Brown (1999) draw an instructive distinction between the "epistemology of possession" and the "epistemology of practice". "Knowledge is about possession…. In all its forms we use it to indicate something an individual or group possesses, can possess, or needs to possess" (p. 388). "Knowing is about interaction between the knowers and the world" (p. 389); it is "dynamic, concrete, and relational" (p. 387). "We must see knowledge as a tool at the service of knowing, not as something that, once possessed, is all that is needed to enable action or practice" (p. 388). From this perspective, knowing and generalizing can be usefully conceptualized as learning and "the resources for learning lie not simply in information, but in the practice that allows people to make sense of and use that information" (Brown & Duguid, 2000, p. 133).

We focus on an eclectic and admittedly ad hoc collection of methodologically relevant perspectives that theorize the dynamic relationships between research and practice with questions of validity and usefulness in mind. The research discourses on which we draw include (professional) learning (Billett et al., 2014; Schwartz et al., 2005; Jensen et al., 2012; Nathan and Sawyer, 2014), organizational studies (Brown and Duguid, 2000, Cook and Brown, 1999; Penuel & Spillane, 2014), policy studies

(Coburn and Stein, 2010; George and Bennett, 2005; Byrne, 2013), social epistemology (Biesta, 2007; Knorr Cetina, 1999, 2006), and design based implementation research (Fishman, Penuel, Allen, & Cheng, 2014) with its ties to both professional and organizational learning. Of course, the studies we've used as examples relevant to data and research use contribute empirical evidence that can inform theories of the dynamic relationship between research and practice. While the range of research discourses is eclectic, and we've only touched on scholarship that might inform an understanding of generalizing as learning, the scholars we cite address a number of common themes.

Perhaps the overarching message is the importance of *human capacity* to produce and use knowledge, individually and collectively, in circumstances always unique in some respects, and the need for conceptual and organizational *infrastructures* that support it. This requires attention to the differing practices and information needs of professionals working in different contexts. As Brown and Duguid (2000) argue, "The same stream of information directed at different people doesn't produce the same knowledge in each. If the people are engaged in different practices, if they are learning to be different kinds of people, then they will respond to the information in different ways" (p. 129). The literature we cite points readers to studies that foreground the practice of professionals in different roles and the ways in which they might work together across boundaries to improve teaching and learning (Coburn & Talbert, 2006; Penuel , Allen, Coburn, & Farrell, 2015). Given the focus of this volume, we privilege literature on the practice of teachers and of those who make policy decisions that impact teachers' work.

We begin with arguments about the appropriate status of general knowledge in local contexts and the ways in which general knowledge might serve practice; then we turn to what this implies about the nature of professional practice (individually and collectively); and finally to the sorts of organizational resources that might support it. We acknowledge the literature we cite provides only a glance towards an extensive and interdisciplinary research base that might support a robust relationship between research practice.

Status of General Knowledge in Local Contexts

As Coburn and Stein (2010) observed, debates about the relationship between research and practice have frequently assumed a linear relationship with the assumption

of "research being conducted and subsequently disseminated to practitioners who apply the findings" (p. xii):

> In its most unembellished form…, this linear model views basic research as leading to applied research, which then leads to the development of products or codified professional practices that are then disseminated to educational practitioners and systems (p. 4).

They note that this assumption has a continued prominence in the 21st century, at least in the US context. By way of example, they cite the phases of research articulated in the Institute of Education Science Requests for Proposals beginning in 2002.

> Funding is organized by phases of research for the improvement of practice, beginning with basic research studies that are intended to uncover new principles of learning, followed by studies that aim to design interventions based on those principles, and leading, finally, to larger-scale studies of the effectiveness of those interventions. (p. 4)

The problem with this linear assumption, Coburn and Stein note, is that "existing research on the relationships between research and practice suggests that, even in successful instances where research influences practice, things work quite differently" (p. 5). The case studies they report "reveal variable, complex and shifting roles of researchers, practitioners, and intermediaries and the multiple pathways through which researchers and practitioners can interact productively" (p. xii). Coburn and Stein argue that "The underlying conceptualization of the relationship between research and practice is important because it influences the way we view the problem and the policy solutions that we offer" (p. 4). Assumptions of a
linear relation have led to policy solutions "that focus mostly on bolstering the quality of basic and applied research, translating research findings into forms usable in practice, and disseminating research-certified programs and practices to practitioners" (pp. 4-5). Research, they worry, "may miss its mark absent attention to the conditions for ongoing professional learning and decision making in public schools" (p. 12).

Turning to the policy context, the US National Research Council's (2012) Committee on "Using Science as Evidence in Public Policy" noted similar limitations in a "two communities" metaphor regarding the relationship between research and practice.

They concluded there "has not been much success in explaining the use of science in public policy" (p. 2): "the research specialty labeled "knowledge utilization" has focused on challenges highlighted by the "two communities" metaphor (researchers and policy makers, each with their distinctive cultures) and proposed various innovations to improve communication and interaction between science and policy—brokering, translation, interaction models." (p. 3). But the relation between policy and science is far more complicated:

> Policy is made in many settings. It evolves from a many faceted social process involving multiple actors engaged in assembling, interpreting, and debating what evidence is relevant to the policy choice at hand, and then, perhaps, using that evidence to claim that a particular policy choice is better than its alternatives. This process is best understood as a form of policy argument or practical reasoning that is persuasive with respect to the benefit or harm of policy actions. Policy argument includes generalizations, extrapolations, assumptions, analogies, metaphors, anecdotes, and other elements of reasoning that differ from and can contradict scientific reasons. From this perspective, scientific knowledge is "evidence" when that knowledge is used in support of statements relevant to policy claims. "Evidence" does not reside only in the world where science is produced; it emerges in the political world of policy making, where it is interpreted, made sense of and is used, perhaps persuasively, in policy arguments. (p. 4).

They note further that while the "approach known as evidence-based policy and practice, focused on improving understanding of "what works," has influenced the production of scientific knowledge, it has made little contribution to understanding the use of that knowledge. Inspired by the seminal work of Weiss (1978 in NRC 2012), they argued that social scientists could not just focus on how to increase the use of research in decision making, but should also consider "to what extent, in what ways, and under what conditions, can social research help" decision makers in making wiser decisions (Weiss,

1978, pp. 26, 78, in NRC, 2012, p. 54).[60] They proposed a framework for an interdisciplinary program of research to provide a more satisfactory explanation and guide for the relationship between research and practice, which draws simultaneously on studies of policy argumentation, psychology of decision-making, and systems perspectives.

What is/should be the status of general knowledge produced in others times and/or places in a given local context? What does this imply about how knowledge can support change? The scholars on whose work we've drawn provide an alternative conceptualization that acknowledges the complexity and the openness of the systems in which teachers and other education professionals work.

As Biesta (2007) argues, "Professionals need to make judgments about 'the most appropriate course of action in the specific circumstances in a context of informal rules, heuristics, norms and values…. [Quoting Sanderson, he continues] 'the question for teachers is not simply 'what is effective' but rather….'what is appropriate for these children under these circumstances?'" (Biesta, 2007, pp. 10-11). This means, he argues, "that inquiry and research can only tell us what is possible—or, to be even more precise, they can only show us what *has been* possible. Research, in short, can tell us *what worked* but cannot tell us *what works*. (16). George and Bennett (2005), writing in the context of public policy, make a similar argument: "The policy maker has to convert the abstract model into a specific strategy that fits a particular situation" (p. 269) or diagnosis of a problem. While good general knowledge can increase the practitioner's chances of being right, "No theory or systematic generic knowledge can provide [practitioners] with detailed, high-confidence prescriptions for action in each contingency that arises" (p. 272) . Given that, "Theory and generic knowledge are best understood as a source of inputs to policy analysis of specific problems….They are an aid, not a substitute …for judgments that decision makers make" (p. 276). They note further that generic knowledge may be most useful when it moves beyond answers to simple cause and effect questions to identify "conditions, processes, and causal mechanisms that link the use of each strategy to variance it its outcomes" (p. 272). Similar circumstances face teachers

[60] Gutiérrez and Penuel (2014) and Moss (2014) note that IES is beginning to signal an interest in contextualizing questions like this through some grant programs and in statements by its advisory board (NBES, 2012) and its recent director (Easton, 2010).

on a daily basis. Moss (2008) summarizes teacher educator Magdalene Lampert's (2001) description of her experience solving the problems entailed in teaching mathematics to fifth graders:

> Lampert routinely attends to evidence with a particular problem or question in mind about what to do next. Thus, evidence is always considered in light of the question or issue it is addressing: what problems to use in tomorrow's lesson, which mathematical concepts to illustrate with a given problem, what students' likely approaches to a problem will be, which student to call on, how to respond to students' misunderstandings, how to support students' willingness to "study" and question their own solutions, how to support students in being a resource for other students, how to communicate progress to students and their parents, and so on…. Her attention to evidence in solving a teaching "problem" was routinely cumulative in the sense of drawing on other available evidence … and anticipatory in the sense of considering how her next move was likely to affect students' learning. (Moss, 2008, pp. 251-252, citing Lampert, 2001)

Biesta (2007) draws on Dewey to describe the kind of epistemology most appropriate for the knowedge-in-action that professional practice entails:

> Knowledge acquired in previous situations — or knowledge acquired by others in different inquiry or research situations — does not enter the process of reflective problem solving in the form of a rule or prescription….. We use "old" knowledge to guide us first in our attempts to understand what the problem might be and then in the intelligent selection of possible lines of action. What "old" knowledge does, in other words, is help us approach problem solving more intelligently. Yet, the proof of the pudding always lies in the action that follows. (p. 17)

Biesta also highlights the importance of questioning ends as well as means and questioning the ways problems are framed. Given these arguments, a number of theorists who share this perspective prefer to replace the term "evidence based practice" with a term like "evidence informed practice" (e.g., Biesta, 2007; NRC, 2012; Nutley, Walter, & Davis, 2006) to signal the appropriate role of general knowledge in professional practice in complex educational systems.

Professional Practice and Knowledge Cultures

These arguments highlight the importance of questions about the cultures in which teachers and other professionals work and the ways in which they enable (or constrain) the kind of professional practice and knowledge-in-action described above. In addressing this issue, Jenson and colleagues (2012) draw on sociologist of science Knorr Cetina's conception of "epistemic culture" (Knorr Cetina, 1999, 2006) which constitutes the "practices in play when knowledge is created, distributed and validated; the products of these processes; and the specific modes of collective reasoning that constitute these practices and guide practitioners' engagement with knowledge" (p. 8). Epistemic cultures have both social-organizational and cultural-cognitive features. At the social/organizational level, there are "a multiplicity of material and symbolic resources …that invites and structures participation and engagement" (pp. 11-12), including those within and beyond the local work environment. "They offer identities and positions from which to act, and they stimulate …[and] regulate…self conduct" (p. 11). At the cultural-cognitive level, there are interpretive frameworks that guide how people reason about and justify knowledge and serve as a basis …for taking on certain responsibilities while …rejecting or delegating others" (p. 12). Epistemic cultures entail "beliefs about knowledge itself; for example about 'the correct distribution of knowledge, the naturalness of access to it, and the particular ways knowledge should be handled and inserted into personal and organizational life' (Knorr Cetina, 2006, p. 37)" (Jensen et al., 2012, pp. 11-12). To better illuminate the usefulness of these concepts, we sketch some of the findings from Jensen and colleagues' (2012) study of the epistemic cultures of four professions in the Norwegian context: accounting, engineering, nursing, and teaching. Their multi-method data collection involved analyses of professional curricula, comparative case studies of how the relevant professional associations organize knowledge, and longitudinal studies of 10 practitioners from each of the four professions involving interviews, focus groups, and work logs. To give a flavor of the sorts of conclusions their research enables, we focus on the contrasts they drew between the epistemic cultures of the teachers and nurses they studied in Norway:

> Nurses are well supported by an extensive knowledge infrastructure, and professional updating is an integrated and regulatory aspect of their profession.

> They use multiple knowledge sources that have been specifically designed for the profession, including manuals, intranet, reference works and textbooks; receive systematic supervision from colleagues; and have access to a wide range of specialised expertise. Knowledge resources are locally adapted and appear to have strong "translation mechanisms" that mediate between globalised, abstract knowledge and the local realities of the hospital wards…. The teaching profession distinguishes itself by being more locally confined, and practitioners in this field describe a limited use of professional knowledge resources from outside the local community. When the teachers refer to research within their own profession or to extended circuits of knowledge, this contact has a more sporadic character.
>
> These differences also manifest themselves as distinct epistementalities [cultural-cognitive dimensions] in the four professions. While teachers emphasize face-to-face sharing of personal knowledge in local communities and claim a freedom to choose methodological approaches on an individual basis, the nurses access collective resources and show concern for ensuring that they are familiar with research-based advancements and collective procedures for good practice. (pp. 17-18)

Jensen et al. go on to describe the ways in which these differences reflect the knowledge strategies of the professional organizations, with the nurses association having established common knowledge resources, and the teaching association promoting the distribution of experiences and reflections of individual teachers. Looking across the four professions, they conclude that "communities that manage to link local practices with extended networks and circuits of knowledge in a systematic way are more likely to develop an epistementality geared towards learning and continuous exploration" (p. 19). These findings highlight the tension between constructing cultures for knowledge and learning which take into account the 'unfolding' nature of professional expertise in a complex world and the need to stabilize knowledge for responsible use" (p vii).

A theme in the scholarship we're reviewed is the importance of the infrastructure supporting professional practice, in some cases to support the uptake (and evolution) of an intervention in context and in others to support continuous improvement and learning.

Aspects of the infrastructure include conceptual frameworks that support noticing and meaningful interpretations of learning; material resources, including technologies, that enable networking and sharing of knowledge; coherent organizational norms and routines that support learning at the school, district, and professional organization levels; flexible access to general knowledge to address problems, support decisions, and guide actions as needed; new roles and relationships for practitioners and researchers to enable mutual learning; and attention to outcomes that include learning by professionals and organizations (e.g., as reflected in evolving knowledge cultures), as well as by their students. The example studies we've provided throughout this chapter highlight many of these features, including evolving relationships and shared responsibilities among researchers and practitioners (e.g., Cobb et al., 2013; Coburn and Stein; 2010; Fishman et al., 2013; Penuel and Spillane, 2014; Wyman et al., 2010).

From this generalizing-as-learning perspective, well warranted knowledge-in-action requires collective attention to the multiple perspectives of teachers, administrators, researchers, policy makers, and other education professionals working within cultures and infrastructures across levels of the system. Dialogue across these multiple perspectives supports innovation and evolution in perspectives: "by looking through a different lens…we may be able to envisage opportunities for action where we did not envisage them before" (Biesta, 2007, p. 19). Coburn and Talbert (2006) reach a similar conclusion based on their two-year study of evidence use in school districts:

> a coherent systemic strategy for evidence based practice may require a system of evidence use that allows for and supports access to different kinds of evidence for different purposes at different levels of the system. Individuals with different work roles have substantively different data needs. A strategy for evidence-based district reform must acknowledge these differences and create mechanisms to bring productive dialogue and coordination across them. Here, congruence does not mean the same conceptions of evidence; rather, it means creating complementary approaches at different levels and functions. (p. 491)

Contextualizing work like this can, in turn, contribute to more general knowledge via a variety of mechanisms. Schorr and Farrow (2014), writing for the Center for the

Study of Social Policy, draw on health-related research to call for "practice-based evidence" alongside other forms of evidence in addressing social problems.

> Incorporating practice-based evidence into ongoing work can take many forms, including reflective practice, the blending of surveillance and monitoring, community-based participatory research, and participatory action research. It can add to the likelihood that changes in practice will be well integrated with local strengths and values. And it can increase the effectiveness of interventions. In putting together practice-based evidence, "the real, messy, complicated world is not controlled, but is documented and measured, just as it occurs, 'warts' and all". (pp. 17-18, internal quotation not cited)

Comparative case study research and theory-based generalizations of the sort described in the previous sections enable such work to contribute, in turn, to more general knowledge.

Analytic Summary

The logics of generalization we reviewed engaged the relationships between theory and cases in different ways: they included more deductive forms of theorizing that conceptualized cases as opportunities to test and elaborate theory, thus privileging commonalities across cases; more inductive forms of theorizing that conceptualized cases as opportunities to build theory, privileging the particulars of each case; and mechanism-based theories or explanations that drew variously on the previous two strategies to address change oriented questions in complex systems. These logics also informed approaches to systematic reviews or research syntheses that sought to integrate knowledge across existing studies, or to use the comparison to illuminate taken-for-granted perspectives and practices for critical reflection. These conceptions of generalization emphasized contributions to collective or general knowledge. Then we turned to the ways in which knowledge (or knowing) might be constructed and put to work to inform policy and practice. There we reviewed theories of professional and organizational learning, considering the status of general knowledge in local contexts, the relevance of local knowledge, the kinds of structures and cultures that support evidence-informed practice in context, and more briefly the ways in which practice-informed evidence might contribute to general knowledge.

The categories we highlighted in the analytic summary of the AMPs section are relevant here as well. These included categories relevant to mapping dynamic systems, exploring different roles for researchers and other stakeholders and different types of collaboration, and enacting different purposes for engaging methodological pluralism. The scholarship reviewed in the CoGs section, especially the subsection on generalizing as learning, illuminated two additional purposes for engaging methodological pluralism to add to our list. "**Support of on-going practice**" emphasizes how knowledge or knowing is put to work in context to support understanding, decision making, and wise action. "**Innovation**" speaks to value of the dialectic between research and practice for enhancing general knowledge and creating novel solutions to educational problems.

In the AMPs section, the focus of the categories in the analytic summary was on conceptualizing programs of research prospectively; here the focus is on generalizing or learning from multiple studies enacted across MTs retrospectively. Taken together, the categories can be used to construct programs of research from existing studies that might not all have been previously conceptualized as relevant to the focal phenomenon, much as we have here with studies of research use and engagement, and then to organize, summarize, contrast, and critique the set of studies so constructed, as we'll illustrate in the next section on "Heuristics for Engaging Methodological Pluralism."

5. HEURISTICS FOR ENGAGING METHODOLOGICAL PLURALISM

Underlying the design of this chapter, and each of the approaches to methodological pluralism we have presented, is a productive tension which we believe can/should be dialectically engaged in any research practice: the tension between critical reflection aimed at learning from differences (Moss, 2005; Moss et al., 2009) and integration or bricolage aimed at building a program of research across those differences (Cobb, 2007). Up until this point, by presenting and comparing a range of different perspectives, we have privileged opportunities for critical reflection foregrounding the differences among Methodological Traditions (MTs), Approaches to Methodological Pluralism (AMPs) and Conceptions of Generalization (CoGs) to illuminate the partialness (and partialities) of any particular methodology. In this section, our focus will be more on integration or considering how they might work together in building a program of research and putting it to work to enhance educational practice, while also acknowledging the critical questions that comparisons among methodologies raise. We emphasize again that our focus here is on *programs* of research—not single studies or even inter-related series of studies—as any given study or series will have to circumscribe the methodologies on which they draw. The emphasis on programs of research situates any particular study or methodology within broader understandings of its affordances and limitations, its antecedents and consequents, and its *contribution* to knowledge about the phenomenon.

A conventional approach to the problem of designing research programs has been to provide general frameworks that scaffold program development, such as the visions of programmatic research we see prominently reflected in the US federal and national context (NRC, 2002, 2004a; IES & NSF, 2013). These visions arguably privilege programs of research intended to culminate in generalizable causal inferences and effective interventions that can be scaled-up for widespread use. For instance, the consequential *Common Guidelines for Education Research and Development Proposals* (2013) developed by the Institute of Education Sciences (IES) of the US Department of Education (ED) and the National Science Foundation (NSF) foreground six genres of research.

> #1. Foundational Research provides the fundamental knowledge that may contribute to improved learning and other relevant education outcomes….
> #2. Early-Stage or Exploratory research examines relationships among important constructs in education and learning to establish logical connections that may form the basis for future interventions or strategies to improve education outcomes….
> #3. Design and Development research…develops solutions to achieve a goal related to education or learning….
> #4. Efficacy Research allows for testing a strategy or intervention under 'ideal' circumstances….
> #5. Effectiveness research examines the effectiveness of a strategy or intervention under circumstances that would typically prevail in the target context….
> #6. Scale-up Research examines effectives in a wide range of populations, contexts, and circumstances. (IES & NSF, 2013, pp. 9-10).

The last three of these six genres, collectively described as "impact studies" entail the methodological tradition we described above (following Shadish et al., 2002) as experimental and quasi-experimental designs for generalized causal inference (GCI):

> Efficacy, Effectiveness, and Scale-up research should use study designs that will yield impact estimates with strong causal validity and that, for example, could meet What Works Clearinghouse standards without reservations (http://ies.ed.gov/ncee/wwc/). Generally and when feasible, they should use designs in which the treatment and comparison groups are randomly assigned. (IES& NSF, 2013, p. 21)

While the authors assert that research is not linear—that, for instance, impact research can identify needs for fundamental research and that these six types do not represent the panoply of useful investigations (p. 8), the document is nevertheless shot through with language that describes stages in a research process culminating in generalizable causal inferences as reflected, for instance, in terms like "early stage" and conclusions like "These guidelines can help practitioners develop a better understanding of what different stages of education research should address and be expected to produce" (p. 24).

Our approach is not intended to reject this vision but rather to situate it in a broader field of possibilities. Our intent is to highlight the limitations of *any* guidelines that privilege (empower) particular methodologies, stages of development, and points of culmination outside of deep consideration of substantive programs of research. If the social researchers whose work we reviewed in the previous sections have taught us anything, it is that educational systems are open and complex, context matters deeply, and change is inevitable. The question is how to conceptualize programs of research (and research policy) that respond resourcefully to these circumstances.

The Function of (Our) Heuristics

At this point, then, our goal is to draw together the resources of the last three sections—the various MTs, AMPs, and CoGs, and the contrasts among them—to suggest a set of heuristics that might support the conceptualization of multi-methodological programs of research focused on particular topics or (sets of) problems. In contrast to the sorts of a priori principles offered by the *Common Guidelines*, we offer something more closely resembling a working set of building blocks that can be used to identify the limits of such a priori principles and to design programs of research that take the unique circumstances relevant to a social phenomenon into account. We choose the word heuristics both to signal goals of learning and discovery and to forestall any notion that there is or could be a correct or complete framework.

In general usage, the word "heuristics" signals strategies that serve as an "aid to learning, discovery, or problem solving" (Merriam-Webster on line), especially exploratory problem solving and that enables "students... to discover or learn something for themselves" (OED online). The term is also associated with efficiency or effort reduction, referring to "simple processes that replace complex algorithms (Newell and Simon, 1992, in Shah and Oppenheimer, 2008, p. 207) or "methods for arriving at satisfactory solutions with modest amounts of computation" (Simon, 1990, in Shah and Openheimer, 2008, p. 207). For mathematician and mathematics educator George Polya (1957), whose theorization of heuristics in mathematics has been drawn on in interdisciplinary contexts, "Heuristic, as an adjective, means 'serving to discover.' Heuristic reasoning is reasoning not regarded as final and strict but as provisional and

plausible only, whose purpose is to discover the solution of the present problem. … Heuristic reasoning is often based on induction, or on analogy" (p. 112).

The social researcher whose use of the term most closely parallels ours is Andrew Abbott (2004) who offers readers heuristics for developing research questions by leveraging comparisons among approaches to social research. Abbott, citing Polya, points out that in the social sciences, unlike mathematics, "we often don't see ahead of time exactly what the problem is, much less do we have an idea of the solution" (pp. 82-83). Heuristics must function to help us generate and frame problems as well as to explore, develop, try out, and evaluate solutions. Thus for Abbott, heuristics help us as social researchers to find "bases for new ideas" (p. 85), free ourselves "from the restrictions of ...[our own] point of view" (p. 86), "master the basic viewpoints and even the heuristic repertoires of *other* stances toward the social world" (p. 86); and use "the different stances to question one another" such that "each stance begins to challenge all the others" (p. 86) ultimately reflecting "far more complex forms of questionning" (p. 87) that draw on combinations of stances. He suggests that researchers progress in their ability to engage heuristics to serve more critical and complex purposes over time.

We intend our heuristics to be of use across the life of a program of research, to support: initial scoping and conceptualization; ongoing mapping of relevant research available in the literature or newly reported; on-going critical reflection on values, contributions, conflicts, and consequences; decisions about what additional research or actions to undertake to enhance understanding and practice; and decisions about when closure or new directions are appropriate. In short, our heuristics are intended to support researchers in taking stock and in making decisions about what to do first and what to do next.

In Table 2 we provide a relatively abstract and sweeping overview (map) of our proposed heuristics for conceptualizing and organizing a program of research and then, in the next subsection, we speak to ways the heuristics might play out in the context of research relevant to our theme of how data and research are used and produced. Consistent with our notion of heuristics, we see the appropriate role for a map like this not as pre-specifying what should be included in a given program of research but rather as supporting researchers in systematically considering a range of ways in which a

phenomenon might be studied and change enabled. Comparisons among the methodological traditions and the approaches to methodological pluralism and generalization we have reviewed suggest consideration of the sets of dimensions we have summarized in Table 2: (a) that we locate the phenomenon of interest in the system within which it functions, analyze that system to illuminate its interacting components, and consider the other systems with which it interacts;[61] (b) that we capitalize on differences among methodologies or the perspectives on social phenomena they foreground to highlight a range of potential studies; (c) that we consider who participates in conceptualizing and enacting a program of research, how, and with what authorities and responsibilities; and (d) that we consider different purposes and strategies for comparison and integration across traditions as our AMPs and CoGs have suggested. The dimensions can be used to organize existing studies focused on a particular phenomenon and to conceptualize new or evolving programs of research.

Table 2: Heuristics for Conceptualizing and Organizing Programs of Education Research

A. MAPPING THE RELEVANT SYSTEMS		
1. Identification of a **focal system and analysis of its components**, including people (agents, participants); conceptual and physical objects; subsystems; levels and types of organization or structure, both formally designated and emergent; and activities and processes; 2. **other related systems** that might be implicated, their relationships with the focal system, the ways in which people and objects cross boundaries between systems; 3. the **evolving interactions and relationships** among these components/systems and the different time scales on which they occur, from moment-to-moment micro-level interactions among people to slower macro-level interactions among organizations and agencies; 4. the **ways in which interactions at one level of a system enable and constrain interactions at other levels**, including how group or system level features emerge from interactions among elements (e.g., routines, building architectures) and how these emergent phenomena in turn shape subsequent interactions; 5. how the focal and related systems **change** over time. [adapted from texts reviewed in complex systems theory (CST) section, especially Larson-Freeman & Cameron 2008, and Lemke & Sabelli, 2008]		
B. CONSIDERING DIMENSIONS ALONG WHICH METHODOLOGICAL TRADITIONS VARY		
CATEGORIES FOR COMPARING MTs	***METHODOLOGICAL TRADITIONS (MTs)***	***DIMENSIONS RESULTING FROM COMPARISONS AMONG MTs***
Aims & Conceptions Of Social	**CAS (ABM)**: Complex	1. **numbers and depth of cases**, from in-depth study of single cases, to studies that compare

[61] There are many theoretical frameworks that could support the sorts of system-level analyses we have relied, here, on complexity theory to support. In our own work, we have drawn on sociocultural theory, especially cultural historical activity theory (Moss et al., 2006; 2008) to assist with systems level analyses. Other relevant theories can be found, for instance, community psychology (e.g., Tseng, 2007), ecological theory (Bronfenbrenner, 1979, 2004), organizational and institutional studies (e.g., Barley, 2008), and a range of social theories, including Bourdieu's reflexive sociology (Bourdieu and Wacquant, 1992), Giddens's structuration theory (1984), or Martin 's social structures and explanations of social action (2011), each with their own emphases and purposes.

Phenomena a) the aims or goals of research within the tradition; b) the understanding or delineation of social phenomena that are the focus of the tradition; **Design & Enactment** c) the logic of inquiry or nature of research design/decisions as a study evolves from initial conceptualization through reporting of a completed study; d) key methods and/or processes through which the research is enacted, including data collection or production, analysis, and interpretation; e) the roles of researcher and researched and the relationship among them; f) the theorists' sense of evolving variations within the tradition and how their work is situated among them **Reporting & Evaluation** g) how the research is represented when it is reported to others; h) how validity, soundness, rigor, worth or quality are conceptualized; and i) conceptions of how knowledge generalizes or is relevant outside the particulars of what was studied within the MT.	Adaptive Systems Theory (Agent Based Modeling) **CCS**: Comparative Case Study Research **CRT** Critical Race Theory **DA (CDA & SFL)**: Discourse Analysis (Critical Discourse Analysis & Systemic Functional Linguistics) **DBR**: Design Based Research **EthR**: Ethnography/ Ethnographic Research **GCI**: Experimental and Quasi-Experimental Designs for Generalized Causal Inference **PAR**: Participatory Action Research **SNA**: Social Network Analysis **SR&M**: Survey Research and Measurement	small numbers of cases, to studies that aggregate large numbers of cases; 2. **emic and etic perspectives**, attending to understandings and intentions or research participants in context (emic) and to the theoretical frames and perspectives that researchers can bring to the context from outside (etic); 3. **nature and aspects of phenomena studied**, whether status at one point in time or process over time; whether relatively decontextualized or studied in relation to other case-specific circumstances and to antecedents and consequents; whether formally constituted or informal and emergent; 4. **time scales & relationships among them**, from more micro to more macro phenomena to relationships among phenomena operating at different time scales; 5. **approach to design and analysis**, from more inductive, to more deductive, to more iterative and dialectical; 6. **conceptions of cause or explanation**, from focus on a single cause to multiple interacting and recursive causes on different time scales; from designed interventions or other external influences to emergent phenomena arising in self-organizing systems; and from those that operate beneath the awareness of actors involved to those based on intention and action-orienting critical reflection. 7. **conceptions of change or progress**, across time scales and levels of the organization or system, and across distance from the phenomenon studied, from more proximal to more distal, 8. **loci for generation of research problems or questions**, from locally generated to theory or literature based to those that evolve in contexts of interest as a study unfolds; 9. **relationship between researcher and researched**, including roles and authorities of those studied at different points in the research process; 10. **standpoints of researchers**, from explicit to implicit, from espousing an ideal of neutrality to espousing an explicit value orientation, from participatory to remote.

C. IDENTIFYING STAKEHOLDERS AND TYPES OF COLLABORATIONS

1. **research communities** that reflect different methodologies, theories, philosophies etc.;
2. **communities of stakeholders** who work with(in) the phenomenon of interest; and
3. **publics** who are impacted by the work;
4. timing and degree of **dialogue and integration**; and
5. **roles and authorities in shaping** the research.

[Adapted from texts reviewed in the multi-, inter-, and transdisciplinary research section (MIT-D), largely Klein 2007.]

D. CONSIDERING PURPOSES FOR METHODOLOGICAL PLURALISM
1. **development,** such that earlier studies inform the design of later studies;
2. **triangulation,** for corroboration or convergence, but also for illuminating differences for further study;
3. **complementarity or configuration**, to illuminate different aspects of the phenomena of interest and the systems within and through which they function;
4. **challenge or confrontation,** to illuminate taken for granted perspectives and allow them to evolve;
5. **support of on-going practice in complex contexts,** and
6. **innovation.**
[Adapted from texts reviewed in mixed methods (MM) section, largely Green, 2007, as well as the MIT-D and conceptions of generalization (CoGs) section.]

Illustrating our Heuristics

In conceptualizing and organizing a program of research, one might begin with a social phenomenon of interest: it might be relatively broad, like a funding priority of a foundation or government research office or relatively narrow, as with a particular intervention, context of interest or locally generated problem, or anywhere in between. As we write, two foundations we know of, for instance, have identified current funding priorities relevant to issues of data or research use and are funding on-going programs of research: The W. T. Grant Foundation's research interest in "the use of research evidence in policy and practice" and the Spencer Foundation's strategic initiative on "Data Use and Educational Improvement." The US Department of Education's Research Advisory Board, the National Board of Education Sciences, has similarly foregrounded "knowledge utilization" (NBES , 2012) as an area in which research is needed. Examples of more narrow starting places are provided by any of the studies we used as examples in the previous sections. The broader foci bound spaces in which relevant research can be located; the narrower problem- and context-specific foci address local needs, provide illustrations of, and may both draw upon and contribute to research relevant to one or more broader foci. The examples used across the previous three sections of our chapter spoke sometimes directly, sometimes indirectly, to how data and research are used and engaged by professionals (teachers, administrators, teacher educators and professional developers, policy makers) working at various levels of the education system and in related professional organizations. In playing out the implications of Table 2 for organizing and conceptualizing a program of research, we take the communities of professional educators working in schools, primarily teachers, as the organizational level

of primary interest. Then, our focal phenomenon is how they use and produce sources of evidence (locally generated data, pre-existing research contributing to general knowledge)—including how they learn to do so—in framing problems, exploring solutions, making decisions, and taking action within their contexts of work.

It's important to acknowledge that our focal phenomenon can be framed in multiple ways within and across MTs. While MTs shape how a phenomenon is framed within any given study, they don't determine its framing. How phenomena are framed is also shaped by the research discourses in which researchers participate including their substantive theories, the value commitments that animate them, and the other stakeholders and contexts with which they are working. Across the examples we've used, we've seen the focal phenomenon framed in terms of whether or how a particular program, policy, or reform effort intended to be relevant across multiple schools leads to one or more outcomes (sometimes studied in large N research as with randomized field trials, sometimes in small N research as the program's implementation is traced in a few local contexts, sometimes as a single case); in terms of how particular professional communities learn to reason about the relationship between teaching practice and students' learning; in terms of how organizational norms and routines enable and constrain teachers' learning opportunities; in terms of how social networks within and beyond a school community and the information that flows within them can both support and reflect teachers' practice; in terms of how teachers' interactions with students (and the evidence sources entailed in those interactions) enable and constrain students' learning; in terms of how teachers' professional organizations make relevant resources and ultimately professional knowledge cultures available; in terms of the antecedents and consequents of inequities in access to resources, to ambitious learning opportunities, or to agentic identities; in terms of how the professional school community interacts with the local education agency, including who and what crosses boundaries, to what effect. Any one or more of these framings could serve as a starting place for a program of research relevant to the focal phenomenon we've described and collectively they begin to flesh out a robust program of research that could be mapped in terms of the dimensions and categories of Table 2. Any research problem could reflect one or more of these frames and could be seen as contributing to a middle range theory (Merton, 1967) or a "local

research frontier" (Mjøset, 2009, where local means specific to a problem, not local context). In the subsections below, we play out the implication of Table 2 illustrated with general reference to the examples cited throughout this chapter.

Mapping the Relevant Systems

As we noted in the AMP Complex Systems Theory (CST) subsection, in conceptualizing a program of research around a given phenomenon, one goal is to bound those aspects of the relevant system(s) necessary to understand how the phenomenon functions and how the focal system(s) of which it is a part might be enhanced. Another goal (often enacted iteratively with the first) is to analyze the focal system, naming its components, the levels and time scales on which they interact, the other systems with which the focal system interacts, and the relationships within and between them (Larsen-Freeman & Cameron, 2008).

Following Part A of Table 2, we might begin by describing likely components of the focal system within which our phenomenon is located. Of course, there is no one correct representation of the components of a system. The particular components chosen will differ in multiple ways by schools and the systems within which they are embedded or with which they interact, as well as by researchers' purposes and perspectives. CST reminds us that these are open systems in constant motion, sometimes reproducing themselves, sometimes evolving to new states that may be less or more productive in terms of valued goals. The mapping of any given system will become more nuanced as a program of research proceeds and its particularities are identified. Here we move up a level of abstraction to suggest components more likely to be relevant across local contexts and research. Components interacting in the focal systems in our examples included

- teachers and school administrators engaging in formally designated and informally formed subgroups;
- working with objects, including physical objects (like digital devices, test reports, samples of students' work, or video records of classroom interaction) and conceptual objects (like subject-matter concepts and representations or analytical language for describing students' learning) either locally developed or inherited and adapted;

- engaging in activities or processes enacting patterns of interaction that reflected norms and routines, again either locally developed (emergent) or inherited/adapted;
- within organizational policies and structures (like schedules, meetings, reporting requirements, evaluation policies); and
- often with members of other groups or communities who regularly interact with teachers (students, researchers, teachers in other contexts, leaders or resource staff at other levels of the education system, coaches, facilitators, or professional developers).

However our focal phenomenon is framed in terms of a research problem, components like these will likely impact the way in which it works within a system and could be productively included in a program of research. Some will become an explicit part of any particular research agenda and others will be left to operate in the background.

Then there are the systems with which our focal system interacts. Whether any given system should be foregrounded in a program of research depends, in part, on how the study is framed and on its relevance to the focal context—how it shapes and is shaped by the individuals and groups that are the focus of the study. Interacting systems in our examples included students in various groups and subgroups working with the teachers; leaders in schools and local education agencies routinely interacting with teachers; formal and informal professional networks within and beyond the schools; and research teams. While not necessarily illustrated, families, neighborhoods, business partners, local schools and colleges where students might subsequently enroll, social work and health organizations, or affinity groups of various sorts might well be relevant depending on local circumstances. While large N studies, of the sort that GCI and SR&M permit, aggregated over or attempted to control for these different influences to enable conclusions across contexts, small N and single case studies informed, for instance, by CCS, CRT, DBR, EthR, or PAR illuminated them and traced how they function. Of course, the various systems that produced the physical and conceptual objects with which teachers interacted are also relevant (including the systems through which policies are produced, tests developed and evaluated, and so on), but what may be most relevant in these cases is how the products that emerged from these other systems were taken up and

interpreted by teachers (in light of all the other components of the systems in which they work). We note as well that understanding how these polices and products are taken up in local contexts could inform the systems in which they were developed.

These systems and components of systems operate on different time scales and at different levels of the organization. In the paragraph above, we've already signaled different levels of organization within the system, including the groups and subgroups (like professional learning communities, classrooms of students, or subject matter departments) through which they operate and within which they are embedded (like local and state education agencies, charter organizations, or national systems). With respect to time scales, our focal phenomenon of data and research development and use was studied, for instance, from moment to moment time scales, where teachers observed what students had just said or done and decided what to say or do next, to slower times scales entailed in planning the next day's lesson or the next sequence of lessons in the next unit based on evidence of students' learning, to yearly or longer time scales where professional development was planned or curricula selected based on system level indicators. Again, which time scales and levels of organization are deemed relevant depend on the context(s) and the framing of the research problem or phenomenon. To illustrate the guidelines about looking at actions and interactions on different scopes and time scales, we turn to the guidelines that arose from our comparisons among methodological traditions.

Considering Dimensions Along Which Methodological Traditions Vary

In Part B of Table 2, we have listed the methodological traditions (MTs), the categories we used to compare them, and the dimensions that resulted from those comparisons. It is through successive iterations among these three features that we selected traditions to study, developed dimensions along which they varied, and settled on categories that enabled fair comparisons. As heuristics for building multi-methodological programs of research, either the MTs or the dimensions could be useful. The initial comparative categories are provided here to support readers in integrating new MTs into the framework, thus expanding the dimensions along which they might vary. In

this section, we focus on the dimensions, grouped in the subsets as represented in the Table 2.

The first entry, on number and depth of cases, is implicated in each of the other entries. We won't discuss it further here except to say that it highlights the complementary values of understanding particular cases in depth as well as understanding tendencies across larger numbers of cases, thus providing different vantage points from which to understand how systems work.

While (almost all) studies include etic components—which researchers and their anticipated audiences bring to a study—emic understandings of people who work within the system studied can be crucial. We saw examples from EthR, CCS, CRT, DA, DBR, and PAR where local meanings were foregrounded, including cases where the intentions/expectations of policy makers or program designers were challenged. Importantly, analyses enacted within these methodological traditions also illustrated how norms and routines, conceptual understandings, and artifacts emerged from patterns of local interactions and how these emergent phenomena in turn shaped interactions among agents. ABM and SNA provided additional tools for anticipating and studying this sort of emergence. Examples enacted within GCI, DBR, including Design Based Implementation Research, and SNA also allowed the tracing of interactions (involving causes and effects) with more cases over longer time scales.

Across the MTs, we saw a range of approaches to conceptualizing and studying causes and explanations. These differences are among the most controversial and heavily debated in the methodological literature. Some argue that GCI, where the counterfactual outcomes of receiving the treatment (for control group members) or of not receiving the treatment (for treatment group members) can be approximated ideally through random assignment, is necessary for warranting causal conclusions. In the context of data use, we illustrated research using GCI with a program intended to support teachers' learning across multiple contexts and we cited other studies where GCI has been implemented, including programs of GCI research where multiple potentially confounding factors were systematically examined as well as meta-analytic research syntheses where effect sizes were aggregated across studies. Questions raised from within other MTs suggested ways in which GCI might under or misrepresent the complex ways in which multiple causes

operate in open and dynamic systems. As CST theorists argued, cause operates in any and all directions, often in ways that cannot be anticipated, where the same outcome can be produced in multiple ways and different outcomes can result from similar configurations of mechanisms. CCS, through process tracing and qualitative comparative analysis, as well as DBR, through iterative mini- and macro cycles, offered extended arguments for alternative approaches to studies enabling causal conclusions. EthR, DA, PAR, and CRT also enabled the tracing of multiple interacting mechanisms over time and across different levels of the system and time scales; illustrated how concepts and categories evolve from and produce patterns of understanding and action; and showed how agents' critical awareness of these forces can enable change. SNA and ABM treated evolving networks, interactions among agents, and emergent phenomena as causes (as well as outcomes). In depth examination of particular cases let us see how any particular causal variable interacts with other variables that combine to (re)produce the state of the system at any given point in time. Thus we saw, for instance, how small shifts in patterns of interaction or introduction of new routines can contribute to productive changes in a group's norms; we also saw how externally designed programs evolved when put to work in local contexts, sometimes to better address local circumstances, sometimes in ways that undermined the intent of designers, as when approaches to inquiry became formulaic.

Importantly, the MTs we reviewed also illustrated the value of attending to different kinds of outcomes, on different time scales, suggesting different (and often complementary) conceptions of improvement. These included attention to shifts in: the qualities of moment-to-moment interactions among teachers or between teachers and students; the norms and routines within teacher communities; the networks and structures within which teachers work; the nature of diversity within the system; conceptions of learning and allocation of responsibilities for learning; shifts in understandings of self and other; and the enactment of agentic identities for supporting positive change. While a key long-range outcome for education research and research on teaching is students' learning, including the outcome of evidence-informed practice, we saw multiple examples of students' learning illuminated by different methodological traditions, from the learning reflected in moment-to-moment interaction to that reflected on annual achievement tests, and from learning that might be conceptualized in terms of knowledge and skills, to

learning that might be conceptualized in terms of shifts in students' identities and participation in practice.

Identifying Stakeholders and Types of Collaboration

Turning to Part C of Table 2, a key question is who should be at the table in conceptualizing a program of research, as the available understandings and experience will shape how a program is conceptualized. In the examples we presented above, we saw the potential not just of different methodologies and theoretical perspectives, but also of active and sustained collaboration with professionals engaged in the practice being studied, and with other stakeholders impacted by the work. These issues were emphasized, in particular, in the discussion of inter- and trans-disciplinary research, in DBR and PAR; and in research syntheses intended to support practice, although they were also present elsewhere. Questions about the norms of collaboration and the authority to shape the research, which are entailed in any project, were explicitly considered as well. The goal was to move beyond taken-for-granted practices of design to illuminate the potential consequences of different choices, for research, practice, and the relationship between them.

Across the examples in previous sections, we saw a range of different relationships between researchers and stakeholders in local contexts, from those where hypothetical scenarios were put into motion by researchers modeling possible interactions among a system's agents; to those where local stakeholders and systems served primarily as sources of evidence; to those where stakeholders participated in shaping the research agenda, formulating problems, or in reviewing and confirming or challenging interpretations of data; to those where they were full partners in all aspects of the research process; to those where researchers served as consultants to assist local stakeholders in formulating and responding to their own questions; to those where the line between researcher and researched disappeared such that everyone's practice provided a source of evidence for critical analysis. These differing roles and relationships, which might complement or challenge one another, offer different possibilities for contributing to general and local knowledge and for enabling change. We have seen, for instance, examples illustrating multiple ways in which the results of published research, including large scale studies where initial involvement of local stakeholders was necessarily

limited, can inform local practice by offering possible courses of action that have worked elsewhere/elsewhen and by suggesting hypotheses for local consideration. And we have seen examples of how single case and locally generated studies, like those of EthR, CRT, DBR, or PAR, can contribute to general knowledge inductively through theory-based generalizations across cases.

Already implicit in these descriptions are a range of possible stances for researchers vis-à-vis their relationship to stakeholders and to the local contexts. Beyond these, we note that the examples have illustrated fundamental (and often controversial) differences in conceptions of objectivity and subjectivity in the research process and their implications for research quality. Some methods/methodologies seek to control researchers' biases, often through independent replication, to enhance objectivity; other MTs seek, instead, to illuminate researchers' biases, often through confrontation with alternative perspectives, to enhance collective critical reflection on their genesis and effects (Moss et al., 2009; Moss, 2005). Again, these differing stances offer various possibilities for contributing to general and local knowledge and for enabling change.

Considering Purposes for Methodological Pluralism

This brings us to Part D of Table 2, which focuses on purposes for methodological pluralism in programs of research. Each of the sections above illustrated different approaches to integration: multi-method research within a dominant MT, AMPs offering various purposes and strategies of integration, and CoGs that supported understanding and learning across studies and methodological perspectives. Looking across them illuminates a range of purposes for engaging in methodological pluralism in building a program of research. Here we highlight some broader themes (adapted largely from texts we reviewed in the MM and MIT-D Research sections, especially Greene, 2007, 2012, and Klein, 2007, 2010, and in the CoGs section, especially generalizing as learning.).

We saw multiple examples within specific programs of research for how one study or set of studies informed the ***development*** of the next study or set (e.g., interviews informing development of surveys, surveys identifying cases for in-depth study, ethnographic research illuminating issues to take into account in design-based research, case studies informing development of interventions intended to be implemented at scale,

and implementation of interventions with the explicit anticipation and study of local adaptation). While there were many examples of ***triangulation*** within MTs, looking across the examples presented suggests the opportunity for triangulation across MTs, including for instance the comparison of large scale and small N case studies focusing on the effects of system wide interventions to support data use. More often than not, however, these comparisons enable ***complementary*** perspectives that, taken together, provide a richer, more nuanced understanding of a given phenomenon, sometimes coherent, sometimes contradictory. For instance, there were multiple examples from CST, mechanism-based generalizations, and research syntheses that illuminated the value of such complementarity. More generally, the examples we have provided throughout the chapter, partially overviewed here, begin, we hope, to illuminate this potential. The emphasis on ***challenge or confrontation*** returns us to the general theme of learning from our differences, suggesting that one way of reading across the examples is not so much to illuminate their complementarities but rather to highlight the ways in which any single methodology provides at best partial understanding and limited affordances for action. Seeking out contradictory findings illuminates new problems for further study. It also provides opportunities for researchers to consider how taken-for-granted perspectives and practices within any given MT, however rigorously conducted, can reflect systematic biases that distort a focal phenomenon and limit possibilities for wise action. While we've tended to frame these purposes in terms of their potential contributions to general knowledge about the focal phenomenon, the purpose of ***support of on-going practice*** in complex contexts highlights the crucial work that professionals and organizations must routinely do to put knowledge to work in making decisions and taking action. ***Innovation*** serves as a place holder for the iterative processes through which local practice can inform general knowledge and, more fundamentally, through which the heuristics map can evolve, with new elements being introduced and existing elements being modified accordingly.

Alternative Approaches for Structuring Programs of Research

As we argued, we intend our heuristics to enable the conception and organization of programs of research that focus on the particulars of the relevant social phenomena and the contexts in which they work. That said, we do not mean our building block

approach to discount the potential *heuristic* value of more general characterizations of programs of research that respond to classes of problems. Our concern is when such general characterizations are treated as a priori prescriptions for how programs of research *should* proceed, as they are with the *Common Guidelines* (IES & NSF, 2013). As Topper (2005) suggests in his defense of methodological pluralism, the goal for any program of research or typology of such programs is to justify "how, why, and in what context it yields a more adequate understanding of social...phenomena [and] how it enables us to cope more effectively with the world around us (or to understand under what conditions or to what degree 'coping more effectively with the world around us' may be something one should not wish to do)" (p. 187). When viewed in conjunction with other such frameworks and mapped against a broader set of possibilities like those in Table 2, their limitations can be illuminated alongside their potential.

For instance, here are two complementary visions from design-based implementation research (Fishman et al., 2013) and the learning sciences (Nathan and Sawyer, 2014; Nathan and Alibali, 2010) that have resolved some of the choices reflected in our heuristic framework in a different way from the *Common Standards*. Fishman and Penuel (2014) argue that design based implementation research should follow four core principles:

> (1) a focus on persistent problems of practice from multiple stakeholders' perspectives; (2) a commitment to iterative, collaborative design; (3) a concern with developing theory and knowledge related to both classroom learning and implementation through systematic inquiry; and (4) a concern with developing capacity for sustaining change in systems. (Fishman, et al, 2013, pp. 136-137)

In their explanation of these general principles, they highlight the relevance of tailoring research-based practices to different contexts. Nathan and colleagues (Nathan & Alibali, 2010; Nathan & Sawyer, 2014) pick up on a similar point and imagine how a program of research that attends to both "systemic" and "elemental" approaches might be developed. Elemental approaches "focus on the component elements of a complex learning environment" (Nathan and Sawyer, p. xx) whereas systemic approaches "analyze learning at the level of the entire complex system" (p. xx). They note that some phenomena are better studied by analyzing the complete system and some by focusing on

individual elements. They propose a "scale down" alternative to the more conventional scale-up approach to research program development:

> The scale- down method begins by studying a system from a systemic perspective—by examining the learning environment in the complex settings in which it naturally occurs (e.g., a classroom). Then, analysis of these systemic observations is used to develop hypotheses for how to improve system performance, first by identifying potential subsystems … that impact system performance; second, by modifying the design and performance of these subsystems; third, by reintegrating modified subsystems into the system; and finally, by observing behavior of the system as a whole in its natural context…..
>
> The aim of scale-down is to improve systemic performance by improving, when possible, the design of subsystems—whether curricular materials, interaction patterns, or teacher behaviors. Refinement of the design and performance of a subsystem can draw on the precision and control that characterize elemental methods, as well as systemic methods of investigation and design. In this way, refinement of a … system is performed in a recursive manner, alternately using elemental and systemic methods when necessary. (p. xx)

Of course, these reflect only two of many possible visions of how programs of research addressing common problems might be designed, but they illustrate the potential for alternatives to a vision that has been prominent in the US federal context. Transdisciplinary research and PAR would, for instance, foreground the central role and authority of stakeholders. And various critical approaches to research, like CRT, as well as versions of EthR and PAR, might also call for research outcomes that illuminate the ways in which multiple (sometimes competing) interpretations and conclusions from available evidence are viable (see e.g., Lather, 2006, 2007) and that situate all programs of research in their historical and political contexts.

These sorts of contrasts highlight the value of seeking out challenges across methodological traditions and of locating any program of research or typology of such programs within a broader map of possibilities. Such commitments prompt critical reflection by illuminating taken-for-granted perspectives and practices so they can be

either re-affirmed with a new self-awareness or enabled to evolve. As Abbott (2001) argues "the focus on the larger framework is not merely intellectually useful as an idea, but also normatively proper as a commitment" so that we can "become explicit about what is implicit in our practices" (p. 5). A robust multi-methodological program of research that embraces the challenge of alternative perspectives not only enhances our understanding and capacity for action within particular programs of research but also sustains our growth and resourcefulness as a field.

In closing this section, we want to remind readers that the list of categories in Table 2 is far from exhaustive and, following Law (2004), "it is very important that it not be seen as exhaustive" (p. 145). As we've argued above, such maps and the research programs they might produce are always historically situated and perspectival; they do political work; and they must be routinely scrutinized for the ways in which they enable and constrain understanding and wise action. Our assigned emphasis on methodology has necessarily glossed over multiple equally important aspects of social research, differences among which could also provoke critical reflection and support integration. As we noted at the outset, any given study or set of studies necessary reflects a confluence of substantive theory, contributing disciplines and professional fields, underlying philosophical perspectives on the nature of social research, national culture, local knowledge and practice, as well as methodological perspectives. Methodologies cross and overlap these categories in varying ways. At a more radical level, Law (2004) highlights multiple goods that might be considered alongside epistemological warrants in designing programs of research and in attending to the specifics of context, including justice, aesthetics, emotion, and spirituality. "Debates of this kind," he argues, "would simultaneously be both broader and more modest than our current discussion of method. They would be more modest because they would arrive at particular conclusions in particular locations for particular studies" (p. 155).

We hope our example will encourage readers to join the conversation by elaborating, challenging, or revising our map, or by developing their own approaches for capitalizing on methodological pluralism. In our overall conclusion, we will consider the social dimensions of social research and consider what it might take to foster and sustain routine practices of methodological pluralism in the field of educational research.

6. CONCLUDING COMMENTS

> "For there is a danger of a *fragmenting* pluralism where ... we are only able to communicate with the small group that already shares our own biases, and no longer even experience the need to talk with others outside of this circle. There is a *flabby* pluralism where our borrowings from different orientations are little more than glib superficial poaching. There is *polemical* pluralism where the appeal to pluralism ... becomes rather an ideological weapon to advance one's own orientation. There is *defensive* pluralism, ...where we pay lip service to others "doing their own thing" but are already convinced that there is nothing important to be learned from them.
>
> The type of pluralism that represents what is best in our … tradition is an *engaged fallibilistic pluralism*. ... Here one begins with the assumption that the other has something to say to us and to contribute to our understanding. The initial task is to grasp the other's position in the *strongest possible light*. Engaged fallibilistic pluralism....is based upon mutual respect, where we are willing to risk our own prejudgments, are open to listening and learning from others, and we respond to others with responsiveness and responsibility." (Bernstein, 1991, p. 339)

We undertook work on this chapter with the premise that methodological diversity in the field of education research is a rich and underutilized resource. Our intent was to support readers in learning from those differences and working across them to enhance both general knowledge and local practice. As Page points out in his book on diversity and complex systems (2011):

> Diversity can provide insurance, improve productivity, spur innovation, enhance robustness, produce collective knowledge, and, perhaps most important in light of these other effects, sustain further diversity. (Page, 2011, p. 3)

However, he also acknowledges that “diversity, for all its benefits, is no panacea: It can contribute to collapse, conflict, and incomprehensible mangles” (Page, 2011, p. 3). The question is how to maximize the potential benefits of diversity while minimizing the

risks. In our presentation, we have tried to balance respectful attention to differences and the critical reflection and learning that promotes with attention to strategies that "get on with it" (Lather, 1986, p. 77) and make collective progress in addressing educational problems. In this closing section, we situate our argument in on-going debates about quality and usefulness in social research, and we briefly consider its implications for research policy and professional development.

Situating Our Argument in On-Going Debates about Social Research

By emphasizing the importance of learning from our differences, we have taken sides in a long-standing debate in the social sciences about how best to manage diversity and avoid fragmentation. One approach, arguably dominant in the current US federal and national context, has been to privilege commonality and work toward an "intellectual synthesis" (Camic and Joas, 2004) to represent and guide social research. The other approach, for which we have argued in this chapter, privileges difference and seeks practices that sustain dialogue and collaboration across different perspectives in addressing social problems. One can see elements of the former approach reflected, for instance, in AERA's (2008) definition of scientifically based research, in the National Research Council's (2002) monograph on "Scientific Research in Education", and in calls for synthesis in key representations of Mixed Methods Research (Tashakkori & Teddlie, 2010). In its more radical forms, this approach has resulted in generalized guidelines about what constitutes rigorous research and a priori decisions about appropriate methodologies. This is arguably the approach reflected most recently in the *Common Guidelines* (IES & NSF, 2013) published by the US Department of Education and the National Science Foundation, where programs of research are expected to culminate in large scale impact studies (consistent with the GCI tradition). Prominent since the formation of the Institute of Education Sciences (IES) by the Education Sciences Reform Act of 2002, these priorities have been reflected (per the IES director's biannual reports) in funding announcements, guidelines for peer review, and funding decisions; in federally funded training programs for researchers; in guidelines for research syntheses developed by the federally funded What Works Clearinghouse, which was intended to make the findings of rigorous cause-and-effect studies easily available to

practitioners, and in "differential consequences for decision makers whose choices are or are not grounded in evidence" as defined in these consequential documents (IES Director's Biennial Report to Congress, 2005, p. 2) (see Moss, 2014; Phillips, 2006; Walters, Lareau, & Ranis, 2009 for an overview).

While such unifying syntheses deserve a full voice at the table where policy is made and programs of research conceptualized, they reflect only one vision for how to manage the diversity of methodological traditions and guide professionals engaged in the production of knowledge. Those who take a more dialogic approach to diversity and pluralism raise concerns about the consequences of working toward a single framework. Such syntheses risk disempowering alternative perspectives by positioning them as supplementary, by blurring boundaries that would illuminate them, and by what Bourdieu (1991) describes as an even more radical form of censorship, ignoring alternative perspectives and thereby keeping them "outside the limits" of the debate. In Lamont and Molnár's (2002) terms, prescriptions of the sort promoted by IES risk turning symbolic boundaries into social boundaries--"manifested in unequal access to and unequal distribution of resources (material and nonmaterial) and social opportunities" (p. 3). Arguing for the "dialogic turn," sociologists of science, Camic and Joas (2004) suggest:

> Rather than decry the multiplicity of theories, methods, and research findings and then seek their integration in a unifying framework, the characteristic of this response is that it welcomes the presence of plurality of orientations and approaches as an opportunity for productive intellectual dialogue (p. 5).

They enumerate the benefits of a more dialogic practice (drawing on the work of Levine):

> By working to grasp "truly alternative points of view"—three benefits accrue: "we learn more about the character and value of the positions we ourselves espouse"; "we learn to respect the position of the other"; and "we learn of other resources for our [respective] projects"—opening the door, in some instances, to fruitful combinations and partial syntheses or, in instances where these options are not feasible, to the critical enrichment and "reciprocal refinement" of diverse perspectives. (Levine, 1995, 1991 in Camic & Joas, 2004, p. 9).

With these scholars, rather than searching for general rules to distinguish rigorous research, we see it as more important than ever for researchers in education to cultivate a

synoptic view of research itself and a sympathetic appreciation of the distinct values and contributions of different ways of studying the phenomena of education. This is not to discount rigor within MTs—every MT can be enacted more and less well—but rather to acknowledge that how quality is conceptualized must be understood from within different MTs and research discourses. Equally important, however, as our opening quote from Bernstein (1991) suggests, *rigor lies in the willingness to confront our own preconceptions about methodology with perspectives and practices of others*. Moreover, it is at the borders of methodological traditions that "powerful heuristic tools to search for innovation" lie (Maldonato & Pietrobon, 2010, p. 1).

The literature we have reviewed pushes us further to consider not just dialogue within the community of researchers but, equally important, dialogue across professional contexts. This is most prominently reflected above in sections on participatory action research, research syntheses that foreground the information needs of multiple stakeholders, and attention to epistemic cultures within and across professions (Jensen et al., 2012; Knorr-Cetina, 1999). These are arguments that were raised in 1999 by committees of the National Academy of Education and the National Research Council (NRC), but were not taken up in federal guidelines, and they have been raised again by the NRC in 2012 in their review of research on the implications of science for policy.

Sociologist of science, Amsterdamska (2008) highlights the importance of attending to knowledge production in all its forms, including the ways in which knowledge is produced in practice:

> From a practice perspective, every diagnostic or treatment decision by a doctor, every choice of policy by a government regulatory agency, and every user's attempt to master new technology can be seen as part of the process of knowledge production. ...Much justified attention has in recent years come to settle on actors who are not scientists and on areas of activity where scientific knowledge, technological know-how, and research are made to intersect with other knowledges, skills, and tasks". (p. 209)

This entails research that is "problem focused, contexualized, and consultative" (Wickson et al., 2006, p. 1047) and that "tackles complexity …and uncertainty…[in] context specific negotiation of knowledge" (Lawrence, 2004, p. 399). Gibbons and Nowotny

(2001) argue that "contextualization of research around the interests of stakeholders fosters a more 'socially robust' knowledge that transgresses disciplinary and institutional boundaries" (p. 67). We have cited multiple examples of such collaborative knowledge production. And we have reviewed comparative and configurational strategies for generalization and research syntheses that privilege multi-methodological research, including research that honors differences across cases and contexts in contributing to general knowledge.

Implications of a Dialogic Approach to Methodological Diversity

In closing, we point briefly to the implications of a dialogic approach to methodological diversity. Implicit in the above argument, but beyond the scope of this paper, is the importance of studying the social dimensions of knowledge production and dissemination, including the preparation of new educational researchers, publication policies and practices, peer review, and research funding. As Longino (2013) defines it: "Study of the social dimensions of scientific knowledge encompasses the effects of scientific research on human life and social relations, the effects of social relations and values on scientific research, and the social aspects of inquiry itself." Such a goal is enabled by dialogue among alternative perspectives as it is by empirical studies of research policy and practice (e.g., Hackett, Amsterdamska, Lynch, & Wajcman, 2008; Camic, Gross, & Lamont, 2011).

We would hope that both beginning and established education scholars would come to embrace the view that a genuinely rigorous approach to social research seeks out challenges from alternative traditions, illuminating the perspectives and practices that we, in a given methodological tradition, take for granted as the way things are done, thus enabling them to evolve. As Warnke (1999) notes, whether we adopt the insights of the other perspective or not, we have learned through our encounter with it" (pp. 131-132).

> "Good interpretations [of others perspectives], …succeed in *illuminating the difference a given text makes to what we thought we knew*—either by giving added confirmation to our views or by asking us to rethink or expand them, *even if we do so in opposition to the text* we are trying to understand. We take the possibility seriously that the text can teach us something, and we therefore put our

> own views and assumptions into play in trying to understand it. (Warnke , 1999, p. 17, italics ours)

Stepping back, a key starting place for research policy and preparation entails "learning how to map [the discourses of educational research] in useful ways …[and to consider] the metatheoretical frame within which one makes strong, empirically responsible, theoretically informed knowledge claims about social research itself" (Reed , 2011). As we noted in the previous section, Abbott (2001) sees attention to the larger framework as an ethical commitment to make explicit what is implicit in our practices. We hope our chapter also makes a useful contribution to that broader goal. Of course, there are many more methodological traditions (methodologies, methods) and approaches to pluralism than those we have illustrated. And our assigned focus on methodologies has pushed alternative dimensions for representing social research into the background (e.g., philosophies, disciplines, professions, social theories, practice contexts). We remind readers that our map, like all such maps, is partial, historically situated, and perspectival. We hope our work supports readers *both* in attending to a larger framework *and*, equally importantly, in considering how to elaborate, challenge, revise, or replace our framework for their own purposes and to incorporate additional perspectives and practices.

In terms of professional preparation, a key goal will be to balance deep expertise in particular methodological traditions with opportunities to learn from alternatives. In addition to having access to broader frameworks within which one's own MT can be located, we would argue from our own experience for the value of learning at least one alternative MT in sufficient depth to understand the challenges it raises for one's own. And, following Bernstein (1991) this might best be conceptualized as learning a "second first language" (p 336, paraphrasing MacIntyre). Experience with multi-methodological, multi-professional ("transdisciplinary") collaborations seems crucial as well. Here, we take guidance from the call for "adaptive expertise" prominent in the Learning Sciences:

> Routine experts develop a core set of competencies that they apply throughout their lives with greater and greater efficiency. In contrast, adaptive experts are much more likely to evolve their core competencies and continually expand the breadth and depth of their expertise as the need arises or as their interests demand. This often requires them to venture into areas where they must function as

> "intelligent novices" who often struggle initially in order to learn new things. (Bransford et al., 2006, p. 26)

Of course, research policy and infrastructure need to be in place to support such learning opportunities. Here we argue that what is most needed are principles to sustain dialogue and learning from differences. This suggests the importance of considering carefully who is at the table when research policy and practice is set and working to ferret out the invisible exclusions (Bowker, 2005) that could limit opportunities for learning. The research we have reviewed also suggests the value of encouraging problem focused research, of making decisions about fruitful methodologies and structures for programs of research in light of particular problems and contexts at hand, and of actively seeking out opportunities for dialogue across different approaches to a problem.

Our hope is that practices of working knowledgeably across research traditions for which we have argued will become increasingly commonplace within our field and that this chapter will contribute to that goal. Embracing the learning opportunities of methodological pluralism will enhance our collective capacity to respond resourcefully to the unknown challenges ahead.

APPENDIX

Ethical Considerations

The topic of ethics in educational research is far too broad and too complex to attempt an adequate treatment within the scope of this chapter, but it is at the same time too important to ignore altogether. These brief comments are intended to suggest the scope of ethical considerations in the conduct of research and to direct attention to several salient issues. In these comments, we have relied heavily on a chapter by Kenneth Strike (2006) in the AERA *Handbook of Complementary Methods in Education Research*.

AERA has adopted and published a Code of Ethics, available on the AERA website[62] and also published in the *Educational Researcher* (AERA , 2011). Similar codes or statements have been adopted by other major professional organizations (e.g., American Statistical Association, American Sociological Association, American Anthropological Association, American Political Science Association, American Psychological Association). The American Anthropological Association maintains an Ethics Blog on their website[63] to promote ongoing public dialogue about real-world ethical dilemmas that arise in practice.

These various codes and statements address a broad range of concerns. Virtually none of these are entirely divorced from the research process, but some are relatively peripheral to research per se. Relatively peripheral concerns might include policies concerning sexual harassment in the workplace, faculty recruitment and hiring, or promotion and tenure processes, for example. Other ethical considerations are more directly relevant, such as those involving confidentiality, anonymity of sources (as appropriate), use of deception in research, conflict of interest, plagiarism, research authorship (e.g., the definition of a creative contribution), secrecy in research (i.e., limiting dissemination of research findings), ownership of research and of research data, editorial responsibilities and peer review. Perhaps most central is the ethical treatment of research participants.

---

[62] See http://www.aera.net/AboutAERA/AERARulesPolicies/CodeofEthics/tabid/10200/Default.aspx.
[63] See http://ethics.aaanet.org.

The 1978 Belmont Report guides the review of research using human subjects conducted by Institutional Review Boards (IRBs) at most universities. Strike (2006, pp. 68-69) summarizes the topics covered in that report as follows:

> The Belmont Report, *Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, is a report of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1978)…. The commission asserted three basic principles along with three primary applications: respect for persons, beneficence, and justice…. Respect for persons involves two subsidiary principles: 'First, that individuals should be treated as autonomous agents, and second, that persons with diminished autonomy are entitled to protections.' … Beneficence … is an obligation that has been expressed in two (inconsistent) basic rules: Do no harm, and maximize possible benefits and minimize possible harms. … Justice requires that benefits and burdens should be equally shared. The report … places special emphasis on avoiding cases where populations who are vulnerable are employed as subjects in research that primarily benefits others or where the class of those who will benefit is much broader than the class from which subjects are selected. … People should not be selected because they are vulnerable or perceived as less (or more) worthy. Benefits and burdens should be equitably shared across lines of race, gender, class, and culture. Research should not add to the burdens of already disproportionately burdened populations.

Research in educational settings may bring particular challenges with regard to protecting vulnerable populations because research participants are often children. In addition, as Strike (2006, p. 71) notes, "A second problem is that voluntary informed consent can be most problematic when one of the researchers is a teacher …. Suppose a teacher is unable to secure consent for one or more children in his or her class—must the research not be done? … Might parents fear repercussions if they withhold consent from their child's teacher? Finally, action research, including that done by teachers, may be quite open ended. Action research, by its nature, may be reinvented in progress. Clarity in detail as to its purposes may be difficult."

The preamble to the AERA Code of Ethics is of particular relevance given the scope and argument of this chapter. Strike (2006, p. 58) draws attention to the preamble's statement of the obligation of researchers to "maintain the integrity of research [by warranting] research conclusions adequately in a way consistent with [the researcher's] own methodological and theoretical perspectives." He goes on to explore the tension between respect for epistemological pluralism and the importance of maintaining intellectual standards, arguing for the importance of peer review and open intellectual discussion and argument within the profession. In this connection, Strike (2006, p. 61) notes the reference in the AERA Code preamble to educational researchers' "competing paradigms." He refers briefly to the complex history of the term "paradigm," and then draws out the implications of the phrase "competing paradigms," stating an extreme position which he then goes on to soften. If the term "paradigm" is used with precision, then "People with different paradigms can be expected to disagree not only about what is true concerning the phenomena of interest but also about such matters as what the boundaries of the phenomena are, how they are to be characterized, what is to count as evidence for claims about them, how evidence is to be collected, and what is to count as a good argument. Such disagreements run deep. … People with different paradigms see the world differently and lack common standards that allow them to resolve their differences." More optimistically, he then goes on to acknowledge that "the claim that paradigms are incommensurable is an overstatement," and offers philosophical support for his view that "conflicts between paradigms can be resolved by argument."

Additional perspectives on ethical concerns come from researchers in other methodological traditions. Cannella and Lincoln (2007), for instance, reflect on the risk of construing the persons studied as "other" and, in some methodologies, of presuming to give voice to their concerns. Pollock (2008) warns against uncritical use of conventional labels for groups defined, for example, by race or culture, and against simplistic generalizations essentializing other groups or their practices or behaviors. She admonishes that within the field of educational anthropology, researchers must stay close to the specifics of particular actions by particular individuals in particular contexts. Fine, Weis, Weseen, and Wong (2000) reflect on the often large difference in status between researcher and researched. Researchers have the power and perhaps the responsibility to

give voice to the concerns of marginalized groups, but this responsibility must be enacted with great care and sensitivity, and with due consideration to the ways research findings may later be taken out of context or misconstrued. Even the written documentation of informed consent may reinforce status differences and make it more difficult to establish trust.

Researchers have an affirmative responsibility to become familiar with the ethical codes of their professional organizations and the institutions where they are employed as well as those of institutions sponsoring research or affording the settings where research is carried out. However, ethics in research cannot be reduced to any set of codes or rules. As stated in the preamble to the AERA Code of Ethics, "Adhering to a set of ethical standards for an education researcher's work-related conduct requires a personal commitment to a lifelong effort to act ethically; to encourage ethical behavior by students, supervisors, supervisees, employers, employees, and colleagues; and to consult with others as needed concerning ethical problems. Each education researcher supplements, but does not violate, the values and rules specified in the ethical standards based on guidance drawn from personal values, culture, and experience."

REFERENCES

Abbott, A. D. (2001). *Chaos of disciplines*. Chicago: University of Chicago Press.

Abbott, A. D. (2004). *Methods of discovery: heuristics for the social sciences*. New York: W.W. Norton & Co.

Abend, G. (2008). The Meaning of 'Theory'. *Sociological Theory, 26*(2), 173-199. doi: 10.2307/20453103

Agar, M. (2004). We have met the other and we're all nonlinear: Ethnography as a nonlinear dynamic system. *COMPLEXITY, 10*(2), 16-24.

Agar, M. H. (1996). *The professional stranger: An informal introduction to ethnography* (Vol. 2). San Diego: Academic Press

Alastalo, M. (2008). The History of Social Research Methods. In P. Alasuutari, J. Brannen, & L. Bickman (Eds.), *The Sage Handbook of Social Research Methods* (pp. 26-41). Los Angeles: Sage Publications.

Alasuutari, P. (1996). Theorizing in Qualitative Research: A Cultural Studies Perspective. *Qualitative Inquiry, 2*(4), 371-384. doi: 10.1177/107780049600200401

American Educational Research Association. (2011). Code of Ethics. *Educational Researcher, 40*(3), 145-156. doi: DOI: 10.3102/0013189X11410403

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Educational Research Association (AERA). (2008). Definition of Scientifically Based Research. Retrieved June 27, 2009, from http://www.aera.net/ResearchPolicyAdvocacy/ResearchPolicyAdvocacyNews/DefinitionofScientifically-basedResearch/tabid/10277/Default.aspx

Amsterdamska, O. (2008). Practices, people, and places. In E. J. Hackett, O. Amsterdamska, M. Lynch, & J. Wajcman (Eds.), *The Handbook of Science and Technology Studies* (pp. 205-210). Cambridge, MA: MIT Press.

Arksey, H., & O'Malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology, 8*(1), 19-32. doi: 10.1080/1364557032000119616

Atkinson, P., Coffey, A., Delamont, S., Lofland, J., & Lofland, L. (2001). Editorial Introduction. In P. Atkinson, A. Coffey, S. Delamont, J. Lofland, & L. Lofland (Eds.), *Handbook of Ethnogrphy* (pp. 1-7). London, UK: Sage.

Axelrod, R., & Cohen, M. D. (2002). Harnessing Complxity. Walnut Creek, CA: InnovationLabs.

Axelrod, R., & Tesfatsion, L. (2006). Appendix A A Guide for Newcomers to Agent-Based Modeling in the Social Sciences. In L. Tesfatsion & K. L. Judd (Eds.), *Handbook of Computational Economics* (Vol. Volume 2, pp. 1647-1659). Amsterdam: Elsevier.

Bammer, G. (2013). *Disciplining Interdisciplinarity: Integration and Implementation Sciences for Researching Complex Real World Problems*. Canberra, AU: Australian National University.

Barley, S. R. (2008). Coalface Institutionalism. In R. Greenwood, C. Oliver, K. Sahlin, & R. Suddaby (Eds.), *The Sage Handbook of Organizational Institutionalism* (pp. 491-518). London, UK: Sage Publications.

Beach, K. (1999). Consequential Transitions: A Sociocultural Expedition Beyond Transfer in Education. *Review of Research in Education, 24*(1), 101-139. doi: 10.3102/0091732x024001101

Becker, H. S. (1998). *Tricks of the trade: how to think about your research while you're doing it*. Chicago, Ill.: University of Chicago Press.

Bergman, M. M. (2008). *Advances in mixed methods research: theories and applications*. London, UK: SAGE.

Bernstein, R. J. (1991). *The new constellation: the ethical-political horizons of modernity/postmodernity*. Cambridge, UK: Polity Press.

Biesta, G. (2007). Why "What Works" Won't Work Evidence-Based Practice and the Democratic Deficit in Educational Research. *Educational Theory, 57*(1), 1-22. doi: 10.1111/j.1741-5446.2006.00241.x

Biesta, G. (2010). Pragmatism and the Philosophical Foundations of Mixed Methods Research. In A. Tashakkori & C. Teddlie (Eds.), *Sage Handbook of Mixed Methods in Social & Behavioral Research* (pp. 95-118). Thousand Oaks, CA: Sage Publications.

Billett, S., Harteis, C., & Gruber, H. (2014). International Handbook of Research in Professional and Practice-based Learning. Dordrecht, Holland: Springer.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education, 5*(1), 7-73.

Boero, R., & Squazzoni, F. (2005). Does Empirical Embeddedness Matter? Methodological Issues on Agent-Based Models for Analytical Social Science. *Journal of artificial societies and social simulation, 8*(4), 1-31.

Bourdieu, P. (1991). The peculiar history of scientific reason. *Sociological Forum, 6*(1), 3-26. doi: 10.1007/bf01112725

Bourdieu, P., & Wacquant, L. J. D. (1992). *An invitation to reflexive sociology*. Chicago: University of Chicago Press.

Bowker, G. C. (2005). *Memory practices in the sciences*. Cambridge, Mass.: MIT Press.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: classification and its consequences*. Cambridge, Mass.: MIT Press.

Branch, M. N., & Pennypacker, H. S. (2013). Generality and generalization of research findings. In G. J. Madden (Ed.), *APA Handbook of Behavior Analysis: Vol. 1. Methods and Principles.* (pp. 151-175). Washington, DC: American Psychological Association.

Brannen, M. Y. (2011). Using multiple case studies to generalize from ethnographic research. In R. Piekkari & C. Welch (Eds.), *Rethinking the case study in international business and management research* (pp. 124-145). Cheltenham, UK: Edward Elgar Publishing Limited.

Bransford, J. D., Barron, B., Pea, R. D., Meltzoff, A., Kuhl, P., Bell, P., . . . Sabelli, N. H. (2006). Foundations and Opportunities for an Interdiscipoinary Science of Learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences*. Cambridge: Cambridge University Press.

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking Transfer: A Simple Proposal With Multiple Implications. *Review of Research in Education, 24*(1), 61-100. doi: 10.3102/0091732x024001061

Brennan, R. L. (Ed.). (2006). *Educational measurement* (4th ed.). Westport, CT: American Council on Education, Praeger Publishers.

Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-Analytic Methodology and Inferences About the Efficacy of Formative Assessment. *Educational Measurement: Issues and Practice, 31*(4), 13-17. doi: 10.1111/j.1745-3992.2012.00251.x

Bronfenbrenner, U. (1979). *The ecology of human development: experiments by nature and design*. Cambridge, Mass.: Harvard University Press.

Bronfenbrenner, U. (2004). *Making human beings human: bioecological perspectives on human development*. Thousand Oaks: Sage Publications.

Brown, J. S., & Duguid, P. (2000). *The social life of information*. Boston: Harvard Business School Press.

Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done? *Qualitative Research, 6*(1), 97-113. doi: 10.1177/1468794106058877

Bryman, A. (2008). The End of the Paradigm Wars? In P. Alasuutari, L. Bickman, & J. Brannen (Eds.), *The Sage Handbook of Social Research Methods* (pp. 13-25). London, UK: Sage Publications.

Bryman, A., Bell, E. A., & Teevan, J. J. (2009). *Social Research Methods*. Oxford, UK: Oxford University Press.

Buchanan, D. A., & Bryman, A. (Eds.). (2009). *The Sage Handbook of Organizational Research Methods*. Los Angeles: Sage Publications.

Byrne, D. (2009). Case-based methods: Why we need them; what they are; how to do them. In D. Byrne & C. Ragin (Eds.), *The SAGE handbook of case-based methods* (pp. 1-13). Los Angeles, CA: Sage.

Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation, 19*(3), 217-228. doi: 10.1177/1356389013495617

Byrne, D. S., & Callaghan, G. (2014). *Complexity theory and the social sciences: the state of the art*. Abingdon, Oxon ; New York: Routledge.

Byrne, D. S., & Ragin, C. C. (2009). *The SAGE handbook of case-based methods*. Los Angeles, Calif.: SAGE.

Camic, C., Gross, N., & Lamont, M. (2011). *Social knowledge in the making*. Chicago : London: University of Chicago Press.

Camic, C., & Joas, H. (2004). *The dialogical turn: new roles for sociology in the postdisciplinary age: essays in honor of Donald N. Levine*. Lanham, Md.: Rowman & Littlefield Publishers.

Campbell Collaboration Steering Group. (2014). Campbell Collaboration Systematic Reviews: Policies and Guidelines (Vol. 1.1). Oslo, Norway: Campbell Collaboration.

Campbell, D. T. (1975). III. "Degrees of Freedom" and the Case Study. *Comparative Political Studies, 8*(2), 178-193. doi: 10.1177/001041407500800204

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull, 56*(2), 81-105. doi: 10.1037/h0046016

Campbell, D. T., & Stanley, J. C. (1963). Experimental and Quasi-Experimental Designs for Research on Teaching. In N. L. Gage (Ed.), *Handbook of Research on Teaching* (1st ed., pp. 171-246). Chicago: Rand McNally & Company.

Campbell, D. T., Stanley, J. C., & Gage, N. L. (1966). *Experimental and quasi-experimental designs for research*. Chicago: R. McNally.

Cannella, G. S., & Lincoln, Y. S. (2007). Predatory vs. dialogic ethics: Constructing an illusion or ethical practice as the core of research methods. *Qualitative Inquiry, 13*(3), 315-335.

Carr-Hill, R., Rolleston, C., Pherali, T., & Schendel, R. (2014). Title Registration for Systematic Review: The Effects of School-Based Decision Making on Educational Outcomes in Low- and Middle-Income Contexts: A Systematic Review. Oslo, Norway: Campbell Collaboration.

Carrington, P. J., Scott, J., & Wasserman, S. (2005). *Models and Methods in Social Network Analysis*. Cambridge, UK: Cambridge University Presss.

Chalmers, I. (2012). A reality check for systematic reviewers. Retrieved from https://editorial-unit.cochrane.org/blog/guest-blog-sir-iain-chalmers-reality-check-systematic-reviewers

Charmaz, K. (2006). *Constructing grounded theory: a practical guide through qualitative analysis*. London ; Thousand Oaks: SAGE.

Charmaz, K. (2014). *Constructing grounded theory*. Los Angeles, CA: Sage.

Christie, F. (1994). On Pedagogic Discourse: Final Report of a Research Activity Funded by the ARC 1990-2. Melbourne, AU: University of Melbourne.

Clarke, K. A., & Primo, D. M. (2012). *A model discipline: political science and the logic of representations*. Oxford ; New York: Oxford University Press.

Cobb, P. (2007). Putting Philosophy to Work. In F. K. J. Lester (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning* (pp. 3-38). Washington, DC: National Council of Teachers of Mathematics.

Cobb, P., Jackson, K., Smith, T., Sorum, M., & Henrick, E. (2013). Design research with educational systems: Investigating and supporting improvements in the quality of mathematics teaching and learning at scale. *National Society for the Study of Education Yearbook, 112*(2), 320-349.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning About Statistical Covariation. *Cognition and Instruction, 21*(1), 1-78. doi: 10.1207/S1532690XCI2101_1

Cobb, P., Zhao, Q., & Dean, C. (2009). Conducting Design Experiments to Support Teachers' Learning: A Reflection From the Field. *Journal of the Learning Sciences, 18*(2), 165-199. doi: 10.1080/10508400902797933

Coburn, C. E., Russell, J. L., Kaufman, J. H., & Stein, M. K. (2012). Supporting Sustainability: Teachers' Advice Networks and Ambitious Instructional Reform. *American Journal of Education, 119*(1), 137-182. doi: 10.1086/667699

Coburn, C. E., & Stein, M. K. (2010). *Research and practice in education: Building alliances, bridging the divide*: Rowman & Littlefield Publishers.

Coburn, C. E., & Talbert, J. E. (2006). Conceptions of Evidence Use in School Districts: Mapping the Terrain. *American Journal of Education, 112*(4), 469-495.

Coburn, C. E., & Turner, E. O. (2011). Research on Data Use: A Framework and Analysis. *Measurement: Interdisciplinary Research & Perspective, 9*(4), 173-206. doi: 10.1080/15366367.2011.626729

Cochran-Smith, M., Ell, F., Ludlow, L., Grudnoff, L., & Aitken, G. (2014). The Challenge and Promise of Complexity Theory for Teacher Education Research. *Teachers College Record, 116*(5), 1-38.

Colyvas, J. A. (2012). Performance Metrics as Formal Structures and through the Lens of Social Mechanisms: When Do They Work and How Do They Influence? *American Journal of Education, 118*(2), 167-197.

Conrad, D., & Kendal, W. (2009). Making space for youth: iHuman youth society and arts-based participatory research with street-involved youth in Canada. In D. Kapoor & S. Jordan (Eds.), *Education, participatory action research, and social change* (pp. 251-264). New York: Palgrave Macmillan.

Cook, S. D. N., & Brown, J. S. (1999). Bridging Epistemologies: The Generative Dance Between Organizational Knowledge and Organizaitonal Knowing. *ORGANIZATION SCIENCE, 10*(4), 381-400.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and Analysis Issus for Field Settings*. Boston: Houghton Mifflin.

Cook, T. D., Reichardt, C. S., & Evaluation Research, S. (1979). *Qualitative and quantitative methods in evaluation research*. Beverly Hills, Calif.: Sage Publications.

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

Cooper, H. M. (2010). *Research synthesis and meta-analysis: a step-by-step approach*. Los Angeles: Sage.

Corbin, J. M., & Strauss, A. L. (2008). *Basics of qualitative research: techniques and procedures for developing grounded theory*. Los Angeles: SAGE Publications.

Crawford, J., & Impara, J. C. (2001). Critical Issues, Current Trends, and Possible Futures in Quantitative Methods. In V. Richardson (Ed.), *Handbook of Research on Teaching* (4th ed., pp. 133-173). Washington, DC: American Educational Research Association.

Creswell, J., Clark, V., Gutmann, M., & Hanson, W. (2003). Advanced Mixed Methods Research Designs In Tashakkori, A., & Teddle, C.,(Eds.) Handbook of Mixed Methods in Social & Behavioral Research (pp. 209-240): Sage: Thousand Oaks, California.

Creswell, J. W. (2014). *Research design: qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, Calif.: SAGE Publications.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30*(2), 116-127. doi: 10.1037/h0076829

Daly, A. J. (2012). Data, Dyads, and Dynamics: Exploring Data Use and Social Networks in Educational Improvement. *Teachers College Record, 114*(11), 1.

Davis, B., & Sumara, D. J. (2005). Challenging Images of Knowing: Complexity Science and Educational Research. *International Journal of Qualitative Studies in Education, 18*(3), 305-321.

Delgado, R., & Stefancic, J. (2001). *Critical race theory: an introduction.* New York: NYU Press.

Denscombe, M. (2008). Communities of Practice: A Research Paradigm for the Mixed Methods Approach. *Journal of Mixed Methods Research, 2*(3), 270-283. doi: 10.1177/1558689808316807

Denzin, N. K. (1978). Triangulation: A case for methodological evaluation and combination. *Sociological methods*, 339-357.

Design Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher, 32*(1), 5-8. doi: 10.3102/0013189x032001005

Easterby-Smith, M., & Lyles, M. A. (2011). *Handbook of organizational learning and knowledge management*. Chichester, U.K.: Wiley.

Easton, J. Q. (2010). Director's Final Proposed Priorities. Washington, DC.: U.S. Department of Education, Institute of Education Sciences.

Eisenhart, M. (2001). Changing Conceptions of Culture and Ethnographic Methodology: Recent Thematic Shifts and Their Implications for Research on Teaching. In V. Richardson (Ed.), *Handbook of Research on Teaching* (Fourth ed., pp. 209-225). Washington, DC: American Educational Research Association.

Eisenhart, M. (2009). Generalization from qualitative inquiry. I. In K. D. Ercikan & W.-M. Roth (Eds.), *Generalizing from educational research.* New York: Routledge.

Eisenhart, M. A. (1988). The Ethnographic Research Tradition and Mathematics Education Research. *Journal for Research in Mathematics Education, 19*(2), 99-114.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). *Writing ethnographic fieldnotes*. Chicago: University of Chicago Press.

Ercikan, K., & Roth, W.-M. (2009). *Generalizing from educational research: beyond qualitative and quantitative polarization*. New York: Routledge.

Ercikan, K., & Roth, W.-M. (2014). Limits of generalizing in education research: Why criteria for research generalization should include population heterogeneity and users of knowledge claims. *Teach. Coll. Rec, 116*.

Erickson, F. (1986). Qualitative Methods in Research on Teaching. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (Third ed., pp. 119-161). Washington, DC: American Educational Research Association.

Erickson, F. (1998). Qualitative Research Methods for Science Education (Part Two). In B. J. Fraser & K. G. Tobin (Eds.), *International Handbook of Science Education* (pp. 1155-1173). Dordrecht: Kluwer Academic Publishers.

Erickson, F. (2011). A history of qualitative inquiry in social and educational research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage Handbook of Qualitative Research* (pp. 43-60). Los Angeles: Sage Publications, Inc.

Fenwick, T. J., & Edwards, R. (2010). *Actor-network theory in education*. London ; New York: Routledge.

Fine, M., Weis, L., Weseen, S., & Wong, L. (2000). For whom? Qualitative research, representations, and social responsibilities. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (Vol. 2, pp. 107-131). Thousand Oaks: Sage Publications.

Fishman, B. J., Penuel, W. R., Allen, A.-R., Cheng, B. H., & Sabelli, N. (2013). Design-based implementation research: An emerging model for transforming the relationship of research and practice. *Design-based implementation research:*

Theories, methods, and exemplars. National Society for the Study of Education Yearbook, 112(2), 136-156.

Flyvbjerg, B. (2001). *Making social science matter: why social inquiry fails and how it can succeed again*. Cambridge, UK: Cambridge University Press.

Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry, 12*(2), 219-245. doi: 10.1177/1077800405284363

Flyvbjerg, B., Landman, T., & Schram, S. (2012). *Real social science: Applied phronesis*. Cambridge, UK: Cambridge University Press.

Frank, K. A., & Yasumoto, J. Y. (1998). Linking Action to Social Structure within a System: Social Capital within and between Subgroups 1. *American Journal of Sociology, 104*(3), 642-686.

Frodeman, R. (2011). Interdisciplinary research and academic sustainability: managing knowledge in an age of accountability. *Environmental Conservation, 38*(02), 105-112.

Frodeman, R., Klein, J. T., & Mitcham, C. (2010). The Oxford handbook of interdisciplinarity.

Gage, N. L. (1989). The paradigm wars and their aftermath: A "historical" sketch of research on teaching since 1989. *Educational Researcher, 18*(4), 4-19.

Gage, N. L., & American Educational Research, A. (1963). *Handbook of research on teaching; a project of the American Educational Research Association*. Chicago: Rand McNally.

Gee, J. P. (2008). A Sociocultural Perspective on Opportunity to Learn. In P. A. Moss, D. C. Pullin, J. P. Gee, E. H. Haerel, & L. J. Young (Eds.), *Assessment, Equity, and Opportunity to Learn* (pp. 42-75). Cambridge, UK: Cambridge University Press.

Gee, J. P. (2011). *An introduction to discourse analysis: theory and method*: Routledge.

Geertz, C. (1973). *The interpretation of cultures selected essays*. New York: Basic Books.

Geertz, C. (1983). *Local knowledge: further essays in interpretive anthropology*. New York: Basic Books.

George, A. L., & Bennett, A. (2005). *Case studies and theory development in the social sciences*. Cambridge, Mass.: MIT Press.

Gibbons, M., & Nowotny, H. (2001). The potential of transdisciplinarity. In J. T. Klein & et al. (Eds.), *Transdisciplinarity: joint problem solving among science, technology, and society* (pp. 67-80). Basel: Springer.

Gieryn, T. F. (1999). *Cultural boundaries of science: credibility on the line*. Chicago: University of Chicago Press.

Gillborn, D. (2013). The policy of inequity: using CRT to unmask white supremacy in education policy. In M. Lynn & A. D. Dixson (Eds.), *Handbook of Critical Race Theory in Education* (pp. 129-139). New York: Routledge.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory; strategies for qualitative research*. Chicago: Aldine Pub. Co.

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

Gobo, G. (2008). Re-Conceptualizing Generalization: Old Issues in New Frame. In P. Alasuutari, L. Bickman, & J. Brannen (Eds.), *The Sage Handbook of Social Research Methods* (pp. 193-213). London, UK: Sage Publications.

Goeken, M., & Borner, R. (2012). Generalization in Qualitative IS Research - Approaches and their Application to a Case Study on SOA Development. *Australasian Journal Of Information Systems,, 17*(2). doi: doi:10.3127/ajis.v17i2.685

Gomm, R., Hammersley, M., & Foster, P. (2000a). Case Study and Generalization. In R. Gomm, M. Hammersley, & P. Foster (Eds.), *Case Study Method* (pp. 98-116). London, UK: Sage Publications.

Gomm, R., Hammersley, M., & Foster, P. (2000b). *Case study method key issues, key texts*. London, UK: Sage.

Goodwin, C. (1994). Professional Vision. *American Anthropologist, 96*(3), 606-633.

Gough, D., Oliver, S., & Thomas, J. (2012). *An introduction to systematic reviews*. London ; Thousand Oaks, Calif.: Sage.

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey-Bass.

Greene, J. C. (2012). Engaging Critical Issues in Social Inquiry by Mixing Methods. *American Behavioral Scientist, 56*(6), 755-773. doi: 10.1177/0002764211433794

Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed-method evaluation. *New Directions for Evaluation, 1997*(74), 5-17. doi: 10.1002/ev.1068

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274. doi: 10.3102/01623737011003255

Greenhalgh, T., Robert, G., Bate, P., Macfarlane, F., & Kyriakidou, O. (2005). *Diffusion of innovations in health service organisations: a systematic literature review*. Malden, Mass.: BMJ Books/Blackwell Pub.

Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., Kyriakidou, O., & Peacock, R. (2005). Storylines of research in diffusion of innovation: a meta-narrative approach to systematic review. *Soc Sci Med, 61*(2), 417-430. doi: 10.1016/j.socscimed.2004.12.001

Greenleaf, C. L., Litman, C., Hanson, T. L., Rosen, R., Boscardin, C. K., Herman, J., . . . Jones, B. (2011). Integrating Literacy and Science in Biology: Teaching and Learning Impacts of Reading Apprenticeship Professional Development. *American Educational Research Journal, 48*(3), 647-717. doi: 10.3102/0002831210384839

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record, 111*(9).

Groves, R., Fowler Jr, F., Couper, M., Lepkowski, J., Singer, E., & Tourangeau, R. (2009). Survey methodology (2nd): New York: John Wiley and Sons.

Guile, D. (2011). Interprofessional Learning: Reasons, Judgment, and Action. *Mind, Culture, and Activity, 18*(4), 342-364.

Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to Practice as a Criterion for Rigor. *Educational Researcher, 43*(1), 19-23. doi: 10.3102/0013189x13520289

Hackett, E. J., Amsterdamska, O., Lynch, M., & Wajcman, J. (2008). *The handbook of science and technology studies*. Cambridge, Mass.: MIT Press

Hale, C. R. (2006). Activist research v. cultural critique: Indigenous land rights and the contradictions of politically engaged anthropology. *Cultural Anthropology, 21*(1), 96-120.

Halliday, M. A., & Hasan, R. (1989). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Oxford, UK: Oxford University Press.

Halliday, M. A., & Matthiessen, C. (2004). *Introducing functional grammar*. New York: Edward Arnold.

Halliday, M. A. K. (1994). *An introduction to functional grammar*. London: E. Arnold.

Halverson, R., Grigg, J., Prichett, R., & Thomas, C. (2007). The new instructional leadership: Creating data-driven instructional systems in school. *Journal of School Leadership, 17*(2), 159-164.

Hammersley, M. (2004). Analytic Induction. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), *Encyclopedia of Social Science Research Methods* (pp. 17-19). Thousand Oaks: Sage.

Hammersley, M., & Atkinson, P. (2007). *Ethnography: principles in practice*. London, UK: Routledge.

Hammersley, M., Gomm, R., & Foster, P. (2000). Case Study and Theory. In R. Gomm, M. Hammersley, & P. Foster (Eds.), *Case Study Method* (pp. 234-258). London, UK: Sage Publications.

Harding, S. G. (1987). *Feminism and methodology: social science issues*. Bloomington: Indiana University Press.

Hedström, P., & Swedberg, R. (1998). *Social mechanisms: an analytical approach to social theory*. Cambridge : New York: Cambridge University Press.

Hedström, P., & Ylikoski, P. (2010). Causal Mechanisms in the Social Sciences. *Annual Review of Sociology, 36*(1), 49-67. doi: doi:10.1146/annurev.soc.012809.102632

Henig, J. R. (2012). The Politics of Data Use. *Teachers College Record, 114*(11), 1-32.

Hirsch Hadorn, G., Biber-Klemm, S., Grossenbacher-Mansuy, W., Hoffmann-Riem, H., Joye, D., Pohl, C., . . . SpringerLink. (2008). *Handbook of Transdisciplinary Research*. Dordrecht: Springer Science + Business Media B.V.

Honig, M. I., & Venkateswaran, N. (2012). School-Central Office Relationships in Evidence Use: Understanding Evidence Use as a Systems Problem. *American Journal of Education, 118*(2), 199-222.

Horn, I. S. (2007). Fast Kids, Slow Kids, Lazy Kids: Framing the Mismatch Problem in Mathematics Teachers' Conversations. *Journal of the Learning Sciences, 16*(1), 37-79. doi: 10.1080/10508400709336942

Howard, T. (2008). Who really cares? The disenfranchisement of African American males in preK-12 schools: A critical race theory perspective. *The Teachers College Record, 110*(5), 954-985.

Howe, K. R. (1988). Against the Quantitative-Qualitative Incompatibility Thesis or Dogmas Die Hard. *Educational Researcher, 17*(8), 10-16. doi: 10.2307/1175845

Hubbard, L., Mehan, H., & Stein, M. K. (2006). *Reform as learning: school reform, organizational culture, and community politics in San Diego*. New York: Routledge.

Huisman, M., & Van Dujin, M. A. J. (2005). Software for Social Network Analysis. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and Methods in Social Network Analysis* (pp. 270-316). Cambridge, UK: Cambridge University Press.

Hunter, A., & Brewer, J. (2003). Multimethod Research in Sociology. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of Mixed Methods Research in Social and Behavioral Sciences* (pp. 577-594). Thousand Oaks, CA: Sage Publications.

Institute of Education Science, & National Science Foundation. (2013). *Common Guidelines for Education Research and Development*. Washington, DC: US Department of Education.

Jang, E. E., McDougall, D. E., Pollon, D., Herbert, M., & Russell, P. (2008). Integrative mixed methods data analytic strategies in research on school success in challenging circumstances. *Journal of Mixed Methods Research, 2*(3), 221-247.

Jarvie, I. C. (2011). Philosophical Problems of the Social Sciences: Paradigms, Methodology, and Ontology. In I. C. Jarvie & J. Samora-Bonilla (Eds.), *The Sage Handbook of The Philosophy of Social Sciences*. Los Angeles: Sage.

Jennings, J. L. (2012). The Effects of Accountability System Design on Teachers' Use of Test Score Data. *Teachers College Record, 114*(11), 1-23.

Jensen, K., Lahn, L. C., Nerland, M., & ebrary, I. (2012). *Professional learning in the knowledge society*. Rotterdam: Sense Publishers.

Johnson, B., & Christensen, L. (2008). *Educational research: Quantitative, qualitative, and mixed approaches*. Los Angeles, CA: Sage.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher, 33*(7), 14-26. doi: 10.3102/0013189x033007014

Jordan, S. (2009). From a methodology of the margins to neoliberal appropriation and beyond: The lineages of PAR. In D. Kapoor & S. Jordan (Eds.), *Education, Participtory Action Research, and Social Change: International Perspectives* (pp. 15-27). New York: Palgrave Macmillan.

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement (4th Ed.)* (4th ed., pp. 17-64). Westport, CT: American Council on Education and Praeger Publishers.

Kapoor, D., & Jordan, S. (2009). *Education, participatory action research, and social change: International perspectives*. New York: Palgrave Macmillan.

Kapoor, D., & Jordan, S. (2009). Introduction: International perspectives on education, PAR, and social change *Education, Participatory Action Research and Social Change: International Perspectives* (pp. 1-11). New York: Palgrave Macmillan.

Kingston, N., & Nash, B. (2011). Formative Assessment: A Meta-Analysis and a Call for Research. *Educational Measurement: Issues and Practice, 30*(4), 28-37. doi: 10.1111/j.1745-3992.2011.00220.x

Kingston, N., & Nash, B. (2012). How Many Formative Assessment Angels Can Dance on the Head of a Meta Analytic Pin:. 2. *Educational Measurement: Issues and Practice, 31*(4), 18-19.

Klein, J. T. (2007). Interdisciplinary Approaches in Social Science Research. In W. Outhwaite & S. P. Turner (Eds.), *The Sage handbook of social science methodology* (pp. 32-50). London, UK: Sage Publications.

Klein, J. T. (2010). A Taxonomy of Interdisciplinarity. In R. Frodeman, J. Thompson Klein, & C. Mitcham (Eds.), *The Oxford Handbook of Interdisciplinarity* (pp. 15-30). Oxford, UK: Oxford University Press.

Knorr Cetina, K. (1999). *Epistemic cultures: How scientists make sense*. Chicago: Indiana University Press.

Knorr Cetina, K. (2006). Knowledge in a knowledge society: Five transitions. *Knowledge, Work and Society, 4*(3), 23-41.

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Ladson-Billings, G. (1998). Just what is critical race theory and what's it doing in a nice field like education? *International Journal of Qualitative Studies in Education, 11*(1), 7-24.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational Researcher, 35*(7), 3-12.

Ladson-Billings, G. (2013). Critical race theory: what it is not! . In M. Lynn & A. D. Dixson (Eds.), *Handbook of Critical Race Theory* (pp. 34-47). New York: Routledge.

Ladson-Billings, G., & Tate, W. (1995). Towards a critical race theory of education. *Teachers College Record, 97*(1), 48-68.

Lamont, M., & Molnár, V. (2002). The Study of Boundaries in the Social Sciences. *Annual Review of Sociology, 28*(1), 167-195. doi: doi:10.1146/annurev.soc.28.110601.141107

Lampert, M. (2001). *Teaching problems and the problems of teaching*. New Haven: Yale University Press.

Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford ; New York: Oxford University Press.

Lather, P. (1986). Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange, 17*(4), 63-84.

Lather, P. (2006). Paradigm proliferation as a good thing to think with: teaching research in education as a wild profusion. *International Journal of Qualitative Studies in Education, 19*(1), 35-57. doi: 10.1080/09518390500450144

Lather, P. (2007). *Getting lost: feminist efforts toward a double(d) science*. Albany: State University of New York Press.

Lather, P., & Smithies, C. (1997). *Troubling the angels: women living with HIV/AIDS*: Westview Press.

Latour, B. (2005). *Reassembling the social: an introduction to actor-network-theory*. New York: Oxford University Press.

Lattuca, L. R. (2001). *Creating Interdisciplinarity: Interdisciplinary Research and Teaching Among College and University Faculty*. Nashville, TN: Vanderbilt University Press.

Law, J. (2004). *After method: mess in social science research*. London: Routledge.

Lawrence, R. J., & Després, C. (2004). Futures of Transdisciplinarity. *Futures, 36*(4), 397-405. doi: http://dx.doi.org/10.1016/j.futures.2003.10.005

Lemke, J. L., & Sabelli, N. H. (2008). Complex systems and educational change: Towards a new research. In M. Mason (Ed.), *Complexity theory and the philosophy of education* (pp. 112-133). Malden, MA: Wiley-Blackwell, 2008.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, Calif.: Sage Publications.

Linn, R. L. (1986). Quantitative Methods in Research on Teaching. In M. C. Wittrock (Ed.), *Handbook of Research on Teaching* (3rd ed., pp. 92-118). New York: MacMillan.

Little, J. W. (2012). Understanding Data Use Practice among Teachers: The Contribution of Micro-Process Studies. *American Journal of Education, 118*(2), 143-166.

Longino, H. (2013). The Social Dimensions of Scientific Knowledge. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Srping 2013 Edition). Retrieved from <http://plato.stanford.edu/archives/spr2013/entries/scientific-knowledge-social/%3E.

Lynn, M., & Parker, L. (2006). Critical race studies in education: Examining a decade of research on US schools. *The Urban Review, 38*(4), 257-290.

Maldonato, M., & Pietrobon, R. (2010). *Research on scientific research: a transdisciplinary study*. Brighton [England] ; Portland, Or.: Sussex Academic Press.

Malpass, A., Shaw, A., Sharp, D., Walter, F., Feder, G., Ridd, M., & Kessler, D. (2009). "Medication career" or "Moral career"? The two sides of managing antidepressants: A meta-ethnography of patients' experience of antidepressants. *Social Science & Medicine, 68*(1), 154-168.

Maroulis, S. (2009). *Agents, Networks, and Markets: Three Essays from a Complex Systems Perspective on Educational Change.* Northwestern University, Evanston, IL.

Maroulis, S., Bakshy, E., Gomez, L. M., & Wilensky, U. (2010). An Agent-Based Model of Intra-District Public School Choice. Evanston, IL: Northwestern Unviersity.

Maroulis, S., Guimera, R., Petry, H., Stringer, M. J., Gomez, L. M., Amaral, L. A. N., & Wilensky, U. (2010). Complex Systems View of Educational Policy Research. *SCIENCE, 330*(6000), 38-39.

Marsden, P. V. (2005). Recent Developments in Network Measurement. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and Methods in Social Network Analysis* (pp. 8-29). Cambridge, UK: Cambridge University Press.

Marsh, J. A. (2012). Interventions promoting educators' use of data: Research insights and gaps. *Teachers College Record, 114*(11).

Martin, J. L. (2011). *The explanation of social action*. Oxford, UK: Oxford University Press.

Martin, J. R. (1992). *English text: system and structure*. Philadelphia: John Benjamins Pub. Co.

Martin, J. R. (1993). Genre and literacy - modeling context in educational linguistics. *Annual Review of Applied Linguistics, 13*, 141-172.

Martin, M., & McIntyre, L. C. (1994). *Readings in the philosophy of social science*. Cambridge, Mass.: MIT Press.

Mason, M., & Wiley online, l. (2008). *Complexity theory and the philosophy of education*. Chichester, West Sussex, U.K. ; Malden, MA: Wiley-Blackwell.

Maxwell, J. A., & Mittapalli, K. (2008). Theory. In L. M. Given (Ed.), *The SAGE Encyclopedia of Qualitative Research Methods* (pp. 877-881). Thousand Oaks: SAGE Publications, Inc.

McFarland, D. A., Moody, J., Diehl, D., Smith, J. A., & Thomas, R. J. (2014). Network Ecology and Adolescent Social Structure. *American Sociological Review*, 0003122414554001.

Merton, R. K. (1967). *On theoretical sociology: Five essays, old and new*: Free Press New York.

Mhina, C. H. (2009). Research and agency: the case of rural women and land tenure in Tanzania. . In D. Kapoor & S. Jordan (Eds.), *Education, participatory action research, and social change* (pp. 155-168). New York: Palgrave Macmillan.

Miller, J. H., & Page, S. E. (2004). The standing ovation problem. *COMPLEXITY, 9*(5), 8-16. doi: 10.1002/cplx.20033

Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: an introduction to computational models of social life*. Princeton, N.J.: Princeton University Press.

Mitchell, J. P. (2007). Ethnography. In W. Outhwaite & S. P. Turner (Eds.), *The Sage handbook of social science methodology* (pp. 55-66). London, UK: Sage.

Mitchell, M. (2009). *Complexity: a guided tour*. Oxford ; New York: Oxford University Press.

Mjøset, L. (2009a). The contextualist approach to social science methodology. *The Sage handbook of case-based methods, London: Sage*, 39-69.

Mjøset, L. (2009b). The contextualist approach to social science methodology. In D. Byrne & C. Ragin (Eds.), *The Sage handbook of case-based methods, London: Sage* (pp. 39-69). Los Angeles, CA: Sage.

Montes, G. (2012). Using artificial societies to understand the impact o teacher student match on academic performance: The case of same race effects. *Journal of artificial societies and social simulation, 15*(4).

Morçöl, G. (2012). *Complexity Theory for Public Policy*. New York, NY: Routledge.

Moreno, J. L. (1953). *Who shall survive?: foundationsof sociometry, group psychotherapy, and sociodrama* (2nd ed.). Beacon, NY: Beacon House.

Morgan, D. L. (2007). Paradigms Lost and Pragmatism Regained: Methodological Implications of Combining Qualitative and Quantitative Methods. *Journal of Mixed Methods Research, 1*(1), 48-76. doi: 10.1177/2345678906292462

Morin, E. (2008). *On Complexity*. Cresskill, NJ: Hampton Press.

Morrison, K. (2002). *School Leadership and Complexity Theory*. New York, NY: Routledge.

Morse, J. M. (2003). Principles of Mixed Methods and Multimethod Research Design *Handbook of Mixed Methods in Social & Behavioral Research* (pp. 189-208). Thousand Oaks, CA: Sage Publications.

Moss, P. A. (2005). Understanding the other/understanding ourselves: Toward a constructive dialogue about "principles" in educational research. *Educational Theory, 55*(3), 263-283.

Moss, P. A. (2008). Sociocultural implications for the practice of assessment I: Classroom assessment. . In P. A. Moss, D. Pullin, J. P. Gee, E. H. Haertel, & L. J. Yough (Eds.), *Assessment, Equity, and Opportunity to Learn*. New York: Cambridge University Press.

Moss, P. A. (2014). Critiques of Education Research. In D. C. Phillips (Ed.), *Encyclopedia of Education Theory and Philosophy*. Thousand Oaks, CA: Sage.

Moss, P. A., Girard, B. J., & Haniford, L. C. (2006). Validity in Educational Assessment. *Review of Research in Education, 30*(1), 109-162. doi: 10.3102/0091732x030001109

Moss, P. A., Phillips, D. C., Erickson, F. D., Floden, R. E., Lather, P. A., & Schneider, B. L. (2009). Learning from Our Differences: A Dialogue across Perspectives on Quality in Education Research. *Educational Researcher, 38*(7), 501-517.

Moss, P. A., Pullin, D. C., Gee, J. P., Haertel, E. H., & Young, L. J. (2008). *Assessment, equity, and opportunity to learn*. Cambridge: Cambridge University Press.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: improving causal inference in educational and social science research*. Oxford ; New York: Oxford University Press.

Nathan, M., & Alibali, M. W. (2010). Learning sciences. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(3), 329-345.

Nathan, M. J., & Sawyer, R. K. (2014). Foundations of the Learning Sciences. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed.). New York: Cambridge University Press.

National Academy of Education. (1999). Recommendations Regarding Research Priorities. Washington, DC: National Academy of Education.

National Board for Education Sciences (NBES). (2012). 2012 NBES Annual Report. Washington, DC: Institute of Education Sciences, US Department of Education.

National Commission for the Proptection of Human Subjects of Biomedical Behavioral Research. (1978). *The Belmont report: Ethical principles and guidelines for the protection of human subjects of research*: ERIC Clearinghouse.

National Research Council. (1999). *Improving Student Learning:A Strategic Plan for Education Research and Its Utilization*: The National Academies Press.

National Research Council. (2002). *Scientific Research in Education*. Washington, DC: The National Academies Press.

National Research Council. (2004a). *Advancing Scientific Research in Education*. Washington, DC: The National Academies Press.

National Research Council. (2004b). *Facilitating Interdisciplinary Research*. Washington, DC: The National Academies Press.

National Research Council. (2012). *Using Science as Evidence in Public Policy*. Washington, DC: The National Academies Press.

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: synthesizing qualitative studies*. Newbury Park: Sage Publications.

Nowotny, H., Scott, P., & Gibbons, M. (2001). *Re-thinking science: knowledge and the public in an age of uncertainty*. Cambridge : Malden, MA: Polity ; Published in the USA by Blackwell.

Nutley, S. M., Walter, I., & Davies, H. T. (2007). *Using evidence: How research can inform public services*: Policy Press.

O'Halloran, K. L. (2004). Discourses in secondary school mathematics classrooms according to social class and gender. *Language, education and discourse: Functional approaches*, 191-225.

OECD. (1972). Problems of Teaching and Research in Universities. *Paris: Organization for Economic Cooperation and Development, 1*, 972.

OECD. (2014). TALIS 2013 Results: An International Perspective on Teaching and Learning. Paris, France: OECD

Olejnik, S., & Algina, J. (2000). Measures of Effect Size for Comparative Studies: Applications, Interpretations, and Limitations. *Contemporary Educational Psychology, 25*(3), 241-286.

Opfer, V. D., & Pedder, D. (2011). Conceptualizing Teacher Professional Learning. *Review of Educational Research, 81*(3), 376-407. doi: 10.3102/0034654311413609

Osberg, D., & Biesta, G. (2010). *Complexity theory and the politics of education*. Rotterdam ; Boston: Sense Publishers.

Outhwaite, W., & Turner, S. P. (2007). General Introduction. In W. Outhwaite & S. P. Turner (Eds.), *The Sage handbook of social science methodology* (pp. 1-2). London, UK: Sage.

Pachucki, M., Pendergrass, S., & Lamont, M. (2007). Boundary processes: recent theoretical developments and new contributions. *Poetics, 35*, 331.

Page, S. E. (2010). Complexity in Social, political and Economic Systems.: SBE 2020 7 White Paper: National Science Foundation.

Page, S. E. (2011). *Diversity and complexity*. Princeton, N.J.: Princeton University Press.

Pascale, C.-M. (2011). *Cartographies of knowledge: exploring qualitative epistemologies*. Los Angeles: SAGE.

Pawson, R. (2004). Mentoring Relationships: An Explanatory Review. Leeds, UK: University of Leads.

Pawson, R. (2006). *Evidence-based policy a realist perspective*. London: SAGE.

Pawson, R. (2013). *The science of evaluation: a realist manifesto*. London ; Thousand Oaks, Calif.: SAGE.

Penuel, W., Riel, M., Krause, A., & Frank, K. (2009). Analyzing teachers' professional interactions in a school as social capital: A social network approach. *The Teachers College Record, 111*(1), 124-163.

Penuel, W. R., Allen, A.-R., Coburn, C. E., & Farrell, C. (2015). Conceptualizing Research–Practice Partnerships as Joint Work at Boundaries. *Journal of Education for Students Placed at Risk (JESPAR), 20*(1-2), 182-197. doi: 10.1080/10824669.2014.988334

Penuel, W. R., & Spillane, J. P. (2014). Learning sciences and policy design and implmenetation: Key concepts and tools for collaborative engagement. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed.). New York: Cambridge University Press.

Phillips, D. (2006). A guide for the perplexed: Scientific educational research, methodolatry, and the gold versus platinum standards. *Educational Research Review, 1*(1), 15-26.

Phillips, D. C. (1983). After the wake: Postpositivistic educational thought. *Educational Researcher*, 4-12.

Phillips, D. C., & Burbules, N. C. (2000). *Postpositivism and educational research*. Lanham, Md.: Rowman & Littlefield Publishers.

Pohl, C. (2011). What is progress in transdisciplinary research? *Futures, 43*(6), 618-626.

Pohl, C., & Hadorn, G. H. (2008). Core terms in transdisciplinary research. In G. Hirsch Hadorn, S. Biber-Klemm, W. Grossenbacher-Mansuy, H. Hoffmann-Riem, D.

Joye, C. Pohl, U. Wiesmann, & E. Zemp (Eds.), *Handbook of transdisciplinary research* (pp. 427-432): Springer.

Pollock, M. (2008). From shallow to deep: Toward a thorough cultural analysis of school achievement patterns. *Anthropology & Education Quarterly, 39*(4), 369-380.

Pólya, G. (1957). *How to solve it; a new aspect of mathematical method*. Garden City, N.Y.: Doubleday.

Pope, C., Mays, N., Popay, J., & ebrary, I. (2007). *Synthesizing qualitative and quantitative health evidence a guide to methods*. Maidenhead, England ; New York, NY: Open University Press, McGraw Hill Education.

Powell, W. W., & Colyvas, J. A. (2008). Microfoundations of Institutional Theory. In R. Greenwood, C. Oliver, K. Sahlin, & R. Suddaby (Eds.), *The Sage Handbook of Organizational Institutionalism* (pp. 276-298): Sage

Rabinow, P., & Sullivan, W. M. (1987). *Interpretive social science: a second look*. Berkeley: University of California Press.

Ragin, C., & Rihoux, B. (2004). Qualitative Comparative Analysis (QCA): State of the Art and Prospects. *Qualitative Methods: Newsletter of the American Political Science Association Association Organized SEction on Qualitative Methods, 2*(2), 3-13.

Ragin, C. C. (1987). *The comparative method: moving beyond qualitiative and quantitative strategies*. Berkeley: University of California Press.

Ragin, C. C. (1992). Introduction: Cases of "what is a Case?". In C. C. Ragin & H. S. Becker (Eds.), *What is a Case? Exploring the Foundations of Social Inquiry* (pp. 1-18). Cambridge, UK: Cambridge University Press.

Ragin, C. C. (2000). *Fuzzy-set social science*. Chicago: University of Chicago Press.

Ragin, C. C. (2007). Comparative Methods. In W. Outhwaite & S. P. Turner (Eds.), *The Sage handbook of social science methodology* (pp. 67-81). London, UK: Sage.

Ragin, C. C. (2008). *Redesigning social inquiry: fuzzy sets and beyond*. Chicago: University of Chicago Press.

Ragin, C. C. (2009). Reflections on Casing and Case-Oriented Research. In D. Byrne & C. C. ragin (Eds.), (pp. 522-534). London, UK: Sage Publications Ltd.

Ragin, C. C., & Becker, H. S. (1992). *What is a case?: exploring the foundations of social inquiry*. Cambridge ; New York: Cambridge University Press.

Reed, I. A. (2011). *Interpretation and Social Knowledge: On the use of Theory in the Human Sciences*. Chicago: University of Chicago Press.

Repko, A. F. (2012). *Interdisciplinary research: process and theory*. Thousand Oaks, Calif.: SAGE Publications.

Rihoux, B., & Lobe, B. (2009). The case for Qualitative Comparative Analysis (QCA):adding leverage for thick cross-case comparison. . In D. Byrne & C. Ragin (Eds.), *The SAGE handbook of case-based methods* (pp. 222-243). London, UK: Sage Publications.

Root, M. (2007). Community-Based research. *The Sage handbook of social science methodology*, 565-577.

Roth, W.-M. (2009). Phenomenological and dialectical perspectives on the relation between the general and the particular. In K. Ercikan & W.-M. Roth (Eds.), *Generalizaing from Educational Resarch* (pp. 235-260). New York: Routledge.

Sabelli, N., & Dede, C. (2013). Empowering design based implementation research: The need for infrastructure. *National Society for the Study of Education Yearbook, 112*(2), 464-480.

Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*. New York: Springer Pub. Co.

Sawyer, R. K. (2005). *Social emergence: societies as complex systems*. New York: Cambridge University Press.

Sawyer, R. K. (2007). Simulating Complexity. In W. Outhwaite & S. P. Turner (Eds.), *The Sage handbook of social science methodology* (pp. 316-332). London, UK: Sage Publications.

Sawyer, R. K. (Ed.). (2006). *The Cambridge handbook of the learning sciences*. New York: Cambridge University Press.

Sawyer, R. K. (Ed.). (2014). *The Cambridge Handbook of the Learning Sciences* (2nd ed.). New York: Cambridge University Press.

Schatzki, T. (2001). Introduction: Practice Theory'in Schatzki, T., Cetina, KK and Von Savigny, E.(eds.) The Practice Turn in Contemporary Theory. New York: Routledge.

Schatzki, T. R., Knorr-Cetina, K., & Savigny, E. v. (2001). *The practice turn in contemporary theory*. New York: Routledge.

Schleppegrell, M. (2012). Systemic Functional Linguistics: Exploring meaning in language. In J. P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 21-34). New York: Routledge.

Schleppegrell, M. J. (2013). The role of metalanguage in supporting academic language development. *Language Learning, 63*(s1), 153-170.

Schorr, L., & Farrow, F. (2014). An Evidence Framework to Improve Results. Washington, DC: Center for the Study of Social Policy.

Schwartz, D. L., & Bransford, J. D. (1998). A Time for Telling. *Cognition and Instruction, 16*(4), 475-522. doi: 10.2307/3233709

Schwartz, D. L., Bransford, J. D., & Sears, D. L. (2005). Efficiency and innovation in transfer Transfer of learning from a modern multidisciplinary perspective (pp. 1 - 51). CT: Information Age Publishing. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective*. Charlotte, NC: Information Age Publishing.

Scoping Studies. (2013). Retrieved 3/4/13, 2013, from http://www.methods.manchester.ac.uk/methods/scoping-study/index.shtml

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

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: an effort-reduction framework. *Psychol Bull, 134*(2), 207-222. doi: 10.1037/0033-2909.134.2.207

Shizha, E. Chara chimwe hachitswanyi inda: Indiginizing science education in Zimbabwe. In D. Kapoor & S. Jordan (Eds.), *Education, participatory action reserach, and social change.* (pp. 139-153). New York: Palgrave Macmillan.

Sieber, S. D. (1973). The Integration of Fieldwork and Survey Methods. *American Journal of Sociology, 78*(6), 1335-1359. doi: 10.2307/2776390

Slavin, R. E., Cheung, A., Holmes, G., Madden, N. A., & Chamberlain, A. (2012). Effects of a Data-Driven District Reform Model on State Assessment Outcomes. *American Educational Research Journal, 50*(2), 371-396. doi: 10.3102/0002831212466909

Smith, M. L. (2006). Multiple methodology in education research. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research*. Washington, DC: American Educational Research Association & Lawrence Erlbaum Associates, Inc.

Snilstveit, B. (2012). Systematic reviews: from 'bare bones' reviews to policy relevance. *Journal of Development Effectiveness, 4*(3), 388-408. doi: 10.1080/19439342.2012.709875

Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of Development Effectiveness, 4*(3), 409-429. doi: 10.1080/19439342.2012.710641

Solórzano, D. G., & Yosso, T. J. (2002). Critical race methodology: Counter-storytelling as an analytical framework for education research. *Qualitative Inquiry, 8*(1), 23-44.

Spillane, J. P. (2012). Data in Practice: Conceptualizing the Data-Based Decision-Making Phenomena. *American Journal of Education, 118*(2), 113-141.

Spillane, J. P., & Kim, C. M. (2012). An Exploratory Analysis of Formal School Leaders' Positioning in Instructional Advice and Information Networks in Elementary Schools. *American Journal of Education, 119*(1), 73-102. doi: 10.1086/667755

Spradley, J. P. (1979). *The ethnographic interview*. New York: Holt, Rinehart, Winston.

Spradley, J. P. (1980). *Participant observation*. New York: Holt, Rinehart, Winston.

St Pierre, E. A. (2002). Comment:"Science" rejects postmodernism. *Educational Researcher, 31*(8), 25-27.

Strike, K. (2006). The ethics of educational research. In J. L. Green, G. Camilli, & P. B. Elmore (Eds.), *Handbook of complementary methods in education research* (pp. 57-73). Washington, DC: American Educational Research Association.

Sue, D. W., Capodilupo, C. M., Torino, G. C., Bucceri, J. M., Holder, A., Nadal, K. L., & Esquilin, M. (2007). Racial microaggressions in everyday life: implications for clinical practice. *American Psychologist, 62*(4), 271.

Supovitz, J. (2012). Getting at Student Understanding - The Key to Teachers' Use of Test Data. *Teachers College Record, 114*(11), 1.

Suri, H. (2012). Epistemological pluralism in research synthesis methods. *International Journal of Qualitative Studies in Education, 26*(7), 889-911. doi: 10.1080/09518398.2012.691565

Suri, H. (2013). *Towards Methodologically Inclusive Research Synthesis*: Routledge.

Suri, H., & Clarke, D. (2009). Advancements in Research Synthesis Methods: From a Methodologically Inclusive Perspective. *Review of Educational Research, 79*(1), 395-430. doi: 10.2307/40071170

Tashakkori, A., & Creswell, J. W. (2007). Editorial: The New Era of Mixed Methods. *Journal of Mixed Methods Research, 1*(1), 3-7. doi: 10.1177/2345678906293042

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: combining qualitative and quantitative approaches*. Thousand Oaks, Calif.: Sage.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, Calif.: SAGE Publications.

Tashakkori, A., & Teddlie, C. (Eds.). (2010). *Sage Handbook of Mixed Methods & Behavioral Research*. Thousand Oaks, CA: Sage Publications.

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

Thomas, J., Harden, A., & Newman, M. (2012). Synthesis: combining results systematically and appropriately. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An introduction to systematic reviews* (pp. pp. 179-227). London: Sage Publications Ltd.

Thomas, J., O'Mara-Eves, A., & Brunton, G. (2014). Using qualitative comparative analysis (QCA) in systematic reviews of complex interventions: a worked example. *Systematic Reviews, 3*, 67-67. doi: 10.1186/2046-4053-3-67

Topper, K. L. (2005). *The disorder of political inquiry*. Cambridge, Mass.: Harvard University Press.

Tseng, V., & Seidman, E. (2007). A systems framework for understanding social settings. *American Journal of Community Psychology, 39*(3-4), 217-228. doi: 10.1007/s10464-007-9101-8

Tsoukas, H. (2009). Craving for Generality and Small-N Studies: A Wittgensteinian Approach towards the Epistemology of the Particular in Organization and Management Studies. In D. A. Buchanan & A. Bryman (Eds.), *The Sage Handbook of Organizational Research Methods* (pp. 285-301). Los Angeles: Sage.

Warnke, G. (1999). *Legitimate differences: interpretation in the abortion controversy and other public debates*. Berkeley, Calif.: University of California Press.

Wasserman, S., & Faust, K. (1997). *Social network analysis: methods and applications*. Cambridge ; New York: Cambridge University Press.

Webb, E. J., Sechrest, L., Schwartz, R. D., & Campbell, D. T. (1966). *Unobtrusive measures; nonreactive research in the social sciences*. Chicago: Rand McNally.

Wenger, E. (1998). *Communities of practice: learning, meaning, and identity*. Cambridge, U.K. ; New York, N.Y.: Cambridge University Press.

Westera, W. (2007). Peer-allocated instant response (PAIR): Computational allocation of peer tutors in learning communities. *Journal of artificial societies and social simulation, 10*(2).

What Works Clearing House. (2014). What Works Clearing House Procedures and Standards Handbook (3.0 ed.). Washington, DC: What Works Clearing House, Institute of Education Sciences, US Department of Educaiton.

White, H., & Waddington, H. (2012). Why do we care about evidence synthesis? An introduction to the special issue on systematic reviews. *Journal of Development Effectiveness, 4*(3), 351-358.

Whittington, J. L., McKee, V., Goodwin, V. L., & Bell, R. G. (2013). Configurational theory and methods in organizational research. (Research in the Sociology of Organizations. In P. C. Fiss, B. Cambré, & A. Marx (Eds.), *Configurational theory and methods in organizational research. (Research in the Sociology of*

Organizations, vol. 38 (pp. 279-302). Bingley, UK: Emerald Group Publishing, Ltd.

Wickson, F., Carew, A. L., & Russell, A. W. (2006). Transdisciplinary research: characteristics, quandaries and quality. *Futures, 38*(9), 1046-1059.

Wiesmann, U., Biber-Klemm, S., Grossenbacher-Mansuy, W., Hadorn, G. H., Hoffmann-Riem, H., Joye, D., . . . Zemp, E. (2008). Enhancing transdisciplinary research: a synthesis in fifteen propositions *Handbook of transdisciplinary research* (pp. 433-441): Springer.

Wilensky, U. (1999). NetLogo. Evanston, IL: Center for Connect Learning and Computer-Based Modeling, Northwestern Univeristy.

Williams, M., & Vogt, W. P. (2011). Introduction. In M. Williams & W. P. Vogt (Eds.), *The Sage Handbook of Innovation in Social Research methods* (pp. 19-24). Los Angeles: Sage.

Wolcott, H. F. (1994). *Transforming qualitative data: Description, analysis, and interpretation*: Sage.

Wolfram Research Inc. (2007). Mathematica, Version 6.0.

Wyman, L., Marlow, P., Andrew, C. F., Miller, G., Nicholai, C. R., & Rearden, Y. N. (2010). High stakes testing, bilingual education and language endangerment: a Yup'ik example. *International Journal of Bilingual Education and Bilingualism, 13*(6), 701-721. doi: 10.1080/13670050903410931

Yin, R. K. (2009). *Case study research: design and methods*. Los Angeles: Sage Publications.

Yin, R. K. (2014). *Case study research: Design and methods*. Los Angeles, CA: Sage publications.

Znaniecki, F. (1934). *The method of sociology*. New York: Farrar & Rinehart.

Zolkower, B., & Shreyar, S. (2007). A teacher's mediation of a thinking-aloud discussion in a 6th grade mathematics classroom. *Educational Studies in Mathematics, 65*(2), 177-202.