The Dissertation Committee for Keith Merriman Sturges certifies that this is the approved version of the following dissertation:

The Cultural Production of the Modern Program Evaluator in Education

Committee:

_____
Douglas Foley, Supervisor

_____
Kevin Foster

_____
Martha Menchaca

_____
Luis Urrieta, Jr.

_____
Pedro Reyes

The Cultural Production of the Modern Program Evaluator in Education

by

Keith Merriman Sturges, B.A.; M.Appl.Anthro.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2011

Dedication

This dissertation is dedicated to my sons, Anthony and Brian Sturges, who sacrificed many holidays, birthdays, and potential hiking trips so that I could persevere with my ambition to complete this work.

The Cultural Production of the Modern Program Evaluator in Education

Keith Merriman Sturges, Ph.D.

The University of Texas at Austin, 2011

Supervisor: Douglas E. Foley

The Cultural Production of the Modern Program Evaluator in Education is a three-year critical ethnographic investigation of the identity production of program evaluators in education. The methodological approach, grounded in Critical Discourse Analysis and analytic induction, includes: 1) open-ended interviews with 20 program evaluators, 2) of which 3 were expanded into case studies, 3) numerous email exchanges, 4) personal reflections from 16 years as a professional program evaluator, 5) field notes and 6) document analysis. Using Holland et al.'s (1998) social practice theory of self and identity, this dissertation outlines the processes, identifies the cultural tools, and provides a concise political-economic history that depicts how social scientists become program evaluators. The goal of this project was to study identity production through discourses and everyday cultural practices as a way to understand how social scientists come to accept, embody, and become passionate about the figured world of contract program evaluation. This includes drawing upon and contributing to existing meaning structures and systems of privilege. The study includes detailed case studies of program evaluators' agentic day-to-day responses to a shifting political economic landscape and competing ideological purposes for conducting evaluations.

Table of Contents

List of Tables

Chapter 1: Introduction

> Relative performance on the AP examinations at TGAP schools continues to remain well below state and national averages. … More disturbing are the trends of the last two years. Since 2003, 20% fewer students took 19% fewer AP examinations at TGAP schools. At the same time, the percentage of examinations with scores of 3 or higher fell from 30% to 21%. Declining performance can be expected with increased participation, but not with decreased participation. It appears that open-enrollment policies for AP courses and financial support for examinations enhanced student access through 2003. The decreased participation and poor performance since 2003 are discouraging. These data raise concerns about the academic preparation of students who are enrolled in AP coursework (Texas Center for Educational Research 2006).

The above snippet of lessons learned, taken from a summative program evaluation report, was followed by recommendations to the funding organization asking that it modify the program to ensure that teachers were actually implementing the reform in the classroom with fidelity, were completing the required professional development, and had ongoing access to university faculty support on a regular basis. The Year 6 TGAP program evaluators reported that TGAP promoted college awareness, assisted teachers in developing the Advanced Placement program, and provided incentives for students to complete the AP exam in six Texas districts serving primarily underrepresented students. Using diverse measures (AP exam-taking patterns and scores, over 400 interviews with

teachers and students, and observations in over 600 AP classrooms), some salient trends became clear. AP programs were established, teachers were trained, and classes were populated with underrepresented students.

But, there is considerable question as to whether these courses were truly advanced. Advanced programs rely heavily on a combination of content-specific knowledge and innovative instructional strategies, such as graphic organizers, hands-on learning, interactive discussions, and self-directed research. Ideally, according to the College Board, AP teachers use strategies that enhance students' ability to learn on their own. But, in over 90% of the 628 AP classrooms, observers saw teachers lecturing to students most of the time. When questions were asked of students, they typically required one-word answers. Teachers seldom guided discussions and rarely used project-based learning. The observations revealed limited use of the instructional strategies described in advanced course guides. While the report offered feedback on the use of particular strategies, nowhere did it describe GEAR UP's unintended effect of widening the disparity between historically underserved students and those who have been well-served through advanced programs.

Observations and interviews revealed teachers' deeply held expectations about students, based on students' social characteristics. If teachers believe certain (read underrepresented) students are less capable because of social positional conditions, such as class or ethnicity, teachers expect less from them, push them less, and steer them away from rigorous coursework. TGAP teachers felt that they needed to "water down" the curriculum for "these" students. A history teacher said, "We worry about putting too

much pressure on students and having them dropout of school. They are at-risk." The "at-risk" designation spoke for itself since teachers perceived a cap to what students could learn. An assistant principal said: "We often squeeze water from rocks here."

Furthermore, in interviews, students detected little difference between advanced and regular classes. They described AP as "practically the same thing as the regular classes, except that they're more strict." A 12th grader said, "My AP history teacher doesn't really teach AP. She just makes you do another paper." Students said most of the challenge in AP was all the "extra" work. And, of the AP exam, students said they were willing to try the exams, although they believed they would not score very high. One said, "Well, it is free." From the report, a reader might overlook how the lack of rigor and lower AP scores contributes to reproduction of the status quo. For students who choose to go to college, lower AP scores are likely to translate into attendance at less-selective institutions of higher education (Lichten 2000).

Problematic with the GEAR UP evaluation report referenced above is that, while it targeted teachers as implementers and students as recipients, it did not scrutinize the reform's production, support structure, or funding mechanism. The program was operationalized according to a logic that defined steps founded on sweeping assumptions about categories of students (e.g. the social-positional intersection of Hispanic and low-SES as "at risk") by enumerating students meeting inclusion criteria and that teachers are not fully competent. And, nowhere did the report raise concerns about the ways program developers' assumptions about low-income, minority students might have influenced its lack of desired impact. The program was designed and implemented in the best interest

of, but without input from, those being served. Students, teachers, and principals neither influenced which data were collected nor how they were used. Thus, in essence, the evaluators were given relatively free range to judge the program at one spot, but not in its conceptualization, production, refinement, or in the negotiations between influential power groups that decided it was the answer to the problem of low college attendance of underrepresented students in Texas. The program's failure to yield desired results—because of the purported shortcomings of its recipients—was followed by its discontinuation.

From a critical lens, program evaluation may deny itself an understanding of social inequities by pre-determining normative rules and favorable dispositions of the ideal student while sidestepping multi-level social inequity. The reference, the "norm," is both a construction of and method for reproducing the dominance of the social positioning of the group who assembled the curriculum, the instructional approaches, and the assessments. In other words, program evaluation does not question what is taught, who is teaching whom, how students are selected and distributed, or how the content material is socially produced. It, thereby, contributes to the logical error of trying to formulate a theory of curriculum from a theory of learning (Apple 1978). The language of learning tends to be apolitical and ahistorical (which, from a social justice perspective, masks the political and economic backdrops of curriculum organization, selection, and effect). For instance, the student achievement measures took (and still take) for-granted the kinds of knowledge and its organization that finds its way into schools. Unexamined are the presuppositions of what curriculum developers consider valuable knowledge.

Every program evaluation report ends with a similar set of straight-forward, data-derived lessons learned and recommendations that are ready to be enacted into policy and program practice by those who make decisions about education reform program selection, deployment, funding, and development. These suggestions for program improvement harmonize seamlessly with the hundreds of pages of meticulously-analyzed, purportedly neutral process and outcome data that were generated at the crossroads of rigorous social scientific method, client specifications, and interactions with participants.

Michael Taussig's notion of the "labor of the negative" comes to mind. This accumulation of all that is generally known, but which cannot be acknowledged, is bound up in a cultural act of learning to be silent in front of a public secret (1999). Beyond its purported neutral stance, contract evaluation "[c]onstructs a model of the world with a system of categories that come to expect certain relationships and behaviors to occur and then experience those categories" (Torres 1999) according to a theory of action. This form of inquiry turns to social scientific procedures with the intention of controlling processes through forecasts, predictions, and manipulations, thus, both drawing from and maintaining existing meaning structures.

While educational reform model developers and education-focused government agencies rely on this scientific rationality to sanction and "certify" what works, the commonsensical technical procedures associated with data collection and reporting operate within the program box. The "truth" that is assembled relates only to the inputs, outcomes, actions, and people at the site of deployment. Omitted are the manifold

political, social, and economic contexts and interactions with other programs, events, and interest groups. Since continued funding is often contingent upon compliance within the framework of accepted and approved educational reform steps, neutral(ized) evaluation projects, ostensibly established to solve practical social problems, mask broader social inequities. The definition of an education reform's success and the creation of interventions intended to meet the learning needs of struggling students is the toil of middle-class knowledge workers.

The space between knowledge producer and subject serves as an incubator of potential power exercises. In essence, it is a power to judge. This power to judge can be interpreted as disguised as habituated objective knowledge. Its processes are deeply embodied as they are repeated until they become so deeply habitual that they appear to be natural or normal. This form of normalized power encompasses not only the production and structuring of knowledge, but deeply-rooted ontological assumptions that operate as commonsense or habit (Sullivan 2006). For instance, although the targets of Title I reform experiments are "at risk" students, the middle-class knowledge workers who evaluate the deployment and impact of reforms are depicted and depict themselves as unbiased change agents. This supplies the ideology of naming the "at risk" and producing adequate, replicable measures of this *other*'s success and failure. This naming is an exercise of power.

Yet, this power is not held in the hands of program evaluators. It is distributed throughout the system (Foucault 1977). Like all social research, the researcher is privileged with a sense of entitlement to enter the lives of subjects. This research is

mediated by a complex political and economic enterprise that, shrouded by discourse, expresses both the interests and power of an educational industry and the commitment to social justice of a liberal ideology. This tension between apparent poles mediates what can be said, by whom, in what terms, and to what audience.

The principal artifact, and loci of this exercise, is the report and its accoutrement. This power is exercised primarily through social scientific textual representations that are intended for consumption by a microcosmic audience and, therefore, do not undergo the scrutiny that basic research undergoes. In the context of postindustrial capitalism, textual knowledge is a major medium of commercial exchange, occupying a critical space in the politics of hegemonic representation. And, the power embedded in textualized rhetoric is disguised in the discourse of educational reform.

In A Political Sociology of Educational Reform (1991), Popkewitz contextualizes educational reform as a social and political practice that emerged during the Progressive Era. In doing so, Popkewitz provides a synopsis of the mercurial term *reform*, which, until the early 1800s,—and founded on the notion that change was tied to reason and "systematic human intervention to social institutions"—was spared for sinners in need of spiritual salvation. The rise of institutions provided a space for the assignment of nation-state collective identities that could be disaggregated into population categories that would enable state administration to the group and to the individual, thereby granting the institution with the power to save those deemed to be in need of saving.

During the late 1800s, the Age of Reform brought with it management structures to orchestrate and regulate activities in public sectors, including mass schooling, with the aim of improving social life. The meaning of *reform* shifted again by the early 1920s to index "the application of scientific principles as the means to achieve social enlightenment and truth."

With the rise of this scientific approach came specializations within education—teacher, administrator, researcher. In his précis, Popkewitz explains that the term *reform* does not have a single meaning, but "that its meaning shifts with a continually changing institutional environment." Nonetheless, *reform* has consistently been based on Protestant moral opinions, most notably in the U.S. Its coupling with social scientific inquiry helped the state apparatus "interpret the more complex social relations and interdependence among communities and to reassert moral, social, and cultural authority through the process of reform" (Popkewitz 1991).

Nearly a decade later, Popkewitz returned to this topic. In the U.S., this discourse promulgated a national image, a norm, a standard (Popkewitz 2000). At the same time, it borrowed from the logic of human development goal setting and achievement. The mechanisms of this state-administrated reform discourse were put into operation with an ideal and a set of methods for measuring strides toward betterment; strides toward attainment of a culturally-produced standard and a rationali(ized) notion of growth. In this context, the program evaluator was endowed with the expert status to make technical judgments about the reform program and both the individual and group (student, teacher, administrator).

I would like to add to Popkewitz' cogent sociological history a précis on the rise of the contract in educational servicing. Although he attributes the moral authority of Protestantism as the origin of reformist thinking, Popkewitz' contention that the contemporary nation-state is the locus of power overlooks the other pressure groups, notably those with political-economic and professional interests in social administration. These include non-profit organizations, professional associations, and private companies that thrive in the realm of education services and materials. Reform applied to education has obvious economic implications—and considerable potential for gains or losses—for textbook companies, for instance. The power of reform does not rest solely with one organization or kind of organization. This knowledge work is supported by a discourse that is premised on the notion that social scientific method can, and is actually used to, measure and select from what curricular and instructional programs work most effectively in public education. This discourse, the Educational Reform Discourse (ERD), is put into operation through various forms of commoditization and exchange, especially legislative, economic, and symbolic exchange. For this reason I use the term *reform discourse*.

"Discourse" refers to the recurrence of statements and constructs across fields and texts (Foucault 1972), which, together, produce systems of meaning and fields of knowledge that disguise and naturalize unequal power relations. Discourses are dynamic and continually regenerated in specialized texts (e.g., evaluation reports) and contexts (e.g., meetings with clients to negotiate evaluation services). Their repetition across spaces, thoughts, and artifacts make them appear ubiquitous and, as such, they seem to be

the natural ontological state. Discourses serve institutional and instrumental purposes through this ability to disguise and shape-shift.

A major task for the ERD, the function that makes it capable of sorting, monitoring, and improving its apparatus, is its application of research. The research portion of the ERD is not single faceted, however, but consists of several bodies, including basic or academic research on potential reform models, accountability systems, and program evaluation. While each of these has a different, and often competing, voice, they share features of the same lexicon, moral orientation, and insistence on their own objective stance in the process of judging the value of reforms, for whom they should be designed, and by whom and under what circumstances they will be deployed. It is this repetition across domains, organizational types, spaces, and texts that produce the ERD's apparent ubiquity and, therefore, the appearance that it is a normal or even natural condition.

An apparent contradiction arises when one juxtaposes the ERD's moralist orientation and its apparent objective orientation. This distancing of operative from the operant is the homology of *reform's* previous iterations. In order to save souls, one must work for the church in the name of God, rather than as an individual. Likewise, in order to manage populations, one must act in the name of the agency or nation-state in the name of the ERD. This distancing entitles one to judge and, I argue, is one of the crucial steps to becoming a successful program evaluator. Foucault described the constructing character of discourse—how social formations (epistemes) and local sites define and position human subjects (1972). When embodied and otherwise normalized, these

discourses become recalcitrant, irresistible *truths*, which, because of how they appear natural, go undetected.

Evaluations are not invisible. But, outside of the contractor-contracted relationship, they are not questioned. In the uncritical narratives unfolding around educational reform (and, therefore, program development), the evaluator's work is depicted as apolitical and disconnected from the social nexus in which she operates. Ernest House, one of the founding members of the American Evaluation Association, comments: "evaluators see themselves as neutral toward social classes and interest groups and espouse a rationale of expertise, service and efficiency—the professional ideal" (1993, 28). The official rules by which she plays, the American Evaluation Association's standards (http://www.eval.org/EvaluationDocuments/progeval.html) along with rigorous methodological approaches, ostensibly prepare the evaluator to be dispassionate and unbiased and therefore, able to manufacture this officially sanctioned, objective knowledge.

Although every federally funded—including those that "flow through" state departments of education—educational reform program requires an external program evaluation, the textual representations are produced and consumed at the margins. Since this knowledge production happens behind the scenes, it is an *absent presence*, to appropriate and repurpose Apple's term (1999). Popkewitz notes (1991) that the politics of scholarly knowledge production (and knowledge consumption) go largely unseen because of a distancing of scholarly research from social interests. But, in the case of program evaluation knowledge work, this distancing that informs public interests

directly, I would argue, is even farther removed from the public because of its technocratic-esoteric status and miniscule audience. More importantly, knowledge claims are the proprietary intellectual property of the client and, thus, seldom made available for mass consumption, much less offered for the scrutiny of other knowledge producers. This is an interesting problem in the context of contemporary legislation and its connection to the ERD, which seek to expose failing children, failing teachers, and failing programs. This could be a tiny fracture in the discourse—the sort of fracture described by Althusser (1970)—while it is impossible to escape ideology, where the system's creation of an illusion of coherence is not yet complete. It opens a peephole into one matrix of power in the ERD.

While the relationship between knowledge production and power is well researched (Bourdieu 1977, Trouillot 1995, Foucault 1972), as is the relationship between that knowledge production and the augmentation of inequity in education (Apple 2004, Popkewitz & Brennan 1998), missing from the literature on public education reform is a critical analysis of knowledge producers that bridges the individual and the institution as mutual conditioners that interact with and enhance the ERD.

The production and consumption of texts, is imbued with such issues as possessing the right to depict, to influence decisions, or to enter "officialized" realms are aspects of power and position (Bourdieu 1977, Foucault 1972). The indexed poststructural work identifies the individual as a subject who is subsumed into a discourse. In other words, while numerous studies link production to producer, I attempt to take that link a step further to consider the way that production contributes to the

production of the producer. The study I offer takes a deeper look at the everyday practices of contract program evaluators as they actively (re)construct their professional identities in the context of evaluation work.

The purpose of this study is neither to depict planned educational change nor to challenge the effectiveness of program evaluation. Instead, it is a cultural history, developed around a set of personal histories, of the knowledge producers and workers who work outside of or on the fringes of academia in the context of the ERD. What I offer is not the definitive problematization of commodified praxis, but a critical suggestive model of the cultural production of a particular brand of post-academic knowledge workers. To do this, I depict evaluator descriptions of the processes through which their identities were produced, tracing the development from one of social scientist to one of contract evaluator. This study relies on a sociocultural notion of identity that is linked inextricably to the political-economic and its discursive context. My aim is to interweave identity narratives and traces of the ERD, while also describing a political-economic history as a backdrop.

I attempt to describe my understanding of how social scientists come to accept, embody, and become passionate about the figured world of evaluation. In this project, I studied three interwoven dimensions of this goal: (1) to describe the identity production of evaluators, (2) to describe the evaluation industry as a "figured world," or perceived space of action in which identity is produced and (3) to describe evaluation's knowledge production as the commodification, manufacturing, and distribution of artifacts that contribute to the maintenance of the ERD. I hope to add to the corpus of critical research

a glance at not only the artifactual traces of social scientific work, but a sensitive and reflective inquiry of the (ethno)-genesis of the producers of that knowledge themselves. This study probes the ways in which actors create and express their own histories and identity narratives and then act accordingly in a figured world in which subjects (people and programs) are analyzed and where a kind of "truth" is negotiated, assembled, and shared. This entanglement of negotiated findings and identity shaping needs to be told by a group that is, ironically, often silenced by its own positionality of privilege and marginality. Evaluators have few places to tell their story.

Theoretical Framework

The professional identity of program evaluators is produced continually through everyday cultural performances—data collection and analysis, report generation, meetings, project negotiations—with constant reference to and use of a variety of processes, discourses, and materialities. This conceptualization of institutional(ized) identity and discourse calls for a theoretical framework that links evaluators' interactions, internalizations, and knowledge production acts as reflections of macro social forces. Critical Discourse Analysis (CDA) is an analytic tool that attempts to lay bare power embedded in institutional structures while it also addresses local or multi-level context, history, and intertextuality that are often neglected. CDA's abductive orientation to knowledge production, which I apply to personal stories throughout this study, provided a means for compiling discourse traces that led back to the power matrices and, thereby, helped produce a logic of the cultural practices of evaluation, its assumptions about its

function, and some of the taken for granteds that are difficult to articulate. In other words, I listened to what was not said as much as what was.

To grasp the modern evaluator and her predecessors, a rendering of the broader historical backdrop provides an understanding of legal, political, economic, and theoretical epochs and, therefore, discursive genres. Contemporary scholars interested in the production of identity vis-à-vis discourse and materiality, such as Holland and Lave (2001), position individual history in the context of broader history. These scholars tend to connect sociocultural practices of identification to broader structural forces. In doing so, they depict a stationary history rather than continue the dialogic coproduction at a macro level, and therefore, sidestep the messy task of dismantling—or even mentioning—official(ized) histories. Clearly, missing from the work on the juncture of identity-making and becoming tamed by discourse (as both the flood of knowledge systems and as history of those systems) is a convincing bridge between critical macro histories of ideas and depictions of localized language, identity, and knowledge production. CDA attempts to address this shortcoming by combining Foucaultian-inspired lines of poststructuralism with neomarxist political-economic analysis to denaturalize everyday practice. It draws from the assumption that both identity production and discursive formation are negotiated, constructed, and produced (Bakhtin 1981, Fairclough 1995, Foucault 1972). It is, therefore, both critical and constructive. And, it defines both broad and local formations.

Ideally, a history of the present would aid in formulating an understanding of the ways in which associated practices have become everyday practice of evaluation

knowledge production. And, in addition to providing a broad discursive backdrop that indexes shifts in practices that have become *evaluation*, genealogy also provides insight into the particular discursive formations that become both inhibitors of and tools with which evaluators self-author. Initially, I mused about the prospect of connecting genealogy to individuals as a discursive matrix. Through interviews with people who entered the field at different times, I would catch glimpses of those shifts in discourse. But, to do so presupposes no infiltration of other discursive shifts in the lives of the old-timers. Thus, a genealogical thread serves as part of a broader analysis of institutional history of the field.

In this study, identity narrative is the coherent self story that locates itself in material, imagined, and relational worlds, which, themselves, are created and told in this discursive frame. As such, the various pieces of the identity story strive for internal consistency. New experiences that do not fit the story, when not ignored, are internalized and incorporated into it, thus altering the story or becoming a point of contention. The story's coherence is the historical sedimentation that decreases the story's susceptibility to deep revision. Its ongoing-ness exists only in relation to a narrative (which gives the impression of permanence). For a proficient evaluator, the attributes that made her a researcher become so much a part of how she sees herself that they seem internalized under her skin. Her everyday performance and increasing proficiency in a perpetually dynamic collective space where the story of that fixedness has meaning that also resonates with other evaluators gradually helped her internalize a sense of belonging to the industry.

As she tells her story to co-workers, she is also reaffirming for herself her evaluator identity by doing what Holland et al (1998) refer to as self-authoring. Based in great part on Bakhtin's use of the term dialogism, self-authoring is a kind of gathering together of messages, voices, and discourses, and addressing both the outside world and the self. Through the dialogic process of self-authoring, evaluators learn to detach themselves from and their reactions to earlier self understandings and learn to perform according to the rules, expectations, and interpretations of the newly *figured world* (e.g., from academia to the corporation). Self-authoring takes place at the nexus of the *figured world*—a collective "everyday" space where evaluators become active and passionate about what they do—and *history-in-person*—a space in which identity, performance, local practice, and "long term transformative struggle" (Holland & Lave 2001) meet to produce "traditions of apprehension" (Holland et al 1998). Together, they provide a sense of unity, belonging, and purpose. I agree with Holland and colleagues that people are both drawn into and help shape figured worlds. Indeed, the figured world and self-scripting, together, form a generative recursive loop. In other words, they are mutual conditioners of individual and group identity.

The production of self narratives comes about through a restructuring of understandings about the world based, in part, on materiality and discourses. Through repeated participation, figured worlds become embodied, and when they do, they also become sociohistoric bases for self-understanding. But, repeated action alone does not automatically induce identity production. It is through the constant elaboration of an identity narrative that people interpret and reinterpret past experiences and revise self-

understandings. Self-authoring requires both a distillation of everyday experience and a way to fit that distillation into existing self understandings. By drawing on the self story vis-à-vis social position, discourse, proficiency, and social networks, a social scientist may choose to become an evaluator. This set of relations between self-authoring and participation in contentious local practice provides the partial fixedness taken from one figured world to another. The self-understanding lends itself to a narrative rendering of the ways in which events are experienced and interpreted by individuals (Urrieta 2006).

This theoretical précis has broached a number of topics requiring further elaboration. I will identify some of my assumptions about the constructs of agency, structure, and, therefore, identity. First, at the intersection of poststructuralist determinism and constructivist free will, humans actively author their identities in the interstices of compulsory action. Furthermore, whether by choice or because they feel or are actually obligated to do so, beliefs are influenced by everyday activity. As they interact with their everyday lives, people use available cultural tools (beliefs, discourses, social position) to develop a sense of what is and what ought to be; a sense that often permeates as commonsense. In this context, the recursive process of identity-making takes place. Identity is a complex, dynamic process that is never complete. It is depicted to self and to others as a self-story with internal consistency. This personal narrative is located in the social and material worlds and draws its coherence from reflections of everyday practice. Given these assumptions, there is constant interplay between a dynamic sense of personal and group identity, materiality and performance, local history and larger historical movement. The embodiment of discourses is the function of

multiple processes—of a combination of discursive inscription and of agentic self-authoring.

I do not attempt to accurately predict how much of identity and performance are due to improvisation and how much is attributable to inscription, since that which appears to be unquestioned complicity might, instead, be a form of agency. Such a rigid, dichotomous distinction might not always be reflective of the process, since it is possible to imagine a dialogic relationship between embodiments of discourses and intentional, conscious forms of action. This was a continuing struggle in this study and not one that I have attempted to settle.

Evaluators, both individually and collectively, are not the passive recipients of orders or products of their social contexts. They also actively, and critically, appropriate cultural artifacts and produce actions (Holland et al. 1998). Program evaluators actively "read" the specialized discourses and practices that make up the work of their field, construct their own understandings of program evaluation, and then formulate actions that are shaped by, and in some instances constrained by, specific situational, or contextual social and cultural forces. Furthermore, a modicum of improvisation resides in available cultural tools associated with the figured world of evaluation and the other situational figured worlds in which evaluators thrive (e.g., academia). These tools provide evaluators a degree of agency with which to fashion their knowledge producing identity.

Methodological Approach and Research Design

My method for articulating identity production consists of a critical analysis of identity narratives resulting from oral histories of program evaluators. Embedded in identity narratives are depictions of the ways in which social scientists come to perceive themselves, and act purposefully, as program evaluators. These narratives offer a sense of how each evaluator-in-the-making orchestrates his or her repertoire of key identity components (Holland, et al 1998; Urrieta 2009). For instance, the narratives render the critical steps experienced as serendipitous doors that open and close (sometimes (re)presented as teleological growth), the key internalized voices that echo advice throughout the journey, titles and tag lines, and sense-making in the face of discontinuity and new experiences. I used an ethnographic approach that draws attention to a cultural landscape where programs are assessed, knowledge is assembled and distributed, and, of particular import to this study, evaluator cultural identities and statuses are produced, performed, and refined with every subsequent performance.

Some scholars regard ethnography in terms of value-free fieldwork. For instance, Brewer defines ethnography as the study of:

> People in naturally occurring settings or fields by methods of data collection which capture their social meaning and ordinary activities, involving the researcher participating directly in the setting…in order to collect data in a systematic manner but without meaning being imposed on them externally (2000, 6).

This stance, which is mired in counter-reflectivity, successfully sidesteps the fact that the identification of "naturally occurring settings" and "capturing of social meanings" is, itself, an ontological imposition. To address the looming prospect of projected ontology, some ethnographers and historians of ethnographic writing offer cultural interpretation (Clifford 1988; Marcus & Clifford 1986). Clifford, for example, beseeches ethnographers to render their accounts with a degree of self-awareness about the ways they transform the data they gather in the field into textualized knowledge. While this stance offers some guidance for textual analyses of meaning systems, the perpetuation of bounded cultural frames distracts research from an analysis of culture as process where identity is constantly shaped and reshaped and where the figured world's operating logic is decoded and put into practice. At the same time, everyday practice tests the boundaries of the ERD and, thereby, both reproduces and lends itself to the discourse's expansiveness.

Cultural representation is tangled up in the ways research is conceptualized, data are collected, and the researcher situates himself (Trouillot 1995). This is a particularly salient caveat since the researcher connection to the research—as both a hidden process and as a process of identity production—is precisely a matter of concern in this study. Although obscure in explaining a praxis-oriented solution to redefining ethnographic fieldwork, Fabian's main contribution, for this study's purposes, is in his contention that representation of the *other* needs to be strengthened by critically reflecting on the activities and processes that produce cultural understanding (1983). This solution is enmeshed in how the ethnographer situates herself by becoming attuned to what she

hears, more than to what she sees, since she is cognizant that she does not possess but *is* communication. Because she recognizes her undeniable participation in the physical, sociocultural, and hegemonic intersubjective ethnographic process, she can no longer responsibly write herself out of the story she tells. Thus, in addition to the call for attention to power relations and historical representation in ethnographic inquiry and depiction, this work, especially when it is conceptualized with the intention of understanding the sense the *other* makes of the world, ought to elevate the epistemological value of narrativity.

With this perspective in mind, I selected an oral history approach that draws heavily on identity narratives instead of participant observation so that I could emphasize self-understandings of participants over my own projections. Yet, no matter how systematic or sensitive my analysis of the narratives I heard, it would be unreasonable for me to imagine that my incorporation into the figured world and, therefore, what has become my naturalized way of doing research, did not, at least implicitly, influence the way I envisioned and interacted with this world. More to the point, my experience as a contract researcher and a student of critical literature combined to form the lens through which I conceptualized this study and, later, interpreted the stories I heard. While I accept responsibility for any bias in the project's design, sampling, and much of its meanderings, I think it is fair for me to claim that no one is familiar with the whole of this figured world's terrain. Because of that partiality, my aim was to construct a composite of perspectives in order to produce a tentative description of the figured world and its manifold processes.

To be more specific about my concern, the collection of life history interviews with program evaluators dispersed across an expansive terrain may raise concern about the potential for a gross limitation in this study's methodological approach. Unlike a sited and historically-situated ethnographic précis, this initial exploration into the cultural world of evaluation relies on temporality, but its physical place is, at best, an *imagined community*, to borrow from Anderson (1991). That, alone, does not make the cultural practices and processes of identity production any less salient. Indeed, because evaluation, as practice, is carried out in multiple sites, much of which is cyber-sited, no single physical location captures its everyday practice. Be that as it may, unlike an ethnographic study that comes to know a space well, we cannot really know to what extent the findings I offer in this study are applicable across all external program evaluators in education in the current configuration of the evaluation figured world.

I suspect, based on feedback from evaluators occupying vastly different positions, that many of the findings are fairly accurate portrayals of the field and practice. Others, however, may be limited in their applicability. For instance, I have no way to predict with any degree of precision how many evaluators fit into some of the categories I offer in this study. The aim of this exercise was to amass enough evidence to evoke and sustain a conversation. In the course of that continued conversation, some of the parameters will, most likely, be identified and some will be refined or modified.

Sources of Information

I used oral history interviews as the primary method. The available literature on identity production provides limited guidance in methodology since most studies are

drawn from research on other topics that have implications for, but are not squarely focused on, identity. I decided to use a framework that would lend itself to analysis across cases and that would allow me to develop illustrative case studies (Miles & Huberman 1994).

The interviews provided material for generating within-case vignettes of the evaluator-in-the-making experience and added depth and life to the overall description of the field. The within-case stories depicted the ways in which evaluators negotiate and rework conflicting elements of their identity narratives and how, over time, they both internalize and contribute to the ERD and the institutionalized practices associated with evaluation (as a subgenre of social science knowledge production). Taking a few steps back to look across those stories yielded multiple perspectives and, at the same time, patterns that highlighted noteworthy stories that distinguish program evaluation eras, the types of institutions that do program evaluations, and the depth and process of individuals' incorporation into the industry. The interviews provided valuable insights into how self-authoring took place, with what cultural tools, and within what constraints.

I crafted questions from both specific identity literature categories that include social position (race, gender, social class), early socialization (values, sense of activism), education (degrees sought and why), and career moves (impetus for selecting this industry over others), as well as broad, open-ended depictions (e.g., Tell me about how you became an evaluator) that relied on the uniqueness of each participant's story. I asked participants to be involved in a minimum of two interviews. In the first interview, I asked participants to sketch out their life histories from early childhood through college

and then to connect those sketches to their careers. I wrapped up the interview with a few questions that were derived from theoretical and empirical assertions about the evaluation industry, especially those related to shortcomings or gaps in evaluation, and advice they would give to budding social scientists considering program evaluation as a career path.

In a second interview, I explored everyday practices of evaluation, which helped me understand the experiences against and through which evaluators continually produce their identities. I asked participants to describe one project from early in their evaluation career from conceptualization through finish. Questions surrounded themes of management (e.g., hiring and task administration), communications (e.g., with funding agencies, the public), data collection and analysis (e.g., case selection, dealing with conflicting findings), dissemination (e.g., quality assurance, "sanitization"), and marketing (e.g., objectivity).

I also asked them to describe how they were incorporated into program evaluation, surprises they encountered on the road to proficiency (and how those surprises were resolved), and how they came to be able to function as members of interdisciplinary teams. Finally, I asked broad questions about the field, such as descriptions of and distinctions between the kinds of organizations doing this sort of work and changes participants noted over time in the practice of evaluation. When participants offered what appeared to be important insights in other areas, I followed those threads instead of the ones I had predicted to be most central.

Using a purposive sample, I selected 20 evaluators. The first criterion was education. I only selected people who held advanced degrees in social sciences[1] (anthropology, psychology, and sociology) and who had worked or do work on contract program evaluations. I did not include evaluators with training in non-social science disciplines, although many practicing evaluators hold advanced training in engineering and education. The rationale for this choice was the focus of this study, which is how social scientists are reshaped/reshape themselves into program evaluators. At a minimum, I selected people who had completed a social science degree either as part of a masters or a doctoral degree. The second criterion helped me get a broad representation of social science disciplines; participants' terminal degrees included four PhDs in sociology, three PhDs in anthropology, two PhDs in education (with MAs in history or sociology), six PhDs in psychology (including social, educational, and experimental), two PhDs in educational administration (with MSs in psychology), one who is working on his PhD in communications studies, and two MAs in policy.

The number of program evaluators with backgrounds in social sciences is not infinite. The total number of program evaluators who work in education might be between 4,500 and 6,000 (I will return to this in greater detail in Chapter 3), including practitioners with non-social science degrees. Funding patterns, from federal, state, and private sources, serve as rough indicators of who is doing what work. Evaluators are aware of the limitations in funding and are familiar with competition. Thus, the figured

[1] My definition of social science encompasses the broad range of disciplines that use empirical evidence and social theory in an attempt to understanding human behavior, sociocultural practice, and process. These include, but are not limited to, anthropology, sociology, history, communication studies, psychology, and policy studies.

world's boundaries and scope are somewhat delimited. Complicating this, however, is the fact that, while many clients maintain vendor lists that include approved evaluators, there is no license or credential (other than degree) for promoting oneself as an evaluator, in general, much less an education-specific evaluator. Therefore, program evaluators are self-labeled, but they are also recognized and conditioned by the figured world.

With this in mind, I recruited the 20 evaluators along three axes. First, I invited participants who became evaluators at different periods marked by shifts in the evaluation industry. These epochs include pre-ESEA I, the mid-1970s institutionalization of the field, the 1980s reorganization of federal agencies, and the period beginning in the mid 1990s that indexes a shift toward both centralization of knowledge policy and—perhaps ironically—neoliberalization of knowledge production via outsourcing. Second, I selected participants according to their level of integration into evaluation. Some had been working as evaluators less than a year at the time of the interviews, while others had been doing them for 40 years. At the same time, some worked as full-time evaluators and others did this work as a supplement to or minor component of other roles. Third, I selected participants by organizational type. Together, these three axes served as a wide net for covering the variety of the evaluation figured world's unique spaces.

Table 1. Sample Summary Table

Participants	Epoch	Level of Integration	Main Type of Institution
5	Cold War	Guru	Regional Ed Lab
5	War on Poverty	Career evaluator	Corporation
5	War on Waste	Part-timer	Not-for-profit Foundation
5	Centralization	Interluder	University department
		Newcomer	University research center
			Self employed (hired gun)

I began recruiting participants through a combination of contacts via publicly available means (e.g., contract research organization websites) and through word of mouth (i.e., snowball sampling). In the former case, I gathered publicly available contact information (such as telephone numbers and email addresses) and contacted prospects directly. I began recruitment with my own professional network by sharing a written description of the study and invited people who met the eligibility criteria to contact me. I quickly branched out with a combination of snowball sampling and various written resources to identify prospective participants (including websites of companies and academic departments, published journal articles and books, and unpublished evaluation reports). The final sample, therefore, included a broad range of evaluators, as is summarized in the participant table in Table 2.

Table 2. Participant Summary Table

Person	Status	Began	Left	Org. type	Summary Notes
Nicole	Guru	1960s	Present	Federal govt., independent contractor, adjunct	After directing early federally-mandated evaluations before becoming an internal evaluator for the federal government. She is now a much sought after independent consultant.
Martin	Career	1960s	1980s	Corporate, academic, indep. contractor	He was as tenure-tracked psychology faculty and then conducted one of the first federally-mandated evaluations while working for a for-profit firm. Later, he opened his own consulting shop.
James	Career/ Guru	1960s	Present	REL, university, independent contractor	After directing a university center evaluation, he worked for an REL for several years. He then became a tenure-tracked faculty member in evaluation. After retiring, he expanded his private services.
David	Career	1970s	Present	Corporate, institute	He worked as a consultant for 4 years on a corporate evaluation and then went to work for that company. Later, he opened a university-based policy institute where his shop continues to conduct evaluations.
Alice	Interluder	1970s	1970s	Independent consultant	She worked as a consultant for 4 years on a corporate evaluation and then spent the rest of her career doing non-evaluation work.
Mark	Guru	1970s	Present	University-based, government, independent contractor	He began as a tenure-tracked faculty member before moving to doing contract work. Not long after, he worked for the federal govt. for more than a decade before working as an independent contractor.

Person	Status	Began	Left	Org. type	Summary Notes
Beth	Career	1970s	Present	University, non-profit	She began as a tenure tracked faculty member. After successfully beginning a research institute, she went on to become the VP of evaluation at a non-profit organization and is now the CEO.
Beatrice	Career	1980s	Present	University	She went directly from her PhD program into a faculty position. She now directs a university center that conducts large-scale evaluations.
Kathy	Side endeavor	1980s	Present	University, non-profit	She went into a non-profit before becoming disillusioned. She then became a tenured faculty member in a university graduate evaluation program. She also conducts small scale evaluations on the side "to stay connected with what she loves."
Anne	Career	1980s	Present	University, non-profit	She worked for an REL collecting evaluation data and then began teaching at a state university. Within the past several years, she switched over to a university-based evaluation center that works on contracts exclusively.
Helen	Career	1980s	Present	Non-profit, for-profit	She began as a research assistant and worked her way up to being the director of research at an evaluation firm.
Nick	Interluder	1990s	2000s	Non-profit, academic	After working as a field evaluator on a university-based study, he became a faculty member of a program evaluation graduate program. He oversees graduate students who do and he writes about evaluations.

Person	Status	Began	Left	Org. type	Summary Notes
Laura	Career	1990s	Present	Non-profit, academic	While she began her career in basic research, she later became an evaluator at a small, non-profit. She left because it was not "scientific" enough and moved to a small non-profit foundation.
Bonnie	Career	1990s	Present	Academic, for-profit, non-profit	After earning her PhD, she applied for a position at a research firm as a research associate. She is now the director of research at that firm.
Mary	Career	1990s	Present	Non-profit, for-profit	After earning her PhD, she did independent consulting for a local firm and ended up getting hired.
Sarah	Career	1990s	Present	Corporate, university	During graduate school, she worked as a researcher for a private firm. She left for ethical concerns, taking an internal evaluator job with a large school district. Later, she moved on to a university evaluation group.
Isabella	Side endeavor	1990s	Present	Non-profit	After serving as a public school teacher, she founded and presides over a non-profit organization. She conducts evaluations as a hired gun for at least 3 evaluation companies.
Paul	Interluder/ Newcomer	2000s	2000s	Corporate	During graduate school, he was an evaluation associate and rose to the level of director quickly. He later quit to work for a state government as a policy expert.
Emma	Interluder	1990s	2000s	Non-profit, SEA	In school, she worked as an evaluation associate and rose to the level of director quickly. She later quit to work for a state government as an internal evaluator.
Camila	Interluder	2000s	2000s	Non-profit	After school, she worked at a non-profit firm, where she worked for nearly 3 years. She left because she felt it was not having an impact. Soon after, she joined a large, corporate contract research firm.

The participant summary table and the categories, although theoretically and empirically driven, will demonstrate a weakness in this study's uniform coverage. This fits a core feature of the study's epistemological orientation as exploratory suggestiveness. But, it also reveals a seemingly natural commoditization of knowledge production. In turn, my use of the table reveals a certain point of tension in my study—that I have made space, internally, for meeting the needs of a consumerist audience—whether as internalized "client" voice or read in the economy of contemporary textual sense-making. Inevitably, this dissertation is written not in contradistinction to the ERD, but as an extension and representation of it.

In the strictest sense, the sampling plan failed to yield cases that could fit precisely into the neat categories. This is because, as it turns out, few organizations represent neat reflections of the stereotypical type; most are hybridized and reflective of vaster, more encompassing, trends. This is evidenced by one simple fact: *facultas paribus*, they compete with each other for the same work and staff. In a political economic sense, as I cover in depth in the following chapter, this might have been deduced. From an identity production perspective, however, it creates an interesting set of references based on zombie categories. One might be compelled to ask whether there was purity of types back then, whenever that was.

The era when participants entered evaluation was also problematic, although, perhaps with a little more up front thinking, this issue might have been avoided. For participants who made a 40-year career of evaluation, the cumulative memory mediated by the present discourse set the lens for talking about past modes of doing evaluation. It

was difficult to break down the process of becoming. Furthermore, the old timers had been interviewed several times and their stories sounded like those told by family members dozens of times with only minor modifications with each retelling. The identity narrative was scripted, revised, and, at the moment, unreflective, which may be symptomatic of the pervasive nature of the ERD. After two such cases, I decided to modify not the sampling strategy but the set of questions I was asking. I began asking more questions about identity production of those who had not told their story and who were still in the process of constructing that narrative. And, those who began their professional work 40 years ago with a program evaluation and then did not do another afterward could reconstruct only so much of that year or two that did not substantively define their careers. Their self- depictions were vastly different from those of newcomers with less than a decade of being an evaluator. These participants could recall, and articulate with great clarity, tensions they encountered along the way and how they resolved those tensions.

Perhaps most compelling, vis-à-vis the initial sampling plan, was the level of integration of participating program evaluators into the figured world of evaluation. The sample yielded quite a range of experience levels—from guru (or people who had spent their careers doing and teaching about program evaluation and whose professional title was none other than *program evaluator*) to career evaluators (people who have spent more than a decade doing contract evaluation work as their mainstay and who have no plans to move out of the industry to side endeavor-ers (or evaluators whose primary occupation is outside of program evaluation, but who do this kind of work on the side) to

those who entered the field only as a brief detour to those who only recently became or are still becoming evaluators.

Retrospectively, in a looser sense, the sampling plan oriented me in such a way as to develop an understanding across temporal, political, and spatial zones. It provided a useful tool for assembling facts and stories for cross-case comparisons in a rough and distant manner. Furthermore, it helped illustrate some of the differences and perhaps more surprisingly, as I will discuss later, the similarities across sub-sectors of this figured world.

I collected the interviews over a period of roughly 10 months. While I completed most of them over the telephone, I was fortunate enough to complete approximately one-third in person, two via email, and, in one case, a participant was kind enough to audiorecord a sort of self-administered interview. In another case, an old timer said that she does not talk with interviewers, but prefers to write out her responses. Each of these modes of data collection resulted in pros and cons. The one-hour telephone interview with corporate executives and high- ranking government and academic professionals were particularly problematic in terms of rapport-building. For instance, asking strangers who are known professionals in the field of program evaluation about their childhood experiences within 10 minutes of the beginning of our first live conversation meant developing some level of instant trust.

Initially, I came off like a strict academic type who would stick with the facts and reveal little about my purposes or myself. Furthermore, I was asking people to reflect, critically, on their career and academic choices, without warning. This approach became

problematic when I found it difficult to contact participants for follow-up interviews. One wrote in an email response to my request for a follow up, “Great, now I get to hate myself all over again,” although she eventually completed the second interview. Ultimately, with enough gentle coaxing and sharing of initial write ups, I was able to talk with all, but one participant at least twice and some as many as eight times.[2]

Because of these snags that I surmised to be reflective of lack of rapport, I decided to engage in interviews that were more conversational. In fact, during one interview, a participant asked me to pause for a moment so that I could tell her about me. I told her a shortened version of my own program evaluator story and then situated how I got myself into the study. Thereafter, I used a similar way of framing the discussion. I also used the “protocols,” as loose guides and asked open-ended questions. When asked for clarification on my questions, I would tell a little more of my story. In addition, because I was interviewing social scientists, I also decided that revealing some of the theoretical and associated literature background for my particular questions was important. When I asked about the relationship between program evaluation and social position, for instance, I noted the broad patterns and had citations ready.

While the interviews provided a holistic picture and a sense of the relationship between each individual participant’s identity and the broader themes related to the figured world of evaluation, I also wanted to present a deeper illustration of the nuances of identity production. After analyzing data, both within- and across-cases, I asked a sub-sample of three participants to help me develop case studies from their stories. These

[2] One participant passed away during data collection.

included two career program evaluators and one example of a social scientist who entered program evaluation briefly and then returned to basic research. These case studies illustrate the processes of identity production, not as discrete outcomes of analysis and interpretation, but as rich oral histories. The case studies exemplify, and sometimes contend with, the existing theoretical frames of identity production. More importantly, they touch upon the key points as a continuous flow rather than as artificially isolated factors of micro-constructs.

In addition to the 20 key participants, I conducted interviews with a handful of evaluation company managers, representatives of funding organizations, and people who spent a portion of their career doing program evaluation. And, I also used an array of documents to assemble traces of the politics, funding streams, and regimes served by evaluation from archival sources representing different periods. To the extent possible, I collected reports that represented each decade—from the school survey moment to present. Along with the interviews, this process helped me construct an historical précis of institutional shifts in the ERD, as it was interpreted and enacted in evaluation, over time.

Analysis

As Harry Wolcott notes, ethnographic analysis is a process of defining, reducing, and creating the illusion of coherence from large amounts of textual, partial, and often contradictory data (1994). Analysis is not a single, discrete stage of the research, but a recursive process that begins during initial data collection and continues well after data collection is completed (Davies 1999). This does not mean that data ought to be coded

immediately (LeCompte & Schensul 1999), since premature categorization may close off competing or alternative interpretations. I disagree with Davies' (1999) contention that categories be developed before data are collected (although, the shape of knowledge is certainly influenced by several rounds of filtering and selectivity in data collection). Instead, tentative categories are derived *from* the data. During early stages of field research, analysis takes the form of questions that are posed and refined into notes, memoranda, hunches, and emergent concepts (Klingner, Sturges, & Harry 2003). What follows those early stages is a process of follow up data collection; refinement of categories, concepts, and relationships; and founded speculation. Indeed, construct building and data collection are linked inextricably (Agar 1986). Analysis entails maintaining a well-organized, regularly and reflectively reviewed data base, generating and refining categories and themes, and continuing interpretation from the initial data collection through write up.

For this analysis, I combined analytic induction and critical ethnography. Analytic induction, developed in the early 1930s as a means of theory testing using inductive reasoning, involves six not always discernable steps (Hammersley & Atkinson 1995; Klingner, Sturges, & Harry 2003). First, a phenomenon is described and defined followed by an exploratory investigation of a few cases. During the second phase of exploration, potential explanatory features are documented. Next, tentative explanations are offered based on common themes across cases. From this, further cases may be studied to test what Glaser and Strauss (1967) refer to as *in vivo* hypotheses. If the *in vivo* hypotheses do not fit the new cases, either the hypotheses are adjusted or the

boundary conditions are changed (e.g., exclusion of cases that do not fit). During the final step, hypotheses are refined until they can be considered "valid." While theory drove data collection, data and their analysis took precedence over existing theory and, therefore, helped me reformulate analytic models and interpretations. I adjusted this series of steps somewhat to compensate for the fact that the research I conducted is based on life history, which is problematic not only for its partiality (due to memory, positionality, etc.), but because of my own potential to skip over key points (Davies 1999) due to my own ontological positioning. Davies suggests reviewing tape-recorded interviews not only for research content but also to "assess the interaction and how it may be affecting content" (1999). This reflexive addition helped me be more cognizant of my positionality and, therefore, attempt to reflect thoughtfully on the ways in which my biases influenced my interpretations.

The analytic inductive framework, especially the systematic and iterative collection and analysis of data, meshes well with critical ethnographic analysis. While this combined approach might be suspect because of its appearance to bring together a theory-building "model" and a robust theoretical lens that does not attempt to disguise its political positioning, I believe the two complemented each other. While the former emphasizes description and interpretation, the latter offers an interpretive filter for pinpointing interplays between individual identity production as an assertion of agency and the construction of that "subject" through the inscription of powerful discourses and socialization into a global political economic system that, because of its ubiquity, is difficult to spot. A CDA perspective, thus, helps detect the ways in which power is

exercised through both knowledge production and identity production. Indeed, as Shumar makes clear, ignoring power helps to maintain its invisibility and, therefore, engage in the unconscious participation in its deployment (1997).

After audiotaped interviews were transcribed and fieldnotes written up, I offered to share my rendering of these events with the participating evaluator as a kind of "member check." (Guba & Lincoln 1981). This allowed participants to modify or expand upon what they had shared during the interviews. Once refinements were made to the text, I entered transcribed interviews, field notes, and reflections into ATLAS.ti for textual and conceptual analysis. For textual analysis, I read the documents to identify and index salient chunks that seemed interrelated (Dey 1993; LeCompte & Schensul 1999). I continually adjusted category definitions to accommodate my understanding of evaluator identity production, as well as recollections and references that connected to the figured world of program evaluation in education. As patterns became increasingly predictable (if not saturated) and categories more definable, I coded relevant chunks of text. I drew some categories, a priori, directly from interview questions and others stemmed from the data.

Those a priori categories were drawn from Bogdan and Biklen (1992): 1) setting, 2) definition of the situation, 3) depictions of routines, 4) ways of thinking about objects and people, 5) process (including changes over time), 6) regularly occurring activities or behaviors, 7) hallmark events, 8) rules about the ways things are supposed to be done, and 9) relationships, which include official and unofficial social structures and networks. To hold myself "accountable" to remain aware of my ontological influences, I added a

tenth code type from Sanjek: a reflexive "portrayal of the ethnographer's path in conducting fieldwork" (1990).

Most categories, however, "emerged" from careful, repeated readings of the indexed chunks of text. I was also attuned to two other kinds of patterns: 1) that what is not said by evaluators can be as meaningful as what is said (deCerteau 1984) and 2) that cases that do not seem to "fit" established categories must be incorporated into the analysis and may alter category descriptions, if not substantially alter the categorization scheme altogether. Once I was satisfied with the clarification of and refinements to the patterns, I proceeded to deeper, connective interpretations across themes (LeCompte & Schensul 1999).

An important aim of the research was to move beyond individual, descriptive categories to examine the relationships between various categories. Intercategorical analysis tends to involve increasing levels of interpretation and explanation. ATLAS.ti facilitates a graphic display of all codes and their relationships. Selecting a code shows indexed chunks of text that led to the production of the code, and clicking on a relationship node (e.g., *a* is a part of *b*) displays analytic memoranda (entered during the textual analysis phase) that describe how coded data were linked. Most important in the conceptual analysis phase are: 1) a rich set of cross-case comparisons that result in themes or meta-categories, 2) the development of case narratives, and 3) the juxtaposition of themes arising from interviews with the emergent, document-based understanding of evaluation. It is at this level that I interwove identity narratives with the ERD.

Reflections of everyday practice were compared with the life histories. The thematic CDA data bridged the positionality of the speaker and the narrative accounts of individual identity transformation to broader discursive formations and institutional contexts. In other words, I interwove identity narratives portraying the socialization to program evaluator with an analysis of identifiable traces of the ERD. Thus, I constantly reviewed the texts not only for category analysis, but to consider how assumptions ingrained in the procedures I used may have affected the data I gathered and the ways I analyzed them (Bourdieu & Wacquant 1992).

Representation and Ethnographic Position

Positionality was particularly challenging in this study, since participating evaluators are social scientists and since I am, too, a contract program evaluator. Unlike qualitative studies on which I have worked, I did not share that ontological distancing that seems inherent in qualitative research. In this project, therefore, I was both an insider and an outsider. After serving 16 years as a professional contract researcher and program evaluator, I became a critical onlooker.

Drawing from what Bourdieu and Wacquant refer to as *epistemic reflexivity* (1992), I attempted to treat my involvement as the ever-present author both as co-constructor and subject of the interview moments. The reflexively situated, self-scrutinizing ethnographer recognizes his undeniable participation in the physical, sociocultural, and intersubjective ethnographic process and, for that reason, he can no longer write himself out of the story he tells (see, for instance, Fabian 1983, Trouillot 1995, Foley 1995). This epistemic reflexivity was a double-edged sword. I became

increasingly cognizant, along the way, that as the interviews became increasingly conversational, my story, too, was being drawn out. To no small extent, there is a blurry division between researcher position (including its relations to power) and, to be glib, sharing a sense of instant camaraderie as a social scientist in the figured world of program evaluation.

To unfold this facet of positionality a little more, I interviewed and conversed with social scientists using social science theory and method to discuss the ways they produced knowledge and how their identity as researcher came about. Aside from the Carmenesque complications—which are, ironically more natural because they lack a social/knowledge divide, an important, to me, implication is that participants are likely be among the audience of dissemination. In this context, participants were not only co-producers of the understandings that unfolded, but have ongoing and direct access to presentations and publications.

This adds a problematic dimension to the usual, seemingly-straightforward separation between insider and other. The problematic is the elevation of mutual concerns or responsibilities pertaining to confidentiality. Some obvious implications arise, not the least of which include the lack of potential for Castaneda-like embellishments and, more importantly, an increased risk of participating evaluators being identified. As noted in my discussion regarding member checks, in an effort to assuage this concern, I shared write ups of interviews and conversations for review, modification, or omission from the study. To address this concern further, I developed write ups to focus primarily on illustrations that depict sub-groups rather than individuals. In turn,

this enabled me to take liberties in producing composite cases that were based on details from multiple participants.

Potentially vitiating of my premise is the identity-shifting psychological effect of my positioning. As I became a critical cultural studies researcher, I critiqued, reflected, self-authored and labeled my way out of program evaluation. And, as I returned to that field to listen to the life stories of evaluators, many of which sounded like my own, I realized that I was part of something that is not perfect, that is not academically glorious, but that oftentimes provides good feedback that addresses policy and practice issues. So, my rendering of this dissertation is my own personal journey away from and then back to an imperfect home.

Significance of the Study and Organization of the Dissertation

Drawing from the assumption that sociocultural practices, such as contract knowledge production, can reproduce social inequity, this project contributes to research that brings together the ways in which discourses help normalize cultural practice and people contribute actively to these practices (and, therefore, the maintenance of those discourses) as they produce their own identities. Such a description and problematization of a taken-for-granted system of power that maintains the methods by which hegemonic practices are exercised subtly over the subjects of educational reform may help identify gaps in a largely tactical knowledge production process and the development of evaluators.

This study is organized by a gradual narrowing perspective of program evaluation and evaluators. In chapter two, I describe the shift in which social scientists involved in

an early epoch, mostly academics representing elite schools of education, conducted contracted studies of urban school districts as basic research assistants and became full-time evaluators. I also discuss how the 1960s overproduction of potential academics and neoliberalism contributed to the rise of research corporations. In chapter three, I provide a broad overview of the industry, which includes a cursory overview of the scope of the current figured world of evaluation, a summary of its historical prototypes, and a brief excursion into its system logic. A major emphasis in this analysis is the commodification of knowledge production. From there, I delve into chapter four where I depict the relationship between everyday knowledge work in this figured world and identity production. And emphasis is on the ways in which ERD and resources promote certain evaluation knowledge-production practices. In chapter five, I offer an overview of the personal histories of evaluators, tracing their accounts that depict how they became evaluators—childhood, academic pathways, and professional experiences. I provide a *habitus*-oriented analysis of evaluators, which becomes a foundation for a typology of program evaluators in education. In chapter six, I narrow the lens a bit more and recapitulate my depiction of evaluators in the previous chapters with three illustrative case studies. I close in a final section with a few reflective meanderings about potential directions for further study.

Chapter 2: The Birth of the Modern Program Evaluator: Historical Background

This historical analysis is situated in marked shifts in the discourses underlying evaluation and the modes of collecting and using data for planned school change. It indexes by who, for whom, by what means, and according to what motivation evaluation work was done at different times. I hope to capture the rumblings of instrumental evaluation shortly before the U.S. Civil War with the rise of educational research and industrialization. Soon after, these threads were interwoven by the Populist movement, Progressivism, and the rise of Taylorist scientific management. As important as are these major shifts, I will showcase less hallmark uses of education data among principals and superintendents for curricular decision making.

Socializing Educational Science

The use of educational data for program decision-making was already evident in the 1840s. Educators and administrators created and administered their own academic tests widely. For example, in Boston written exams and surveys were distributed regularly by 1845, serving as the basis for personnel decisions, assessments of subject mastery and for the development of curricula. But, this use of data was not limited to local educators. Legislators and lobbyists attempted to secure the federal government's role in overseeing redistribution of resources and, in education in particular, instructional and curricular quality. As part of Reconstruction, the federal government created the Department of Education in 1867 to collect information on schools and teaching that would help states, especially those in the South, establish effective school systems.

During this period, researchers surveyed and observed differential educational access by class and race, making cases for federal intervention into local education. Between 1872 and 1890, the Hoar and Blair Bills, both of which advocated for greater federal oversight in education, proposed federal funding with stipulations for thorough, efficient education with minimal educational standards (Lee 1949). Framers of the Hoar Bill used assessment data to illustrate that "literacy has been largely increased in the southern States in consequence of the events of the war" (National Education 1871) particularly for "colored" children.

Embedded in the Populist movement of the 1890s, the idea of equality of condition was posited as an ideal state of social harmony across social classes. Worker and farmer protests were perceived by gentry as threats to the social order and, possibly, telling of a move toward socialism. To reduce the effects of this threat, federal and state policy would be focused on the reallocation of tangible and cultural assets according to formula that would, supposedly, reduce disparity. This fervor went on hiatus when the Populists were defeated, however. And, instead of direct action, social science would help transform society through mediation and arbitration. Sociology arose in the 1880s with the belief that society could be shaped and engineered through knowledge and the disinterring of social laws. Statistics, as a strategy for social administration, was "deployed during this period to classify people into groups as a monitoring system amidst health epidemics, increases in industrialization, and urbanization" (Popkewitz 2000). Social problems were to be controlled through a joining of social policy, social administration, scientific reason, and moral/religious dimensions. Education evaluation

and research would, purportedly give clarity and direction to school planning (Popkewitz 2000), thus nourishing the roots of the ERD.

But, not all social science repurposed itself in the liberal approach to engineering an education system that could address social conditions. Some veins of social research called into question racial bases of differential economic and educational situations. Goddard, the director of the Vineland Training School for the Feeble-Minded, re-appropriated the Binet IQ test to support a hereditarianism theory of fixed intelligence (House 1993), for instance. Although this was not the only view, it quickly dominated social policy. The instrument was used as a "scientific" justification for mass deportations of "lesser intelligent" immigrants and sterilization laws (e.g., Indiana 1907). More important for the current purposes, this test was used to sort races according to educational need—African American, Hispanic, and Native American students, the argument went, would be better served in separate schools. Some social scientists argued against this genetic explanation for educational achievement differences (Vanderwalker 1898; Boas 1911).

The Progressive Movement and Global Conflict, Part 1

By 1920, schools had become increasingly organized around Taylor's Principles of Scientific Management, which were first touted as a means of increasing wages and lowering costs for railroads in 1910. According to Taylor, two interrelated problems were at the basis of inefficiency: faulty management and a slack workforce. At the core of these was what Taylor felt to be a haphazard apprenticeship method of training. Taylor would use scientific methods to identity the single best way to do any task

(Callahan 1962) by breaking down each task to its smallest component steps and analyzing a task for the most efficient method. This process would be repeated and refined until it could be standardized. Then, a system of training and appraisal could be developed and administered around those standards—individual workers were expected to meet the standards or they faced additional training or punishment. Each piece of the system gave rise to specialized organizational roles—manager, trainer, appraiser, and, quite important to this study, the planning department.

Perhaps as important as the concoction of scientific management was its popularization. Newspapers and magazines offered articles on simplified principles for everyday use—giving rise to efficiency in cooking, cleaning, home economics. Moreover, Emerson published popular pieces that "not only spoke of the efficiency movement in glowing terms but continually presented it as a panacea for the ills of mankind" (Callahan 1962, 25). The movement was linked gradually to religious revival. Progress and improvement could come about only through the elimination of waste and inefficiency. These ideologues coincided with a witch hunt for wasters. Schools became a prime target, since, according to popular media, they had become adept at resisting change. Economists wrote columns demanding that schools "provide evidence of their contribution to society or have their budgets cut" (Callahan 1962, 47-48). The NEA, as well as federal and state governments, joined the bandwagon. Suggestions for improvement included adjusting curricula to better fit the economy, tying operations and salary funding to improved efficiency, encouraging parents to be watchful of schools, and the establishment of efficiency bureaus in school districts. In addition, the National

Council of Education developed a set of "Tests and Standards of Efficiency of Schools and School Systems" (Callahan 1962).

Given the seemingly ubiquitous drive for efficiency, it should be no surprise that management provided the backdrop for developing a standard hierarchy and set of processes to organize teachers and students in space and time according to the tasks they were required to perform in order to achieve and maintain standards in education as they had elsewhere. Oakes contends "businessmen who sat on school boards pushed for its adoption in the operation of schools and school systems" (1985). In many urban districts, teachers were required to employ standardized curricula and group children by age. Principals took on adapted methods of industrial assembly line and office management techniques. Oakes contends that the notion of school failure was born in this context, since it was impossible to have standards without failure (1985).

Comprehensive high schools were established, providing different tracks for different students based on a variety of measures. Some tracks were intended to prepare youth for college and others were geared to vocational development. Psychologists used intelligence and norm-referenced tests to determine how each child might best fit into the system, further establishing a sense of objectivity, efficiency, and meritocracy to the placement procedures. Carroll indicates that, "By the turn of the century, the common school curriculum, which still had the residual trappings of a religious canon, had taken on scientific authority as the normative standard against which all learning was to be measured" (1992).

In 1903, Joseph Rice reported results of mathematics and spelling tests, which he devised and administered throughout the large urban school districts (Callahan 1962). When published, the report fed the debate about the ineffectiveness and inefficiency of American schools. While the report was brought into technical question, it indexes one of the first uses of in educational research of standardized academic skills assessments as an outcome indicator. It also marked what many scholars believe to be the first program evaluation (Madaus, Scriven & Stufflebeam 1983; Travers 1983; House 1993). By 1913, several standardized instruments were used (handwriting, mathematics, and English) in several large urban districts. School surveys were developed to provide a wider scope, as well as an outside *expert* view of the quality of instruction, learning, and curricular. By 1915, the National Association of Directors of Educational Research (later AERA) was formed with the explicit purpose of developing tests and measures.

School Survey Movement

The school survey movement is a precursor to the modern evaluation. Like evaluations, as Caswell (1932) reports, "There is not agreement as to what constitutes an educational survey." A few common characteristics included "comprehensive study and evaluation…of all factors relating to the practices or programs questioned" and "Both immediate and future needs are stated as they are indicated by the status study and research findings" (Caswell 1932, 179). Thus, surveys were holistic, comprehensive census-like studies that measured the "efficiency" of school operations to include facilities, community economic viability, teacher pay and performance, curricular breadth, health and hygiene, and many other facets (Lagemann 1997; Caswell 1929).

By 1931, more than 625 school surveys had been completed in the U.S. (Caswell 1929). Like evaluation, the idea behind the survey movement was that education could be improved in efficiency by the study of "facts" pertaining to everyday schooling practice. A widely-distributed belief was that school surveys would provide the facts needed to construct sound policy and inform practice (Lagemann 2000) like some brands of program evaluation, method outweighed theory in terms of practical significance. Paul Hanus, a pioneer of the movement, argued that the application of exact quantitative measures to school management problems outweighed any need for theory (Hanus 1920).

Thus, scientific discourse was reinforced with two intersecting legitimizing principles. The first of these was the rise of social measurement and its necessary progeny the "outcome" measure (i.e., achievement indicators). Most school surveys included questions about student achievement. This line of thinking was influenced, heavily, by Thorndike, whose "scientific" measurement approach was broadcast widely along with his eugenicist beliefs about "racial betterment" (Apple 1999). With this new scientific approach, statistical analysis quickly replaced historical analysis (Lagemann 1997). Increasing technical specificity of research methods was extended into specific content areas including reading, writing, spelling, arithmetic, handwriting, and history. By 1928, more than 1300 standardized tests had been developed thanks, in part, to the school survey movement (Lagemann 2000). Hanus, in his survey of New York City public school, for instance, used a mathematics standardized test constructed by one of Thorndike's students.

In addition to these, measures borrowed from the budding field of psychology, Taylor's techniques for measuring efficiency in industry, and Booth's 1889-1903 study *Life and Labor of the People of London*, as well as some original methods, provided the bulk of data collected in school surveys (Lagemann 2000). Making this effort practical, palatable, and practicable was the efficiency movement in which progressive educators aspired to be perceived as "efficiency experts." In this context, efficiency was measured in terms of finance, in addition to student subject mastery. The bottom line was a decrease or at the very least no increase in taxes (Callahan 1962). According to Lagemann:

> The school survey movement would not have been possible without tests and statistical devices that allowed researchers to measure the achievement of students and the costs of instruction and, then, through comparative statistical analysis, to determine which practices were apparently most effective, least costly, and therefore most efficient (1997, 6).

The convergence of scientific method and efficiency was held together by several key interests. First, Tyack contends that the school survey was a vehicle for transmitting the program of the administrative progressives—who believed that public schools' primary purpose was to prepare students for real-world tasks (i.e., the work force) (1974). At the same time, the popularization of the school survey movement went hand-in-hand with discussions about national standards. More specifically, publications upheld school surveys as a mechanism and an indicator of the readiness of America for standards for all

students (except for the "mentally handicapped") (Caswell 1932). According to Travers (1983), while the Federal Bureau of Education wrestled with the establishment of standards of achievement, it eventually dropped the idea while, instead, focusing on standards related to curriculum content, time on task, and curricular range. In addition, the Bureau became engaged in discussions about teacher quality, which gave rise to discussions among chief state school officers about the use of a standardized test for teacher certification (1983).

Costing more than $48,000 was the one-year Cleveland survey, which began in 1915. Initiated by the Cleveland Engineering Society, that alleged the school system was operating at less than 50% efficiency, the Survey Committee of the Cleveland Foundation was assembled to carry out the study (Ayres 1917).[3] Ayres led the committee, which was comprised of 30 external consultants (most of whom were faculty or students of the University of Chicago (Lagemann 2000). Most held advanced degrees (22 had PhDs) in school administration and psychology and had been trained in the rising body of educational literature (e.g., Thorndike), research methods, and efficiency (Ayres 1917). This figure's importance might be highlighted by the fact that at the time, advanced degree-granting intuitions in education were beginning to take root (0 in 1910, 34 in 1920 and 53 in 1930) (Ayres 1917). In addition, some committee members were superintendents of other urban school districts. School surveys both provided work for this expanding field and provided empirical fodder for the expanding body of education

[3] Lagemann contends that the survey was initiated by the Cleveland Foundation because of dropouts, but, Ayres' 400-page report is clear about the rationale and the formation of the Foundation AFTER the survey began.

research. The committee published more than 25 individual survey reports, which it made available to the public through sales and public meetings (Ayres 1917).

The Cleveland survey identified "10 principal factors in addition to the characteristics of the children themselves that determine the quality of results and efficiency of work of a school system" (Ayres 1917). The committee concluded that:

> The business management is honest and efficient. The teaching staff is of inherently good quality. The school plant is of exceptionally high grade. The community genuinely desires good schools. With respect to the other six factors, the findings of the survey have recommended readjustments that would, in the opinion of the staff, greatly strengthen the school system of the city and largely increase its effectiveness for community service (Ayres 1917).

Tied to the vocational focus of the times and operating from the "fundamental principle of the curriculum should be that effective teaching is preparation for adult life through participation in the activities of life," the committee made several recommendations for improving the curricula. For instance, "Work in spelling needs further modernization through concentration on the words most frequently used in adult life and through the development of a habit of watchfulness over spelling during the process of writing" (Ayres 1917). Furthermore, an assimilationist agenda was advanced repeatedly in hopes that the curriculum would aid this process:

> Especially in the lower grades it is evident that the reading matter supplied to foreign children is often ill adapted to their needs. Attention to the needs of

> foreign children ought to bring into the schools more studies of a social type which will acquaint all the children with the organization of the city and the duty of the individual to the community (Ayres 1917).

In summary, and importantly for the purposes of this study, the Cleveland survey reinforced the "need" for external experts and a division of labor between administrators and general educators (Lagemann 2000).

According to Callahan (1962), external experts were hired to conduct school surveys because of the belief that they would have no interest in local politics. According to Lagemann, various foundations (e.g., Rockefeller, Russell Sage, and Carnegie) urged school authorities to consult with external experts for assistance (1997). Furthermore, the external survey became a method for district administrators to avoid the sometimes damaging school surveys imposed upon them by such organizations as the National Manufacturers Association. Not long after the outset of the school survey movement, external measurement expertise and a non-biased stance became normalized practice in the assessment of educational change.

Within a few years of the Cleveland survey, a prominent educational psychologist, Charles H. Judd, led the school survey of Saint Louis, MO with a team of 14 assistants, also comprised of faculty and students. The survey was initiated by the state's Board of Education for the purposes of "securing a body of facts for presentation to the public, primarily to aid in the passage of a bond issue of $3,000,000" (Judd, et al 1918). With a cost of $9,780, the single-year survey examined overall elementary, special, and high school education, as well as the physical plant. This team was made up

of various Superintendents, Commissioners, professors, students and specialists (health, architecture). Unlike the Cleveland study, professors came from a wide-range of institutions including Harvard, Cambridge, and Wisconsin, although most represented the University of Chicago.

The survey committee found the Saint Louis School District to be a "well organized and efficient system" while pointing to several "duplications and incoordinations in a number of respects" (Judd 1918). The survey staff contended, ultimately (and expectedly, given the funding source), that the schools in Saint Louis were deserving of bond funds to improve their facilities and processes. A concern, however, was that the elementary schools were wasting money on "fads and innovations which do not have the sanction of educational experience" (Judd 1918). The problem, according to Judd and his team, is the misallocation of funding, especially for "special classes for defectives and those who are below physical normality. This would mean a regression in public hygiene that an enlightened citizenship would regard as a misfortune" (1918, 12). The team offered a clear set of choices—either more funding or turn these "defectives" to the streets. This perspective, not surprisingly, matched that of the funders.

Important is a set of detailed descriptive tables that compared the average classroom size and building age in various urban districts, underscoring that Saint Louis is not providing what it ought to provide its public. It is important to note that while secondary sources that describe these surveys mention the use of statistics, neither the Cleveland nor the Saint Louis reports mentioned them, much less provide analytic tables.

Indeed, the only "statistics" is a single table that reports the median age of students. The inflation of evidence is key since these authors seem to promote an image of educational research that has "always" relied on statistics.

While Lagemann (2000) offers a somewhat static view of the movement in which experts maintained their expert status, the School Survey Movement went through what appears to me as three phases. Initial surveys, beginning with Boise, Idaho in 1910, were used to identify weaknesses in educational operations and were "conceived as a method by which experts outside the school system checked up on the administrators and teachers" (Caswell 1929, 111). While a professor at Chicago, Judd conducted surveys for school systems for between $2,000 and $10,000 (equal to his annual salary at Chicago). Other academics also participated. This work soon found a public venue in journals such as *Elementary School Teacher* and the *Educational Review*. Like program evaluation at present, the focus of the journal articles was almost always on method and tactic rather than results. Important to note here is that the School Survey Movement represented a wide range of methodological approaches and levels of technical expertise.

As reports describing the measurement process were published and distributed widely, the second phase enabled community groups, foundations, government agencies, and special survey commissions (usually comprised of local civil leaders) to conduct their own school surveys. Caswell points out, "it was of little importance who made the measurements just so long as they were made" (Caswell 1929, 111). But, without

"expert" guidance, interpreting results, developing plans and implementing change were severely hampered.[4]

The third phase, in what appears to be a dialectical synthesis, "includes both the idea of the application of a group of administrative methods of measurement to determine the status of a given problem in an individual school system and service of an expert in that particular field to aid in the solution of the problem" (Caswell 1929, 112). Paul Hanus, for example, was a champion of this approach in his survey of Montclair, New Jersey. While this marriage of views is conceivable, a more critical lens would question whether methods could have been made more accessible to educators. It is important to recall here the academic dominance of scholarly publishing (Bourdieu 1977).

Over its lifespan, the School Survey Movement was enacted through the efforts of a constellation of methodological specialists and educators including university professors, administrative progressives in urban school districts, reformers from civil organizations, foundations, and the U.S. Bureau of Education. In its lattermost phase, the surveys were initiated by community groups or local education administrators. The Federal Bureau of Education picked up the idea of the school survey and positioned itself at the center of external surveys between 1910 and 1920 (Travers 1983). By 1916, it had become a massive effort of the Bureau as it administered statewide surveys in Arizona, Colorado, Delaware, Iowa, North Dakota, Tennessee, Washington and Wyoming. The Bureau also conducted system-wide surveys for several major cities, itself to fulfill its mandate to collect "whatever statistics necessary to maintain efficient schools." The

[4] Much like the second wave of program evaluation labeled by Fetterman as *empowerment evaluation*.

Bureau represented itself as an impartial outsider. Again, this is similar to the current few decades in terms of program evaluation.

Program evaluation, school evaluation and school surveys represent similar types of applied research that uses social scientific methodologies to describe the implementation and impact of educational reform efforts for the purposes of refinement, funding, expansion, and so on within a constantly emerging ERD. All these have philosophical foundations in empirical positivism, but unlike basic research, they are, foremost, pragmatic and rely on the combination of operationalism (which offers, through reduction and essentialization, replicability) and parsimony (a preference for the simplest of available adequate explanations for behavioral phenomena). In other words, through systematic research of observable, presumably objective and predictable chains of events, these kinds of analysis offer pragmatic (i.e., useful) decision-making information. Scientific and technical talk in advanced capitalist society has greater legitimacy than does ethical talk (Foucault 1977; Apple 1978). Social justice, for instance, cannot be oprationalized with an input/output perspective. Yet, the ERD is clearly couched, as mentioned in Chapter 1, in tenets of liberal progressivism.

As Callahan argues, the School Survey Movement contributed to the adoption of standardized tests and teacher-rating procedures (1962). While both of these have serious implications for teacher work, they also, especially when combined with the focus on finance, have direct and severe impacts on administrators, particularly those of higher rank. Furthermore, school surveys set into motion the cooptation of educator time for painstakingly documenting or accounting for every moment, which competed with (and

still does) teaching time. Unlike "pure" empirical research, the three veins of applied ERD evaluation are put into policy and administrative practice. According to Moley (1923), for instance, more than 75% of the recommendations from the Cleveland survey were put into effect. While the school surveys may have impacted operations in many urban school districts, the fundamental shift in knowledge/power from that of do-it-yourself surveys to external expert meant that educators would come to rely on knowledge brokers to help them understand results. Tyack, for instance, reports one prominent educator saying the findings "cannot be interpreted and no one with the least grain of sense would attempt to read [the report]" (1962).

School surveys also permitted comparability—broad similarities in educational issues—by fixing some factors considered important to educational processes. School surveys conducted in Michigan (Grand Rapids), Cleveland, and Indianapolis provided valuable insights into student learning. For instance, Gray's Indianapolis and Judd's Cleveland surveys both studied the amount of time teachers devoted to reading (Travers 1983). While Gray found that students read less than 30 minutes per day in the classroom by the time they reached upper primary and middle grades, Judd found that Cleveland students read nearly two hours each day. When correlated with reading achievement, the outcome differences underscored the need for additional reading time.

Another contribution of the School Survey Movement was laying out, step-by-step, methodologies and procedures that researchers used. In many cases, the designs were so well described that others were able to replicate the processes in their school districts (if done internally) or for their contracts (if done by an external expert). For

instance, Gray's (1917) Indianapolis school survey was greatly influenced by Judd's write up of his 1915 school survey of Cleveland, which he published in the *Fifteenth Yearbook of the National Society for the Study of Education*. This, in turn, provided insights into the development of tests and measures and achievement testing—both surveys focused on student achievement tied to teacher learning objectives. The replication of earlier prototypes, therefore, is not tantamount to adhering to a paragon as much as it is to creating an illusion of powerful scientific practice through repetition, if not habituation.

Published survey reports were used as texts for the inchoate and growing educational administration and research fields. According to Caswell (1929), this had a major impact on course offerings in education in 12 top universities (from 132 in 1900 to 1636 by 1930). Furthermore, in his Indianapolis survey, Gray correlated socioeconomic variables with reading achievement. It is important to note that, during this era, genetic determinism, intelligence, and socioeconomics were linked. Gray, however, along with Judd, followed the belief that intelligence tests reflected only the immediate level of functioning of the individual student. Thus, these school surveys may have contributed to the rise of the sociology of education, as both an implement of the ERD and as a critique of the status quo.

Finally, as Lagemann argues, the increasing status as external expert effectively isolated education academics from academics of other disciplines (1997) while creating a niche for them among university faculty (Lagemann 2000). Because educationists located in higher education spent much of their time collaborating with superintendents

and other educators, as well as gathering empirical, applied data, there was limited elaboration of theory (Lagemann 2000). Clearly, their chosen constituency would be the public and their goal would be to address the demands for efficiency. One implication of that was a split between the previously recognized subject-area experts from arts, sciences and humanities who had dominated national curriculum committees until about 1910 (Lagemann 1997). Faculty from education began replacing their non-educationist peers on these committees.

According to Lagemann (2000), the "scientific study" of learning was made possible by the School Survey Movement because it enabled wide testing of students using standardized tests and intelligence measures. While educational psychologists were called upon to aid in administering and analyzing these data, the American Psychological Association began pressuring psychologists to move beyond testing as a "mere practical device" (Terman, APA president cited in Lagemann 2000). Psychologists were urged, instead, to examine race differences, mental growth, and so on. Further supporting this move, AERA (American Educational Research Association) became an "expert only" society in 1921. Yet, despite this called-upon shift in emphasis, many scholars continued to work on studies like school surveys through the 1960s according to Kliebard (1999) and, as I would argue, continue today.

Global Conflict Part 2, Economic Depression, and Social Programming

In some ways, the various pieces of school research—internal evaluation and accountability, external expert evaluation, research and development contract work, and academic education research—fit into the ERD as overlapping facets. Yet, each piece

increasingly filled its own niche. Indeed, the School Survey Movement helped solidify research about and for educational reform. Although it is often touted as a result from the legendary piece of legislature, the term "program evaluation" predates the Elementary and Secondary Education Act (ESEA) by nearly half a century.

Several hundred articles and technical reports on program evaluation indexed in the Education Resources Information Center (ERIC) were written between 1900 and the 1930s. Take, for instance, this title: "Identifying the Effective Instructor: A Review of the Quantitative Studies: 1900-1952" (written in 1953), an expected bridging of scientific management adapted to employee appraisal and student learning. Another report recommends methods for improving a geometry curriculum: "A General Survey of Progress in the Last Twenty-Five Years. National Council of Teachers of Mathematics (1926)." These works cover a wide range of topics, purposes and methods including military training, racial attitudes toward specific curricula, reviews of testing instruments, curricula for "disadvantaged" youth, specific content area evaluations, city-wide program evaluations, school facility evaluation and ways to perform evaluations (intended for district and school personnel).

Ralph Tyler, who began his writing in the late 1930s, is often credited with being the founder of evaluation in education (House 1993). Tyler used an objectives-oriented approach in which ones states behavioral objectives, measures, and then determines whether the objectives are attained. He viewed curriculum development as a cycle—the quality and impact of curricula are monitored by carefully observing the outcomes. Then, data from these observations are used to fine-tune the curricula. Clearly, Tyler

drew from what had become institutionalized practice in industry and military and drew heavily on the emergent body of evaluation literature. This also marks a shift from evaluation as local process to its external expert status. This also marks a shift in how research would be used. More specifically, the refinements to experimental designs shifted emphasis from educating the public about social problems and their solutions according to scientific management principles to leading the public by brokering highly technical research (Manicas 1987). Behaviorism became the dominant paradigm leading to the development of Iowa statewide testing in 1929 (soon emulated in other states). Several private and philanthropic organizations funded evaluations (Ford and Rockefeller funded evaluation studies while organizations such as ETS supplied the outcome measures). Thus, increasingly, the power to assess, document, and make sense of reform was withdrawn from local educators and became the financial and hegemonic mainstay of external entities.

In line with the struggle to relieve local educators of the power to evaluate, Federal involvement in education also persisted. With NYA, education commissioner Studebaker conducted small-scale evaluations to demonstrate the need for education to concentrate on job preparation. In line with the neoefficiency version of ERD, he also urged superintendents to reorganize funding and curricula to become more efficient. Thus, some of the modern threads began to coalesce. There are, at once, elements of Taylorism and Tylerism (e.g., timing for vocational instruction sequences), as well as use of data to improve student learning beyond efficiency. But, it would be naïve to assume these were the sole drivers for Federal involvement. Key organizations

were intent on increasing the government's role and eliciting policy that would secure positions that would support those organizations' agendas.

The NEA, for instance, which had begun lobbying before WWI, became a major advocate during WWII of the idea that federal funding be introduced into education, especially with regard to vocational education, aid for "educationally disadvantaged" children, and funding for educational research and evaluation. Federal support did come, but at first, it was used to ameliorate effects directly attributable to the federal government itself. The Lanham Act in 1941 and the Impact Aid laws of 1950 eased the burden on communities affected by the presence of military and other Federal installations by making payments to school districts. And in 1944, the GI Bill authorized postsecondary education assistance that would ultimately enable millions of veterans to attend college. Not long after, Federal involvement expanded into a greater role in school reform.

Until the 1950s, educational reform rested largely with state and local officials. Program evaluations were conducted independently, by in-house agencies (small scale), or under foundation contracts (e.g., Ford). But, a decade before federally-required program evaluations, the Cooperative Research Act of 1954 had its basis in the premise that scientific method applied to education would lead to great improvements in educational reforms. Relatively small-scale evaluations were conducted as part of curriculum development and teacher training. Thus, evaluation emerged as a distinct area of *professional* practice in the post-WWII years in North America. Three strands that were most important were the scientific evaluation of educational innovations (e.g. the

effectiveness of new curricula in schools); linking evaluation to resource allocation (e.g. through a Planning, Programming and Budgeting system—cost-benefit and economic appraisal methods) and the evaluation of anti-poverty programs (e.g. the Great Society experiments). A liberal ideology in the U.S. supported beliefs that economic growth would be infinite and that poverty and thee effects of social class on educational attainment would be eradicated (House 1993). This notion of continual expansion was shared by Truman, Eisenhower, Kennedy, and Johnson. Accordingly, social problems could be resolved as were industrial ones:

> The problem is first identified; programs are designed to solve it by government enlightened by social science; money and other resources…are then applied to the problem as "inputs"; the outputs are predictable; the problems will be solved (Hodgson 1978, 76).

To help administer that growth, evaluation was touted as a means to determine, scientifically, which programs maximized desired outcomes and which did not.

Concurrently, education research had become a discipline with experts in content and methodology with "objective" tools. According to Watkins, this professionalization displaced teachers, principals, and superintendents who based planning and curricular changes on their own research and evaluation (2001). Various social sciences were involved in evaluation studies during the 1940s and 1950s. One how-to article from 1959, for instance, comments, "Evaluation research must use measurement techniques and instruments (some borrowed from the fields of psychology, sociology, and

anthropology); after-only design may be adequate with clear-cut dependent variables but before-and-after design is more rigorous." Ample evidence also suggests the Office of Education was conducting in-house program evaluations during the 1930s, 40s and 50s beginning with NYA programs (although they too were critiqued for their lack of technical expertise). The rise of the scientific method for social administration as a fundamental facet of the ERD created a hierarchy of practitioner and researcher, effectively silencing teachers and other educators (Popkewitz 2001), which also laying the foundation for a lucrative industry.

The NeoModern Era

Evaluation's genealogical prototypes in the U.S. include the inchoate educational science of the late 1800s, the adolescent social sciences, the rise of efficiency experts and research contractors in large urban school districts during the early 1900s, and the proliferation of standardized testing instruments during the 1930s (Travers 1983; House 1993). Referred to by Guba and Lincoln (1989) as *first generation* evaluators, work consisted of school measures and intelligence tests of reform effectiveness. Evaluation remained a locally-initiated and funded practice undertaken by university faculty and foundations (e.g., Ford and Spencer) until Federal involvement in education increased during the Cold War. Evaluators could determine a program's inner and contextual values. One logical way to operationalize federally-initiated school change was to use scientific methods to measure change so as to maximize desired outcomes. As mentioned above, evaluations, in the form of surveys and site visits by federal agents, became federally mandated with the passage of the Cooperative Research Act of 1954. By the

mid-1960s, it became increasingly clear that, with expansion of War on Poverty programs, federal agencies lacked the internal capacity to conduct evaluations. Furthermore, public reports of program developers' self-aggrandizing evaluation results raised concerns about bias.

The War on Poverty and ESEA

The Spencer Foundation was established in 1962 as a spinoff of Spencer's firm, Science Research Associates (SRA). That same year, Spencer, as President of SRA, testified before Congress and made explicit his belief in the reformative power of well-funded education research: "In my judgment, hard-minded, sensible investments in educational research can provide the most effective single method of strengthening our schools." In the same year, the federal government began isolated large-scale funding for program evaluations. Before that, evaluation had been considered a "cottage industry" focused on small programs and was usually funded by social agencies or foundations (Weiss 1987) or was conducted by school districts as internal improvement efforts. This new funding stream for evaluations was tied, almost inextricably, to social programming for the disadvantaged (Weiss 1987). The first federal program *requiring* an evaluation was the juvenile delinquency program enacted by Congress in 1962, setting the tone and framework, even if only skeletal, for subsequent evaluations. This thinking about evaluation was voiced in Mark F. Kennedy's wish to "do something about the problems of inner city youth and asking for facts and figures to support a program" (Chelimsky 1991).

No small thanks go to the U.S. Department of Defense for putting program evaluation on the Office of Education's agenda. During the 1950s and 1960s, the Department of Defense had developed mechanisms and quantitative processes to guide program budgeting and resource allocation. Many HEW directors and key staff members had prior military or military contract experience (Abert 1979). This was leading up to big plans. In the midst of Johnson's War on Poverty emerged a new type of public professional equipped to deal with social "ills" and deficiencies arising in the new post-industrial political landscape. Large-scale federal programs to ameliorate poverty, provide counseling for families, and educate a viable workforce employed thousands of professionals. At the nexus was an effort to hold educators accountable for educating children living in poverty.

The deployment of external evaluation became solidified in the early 1960s following the widespread argument that psychological deficit explanations of poverty were perpetuated by local use of flow-through funding (originating at the federal level). Subsequently, federal funding was redirected to community groups, instead of school districts, with the stated intention of changing social conditions (with little or no mention of the budding corporate parks lobbying for and benefitting from these policies). External evaluation, it was argued, would then test the effects of the programs that were funded, revealing what worked and what needed refinement. Weiss explains the logic behind the discourse: "Once systematic and objective data were available … policymakers would use the information to improve the effectiveness of programs and thereby improve the lot of the poor" (1987).

In 1965, Johnson launched the Elementary and Secondary Education Act (ESEA),[5] a highly unusual piece of legislation for its scope since it was signed into law only 89 days after it was introduced into Congress. Developed under the premise that children from low-income homes required more educational services than did children from affluent ones, K-12 schools were funded to provide guidance and develop compensatory programs for low SES students. ESEA also represented the first major piece of federal social legislation that included a controversial mandate for evaluation stated in general language drawing heavily from the Cooperative Research Act (1954).

Senator Kennedy held back the Bill until evaluation was added. According to House, Kennedy believed educators would not use feed money for disadvantaged children (1993) and that educators were to blame for the low achievement of students. Accordingly, "his remedy was to force the schools to provide test scores...to parents in the belief that the parents could then monitor the performance of their schools and bring pressure to the schools to improve" (Ibid, 17). Others supported the Bill because of its promise to help identify effective practices and pathways to wide-sweeping educational reforms and under the auspices and "expert" guidance of the Federal government plus its selected contractors.

The evaluation requirement had three explicit purposes: 1) to identify the most efficient approaches to educating disadvantaged students; 2) to arm parents of disadvantaged children and communities with facts that would empower them to push for

[5] ESEA also appropriated funding for 12 Regional Education Laboratories and 8 university-based research centers to conduct basic and developmental research in education (Gordon & Gordon 1979).

better education and, thereby, enable open debate; 3) to use information on programs and their effectiveness as a means of upgrading schools believing that performance comparisons in evaluations could be used to encourage schools to improve student outcomes. The then Commissioner of education, Keppel, believed that public reporting would also stimulate competition which, in turn, would improve performance. Interestingly and importantly, all these views required "objective" measures and assumed that people would make decisions based on those data.

ESEA also required the Office of Education to develop evaluation standards and models for state and local agencies, and required the Office to provide technical assistance so that comparable data would be available nationwide, exemplary programs could be identified, and evaluation results could be disseminated (Abert 1979). The Office of Education had only 30 days to arrive at these guidelines and standards (House 1978). Among the evaluation models deemed acceptable were the "Planning-Programming-Budget System, the Performance-Based Teacher Education, the National Assessment of Educational Progress, accountability, performance contracting, and behavioral objectives," methods which overlooked the political and cultural complexities of the education contexts (Eddy 1976). Resulting evaluations tended to be large-scale involving, such as did Head Start, hundreds of sites at once as well as thousands of smaller agency-based evaluations for specific programs.

An important source of information about the coming together of HEW's evaluation unit is former Secretary of Health, Education, and Welfare, James Abert's, three-volume memoir, Program Evaluation at HEW: Research Versus Reality (1979).

The set amounts to a critical history of the agency's development from the 1967 discussion about evaluation to its full implementation after appropriations were made by the Congress. Ten years after the evaluation mandate, Abert provided details from meetings, documents, and financial reports to describe the micropolitics of decision-making within the federal government with regard to program evaluation.

Before 1969, HEW formed an under-funded "embryonic evaluation management organization" (Abert 1979). HEW's program evaluation work began August 8, 1967, when David Gorham, the then assistant secretary for program coordination, sent a memorandum to the secretary requesting that an apparatus be established to link funds to review of programs. Several months later, agency heads met with Gorham to discuss the planning problems throughout HEW, the difficulty of completing objective evaluations within agencies, the lack of technical methodological training among current staff for completing evaluations, and the lack of reliable information to aid decision-making. When the plan materialized in 1968, it called for the establishment of explicit objectives for evaluation, evaluating the results of evaluation activities in terms of the objectives, and funding and review of evaluations according to an annual cycle.

The next step in this multi-layered blueprint for measuring program effectiveness was to determine how funding would be allocated and administered (Abert 1979). Two plans were tabled. The first called for using the office of the secretary to manage spending and make decisions. The second called for giving relative autonomy to each unit in the Education Department, permit the units some authority in decision-making, and accept what they said needed to be done. A key argument was that while the first

plan would target evaluations to the needs of top policy officials and the president, the latter would have a greater bearing on entitlement spending and findings would be more likely to be useful not only to policy officials, but other agencies and the private sector (Abert 1979). The second plan was adopted.

Oversight would become the responsibility of Assistant Secretary of Planning and Evaluation (ASPE). ASPE would help agencies develop evaluation plans and require each to list its programs, detail the objectives of the programs, and plan contract and in-house evaluations. Indeed, 75% of the operating evaluation budget went to individual units and the remaining 25% went to fund ASPE—a breakdown, which according to Abert was completely arbitrary, but not publicly questioned (1979). ASPE would serve as broker for evaluation findings, a provider of independent evaluations of evaluations (a sort of internal affairs?), orchestrate cross-agency evaluations, and take on the efforts that were too controversial for other agencies. Its most important role, however, was to examine problems in terms of cost-benefit.

ASPE also took on the responsibility of developing an annual Evaluation Plan for HEW and preparing the Evaluation Digest to show the results of federal evaluations. These documents were used to create procedures for the selection of contractors, monitor evaluation projects, and document activities. This initial setup was not expected to be perfect. Abert comments, "Better targeting was expected to come later, partially when there was staff to achieve it, but more importantly, when one knew with more precision what better targeting meant" (1979). ASPE program workers soon realized that there no common measure of output across programs, which stymied their efforts. The

Department hired General Electric to conduct a cost-benefit analysis of the Office of Education to determine what could be adjusted to provide for common, measurable outputs. School data seemed to be the persistent problem and ASPE staff reasoned that evaluations should be centralized. As a result, ASPE began conducting national surveys in addition to the on-site evaluations.

Of course, the evaluation mandate did not go uncontested. For one thing, different viewpoints surfaced regarding the purposes of the evaluation requirement. An underlying similarity of these, however, was the expectation that evaluation, as a central component for understanding and monitoring change, would identify what worked and, therefore, lead to successful reform. There was also a common assumption that evaluation activities would generate objective, reliable, and useful reports, and that findings would be used as the basis of decision-making and improvement. But, in 1966, the Coleman Report argued that school improvements (i.e., higher quality of teachers and curricula, facilities, and compensatory education) had only a modest impact on student achievement. The report demonstrated that the only variable that consistently correlated with student achievement was the socio-economic status of the student's family (1966). Shortly after the report and continuing through the mid-1970s, educational reform programs targeting low-income students sprang up throughout the U.S.

Despite contention and disagreement about the purpose of evaluation, the ERD persisted. In 1969, as the Office of Education began its implementation of the ESEA mandate, the then Secretary of HEW, Finch stated before the House Committee on Education and Labor "Evaluation will provide information to strengthen weak programs

and drop those which simply are not fulfilling the objectives intended by the Congress when the programs were enacted" (1969). The War on Poverty, unlike previous attempts at reform—in the federal government's and popular media's memory—would learn from mistakes and successes. But, negative findings seldom led to programs being shut down and positive findings seldom saved programs. Weiss, an evaluator cum Harvard professor, wrote:

> At the start of the Great Society, we expected rational thought to sweep away the pathologies of behavior that were rooted in the past. Rational thought in programming would design activities that would enable the poor to enter the mainstream of middle-class America. Rational thought through evaluation would provide direction to expand, modify, or terminate program activities and thus raise their level of competence. Rational thought through economic guidance would keep the economy in a state of steady expansion (1987, 45).

Negative findings did affect overall funding, however. When the 1968 surveys showed no significant gains, senior HEW staff further reasoned that educators did not know how to teach (House 1979). Furthermore, the 1969 surveys have never been released to the public. Nixon used the findings to reduce Title I funding. But, more importantly, a senior ASPE staff member (Rivlin), argued that social service programs needed to be reorganized so that they could answer evaluation questions (House 1979). House argues, and I agree with him, that "In the Rivlin philosophy, there was a parallel between producing social services and manufactured goods" (1979, 29). It required

strong, centralized scientific management and a logic chain as such—the only true knowledge is a production function specifying stable relationships between input and output, which is discoverable only through experimental method. Of course, if the research design is the primary driver for program organization and delivery, one might expect greater "impacts."

In the coming years, ASPE directed several large-scale evaluations including Follow Through, which turned out to be nightmare for all involved. SRI was awarded the contract, but when, after two years, no significant results were found, the Office of Education blamed SRI for using nonstandardized results and for identifying too many contextual variables (although this was part of the original proposed scope of work). Then, ASPE hired the Huron Institute to continue the study with a focus on outcomes and not the political context. After a year of strict focus on outcomes and OE-selected instruments, parents revolted claiming they were supposed to have a greater role. Abt Associates took over the contract from 1974-77 and were unable to isolate good predictors from site to site. Thus, after millions of dollars and a decade of evaluation, ASPE was unable to identify a method for producing a consistent outcome measure. House contends that this search for a few outcomes had two unintended effects. It removed local politics and parents from the grid and, as had been Rivlin's plan, it reduced the complexity of the individual reform models to facilitate measurement.

A note on how this planning function was played out is important. Overall, a few internal staff members trained in a social science conducted the planning and monitoring. For instance, for elementary and secondary education five professionals acted as project

officers (each overseeing 2 contracts each year), monitored program implementation, and secured and reviewed program evaluation reports (Abert 1979). Once the reports were received, staff disseminated findings to "relevant decision makers and through executive summaries, policy implications memoranda and informal means" (Abert 1979). Furthermore, the executive summary was distributed to Congress, OMB, key officials, "chief state school officers, regional commissioners, and selected interest groups" (Ibid), as well as to ERIC.

During this initial setup, another key decision was that evaluations would be conducted at all schools receiving federal program funds. This stemmed largely from debates about being singled out if only some were selected. Furthermore, Abert comments, "It was felt that the sincerity and credibility of the evaluation program would 'sell' better in house if virtually all programs were subject to the evaluation requirement" (1979). This ESEA Rider became effective in 1971. Gradually, evaluation funding increased, a fact tied to the GAO and Congress "insisting that agencies produce real evidence of the effectiveness of their programs" (Abert & Kamrass 1974). In 1968, the Office of Education requested $2.5 million (half was appropriated) and in 1969 while the requested amount was increased to $6 million only $1.5 million was appropriated. But, in 1970, Congress was more receptive, forking over $9.5 million for evaluation, although the appropriation came late in the year (September).

The appropriations beginning in the 1970s, however, become a bit misleading. Not included in the allocations were set-asides for planning and evaluation, which included nearly six million dollars each year (1971-1973). Planning and evaluation funds

covered nearly $1 million a year for educational policy research centers and consultant set-asides were often in excess of two-thirds of a million dollars. Thus, the real operating budget available to the Office of Education for evaluations hovered around $6 million between 1971 and 1973. In addition to these, special programs, such as Experimental Schools, received separate line-item funding.

But, not the entire evaluation mandate revolved around education. Across the federal government, program evaluation resurfaced in a greatly modified form of the military survey that was popularized from the 1930s through the Second World War. Administrators within the executive branch of the federal government were pressured to legitimize and document their actions. This gave rise to specialized units within the GAO and the installation of state-level evaluation units to carry out this set of functions. It also, however, created the infrastructure required for federal flow through funds and the apparatus that would enable the expansion of external contracting for evaluation services. Furthermore, it created a context in which there was enough similarity in the kinds of data sought for each agency program that a pool of consultants and full-time contract workers could redefine their expert subject-area status easily.

Table 3. Appropriations for Evaluation in the Office of Education (FY 1968-73) in thousands.

Category	**1969**	**1970**	**1971**	**1972**	**1973**
Elementary & Secondary Ed	200	3385	3825	3825	2075
Education for the handicapped	0	425	550	550	550
Vocational & Adult Ed	0	900	900	900	900
Higher Ed	0	900	900	900	900
Ed for PD	0	477	1000	1000	800
Libraries & Ed Communications	0	189	400	400	400
Research & Development	1250	2796	4900	3650	4580
TOTAL	1450	9512	12475	11225	10205

Throughout the 1960s, the federal government gradually increased its role in conducting, funding, and overseeing program evaluations. And, perennially, throughout the decade, it sought justifications for the lack of desired program impact. Critical evaluator who is a professor at Harvard, Carol Weiss comments that in the 1960s "We had signed on as evaluators with the intent of contributing to the improvement of social programming, but we seemed to wind up giving aid and comfort to the barbarians" (1987, 41). Furthermore, widespread support for evaluation did not materialize at the local level. Rather than finding accessible and liberating evaluation results, the local communities grew concerned that federal requirements for reporting would eventually lead to more federal control over schooling.

The 1960s also index some infrastructural and institutional shifts for the industry and figured world of program evaluation. The ESEA evaluation mandate led to thousands of evaluations, which made it difficult for university-based and other part-time researchers to keep up with "tight schedules and large-scale research management

demands that the new evaluations required, so new research organizations were established and some old research organizations such as Rand and Systems Development Corporation, went into the evaluation business" (Weiss 1987). Before 1962, there were no journals, courses, or professional organizations for program evaluation. The first book on evaluation during this period was published in 1967 by Suchman (Evaluation Research). More important to the field was Campbell and Stanley's 1963 groundbreaking book titled Experimental and Quasi-Experimental Designs for Research, which purportedly provided guidance on conducting scientific approaches to determining which programs worked and which did not work. Furthermore, at least six evaluation journals emerged in the mid-1960s. One article in the first issue of Journal of Evaluation and Program Planning argues that because the field had an infrastructure, abstract principles and a body of specialized knowledge, was valued by the public and government, and showed signs of altruistic service to the community, it was fast becoming a discipline (Morell & Flaherty 1978).

Between the late 1960s and the mid-1970s, only a few evaluations were conducted according to Congressional directives. More often than not, program evaluations were completed to satisfy the mandate. In the early 1970s, concerns were raised about the *real* purpose for evaluations. According to Freeman (1974):

> [T]hese requirements in many cases are adhered to only in form and not in substance. In such cases, just about anything may go. The Office of Management and Budget seemingly cannot be too rigid about enforcing the requirement, for

> key and important operational programs simply would not continue if a strict commitment to congressional intentions was mandated.

Behind the mechanical, scientific process to determine program efficacy was one that might be regarded as willy nilly. Accordingly, decisions about the extent to which a program is evaluated, by whom and by what means became increasingly closely linked to program officers' ideas of what constitutes evaluation and the form of the program. As Freeman contends, decisions about evaluations depended upon:

> a secretary of a department or his office, or the administrator of an agency or someone down the line, or a professor who has a special methodology he wants to try, or a practitioner who believes or does not believe in what is going on, or even an individual politician or newspaperman with a pet project or idea (Freeman 1974).

Another concern during the early and mid-1970s was with the monitoring capabilities of federal agencies. According to Abert, few lower level program managers had the experience or training to guide projects, make assessments of progress, and provide feedback to those conducting evaluations (1979). Large-scale evaluations, such as Sesame Street (done by Russell Sage and ETS), Experimental Schools (Cambridge, Abt), and Head Start (Campbell & Associates) received external monitoring support. But, smaller initiatives were more difficult to oversee.

Furthermore, methodological approaches to evaluations shifted considerably during the early years after ESEA was passed. For instance, the first studies considered

problems such as the effects of compensatory education programs writ large. But, by 1971, the Office of Education was required to report to Congress annually on specific programs being funded. Thus, larger issues and broad strategies were folded into highly-focused measures of single programs' effectiveness. This switch in focus paralleled a decrease in funding for research for research and development (e.g., military contracts) and the emergence of a new contract industry. With increases in federal spending on mandated evaluations and a short supply of staff, many corporations turned to evaluation. Others, however, were developed just before the passage of ESEA, demonstrating a likely anticipation of the law.

By the early 1970s, the for-profit consulting world dominated contracted program evaluations with 45% of the market share. This was followed by non-profit organizations (29%), universities (21%), state and federal government agencies (4%), and independent consultants (1%). As mentioned, one contribution to the decision to seek external contractors was based on the fact that HEW was severely understaffed (Abert 1979). According to Abert, there was no way to keep the mood of evaluation going while trying to build a professional infrastructure. Abert reflects:

> Outside contracting could be done more expediently and with more assurance of obtaining some product within a reasonably short time frame. This is in contrast to the dubious prospects of obtaining authorizations for in-house staff expansion and then, what is perhaps more difficult, filling the billets with the appropriate types of people (1979, 25).

Appropriations for ESEA in the early 1970s made possible emergence of externally-conducted program evaluation as a distinct professional practice. For the first time, and coinciding with the overproduction of PhDs during the 1960s (Shumar 1997), the decrease of Cold War military contracts and the increase in federal involvement in applied social services, many social scientists began working full-time as contracted researchers. This provided the conditions for researchers to become flexibly specialized knowledge experts on a variety of issues, (e.g., curriculum reform, instructional strategies), fields (e.g., education, health, criminal justice), and, therefore, contracts.

The tidy sum appropriated from Congress was for contracts and grants rather than staff buildup. The office of the secretary's evaluation staff was not even able to fill its 18 slots given the dearth of qualified researchers willing to accept government work over corporate or academic work. A concern for HEW in the early 1970s was the outmigration of agency personnel to for-profit and non-profit consulting firms. Companies enticed professional evaluators with higher pay. With these former agency workers, firms could produce proposals likely to score high because of the insider information (Abert 1979).

After a heated debate between bureau chiefs, the Commissioner of Education (Marland) approved a plan for centralized control of program evaluations in 1971 moving away from individual agency control. In 1973, given that some major program evaluations were specifically required by the program's authorizing legislation (e.g., Experimental Schools, Follow Through), evaluation became a single-budget-line item. An implication of this was a centralization of evaluation staff members. Abert recalls

that individual agency-sponsored evaluations led to a "potpourri of small studies" (1979). In many cases, agency staff conducted the program evaluations themselves.

In terms of funding, The House Committee report in 1974 contained the following: "Evaluation reports on the various programs have yielded disappointing results." Funding decisions, according to Freeman (1974), are often made by internal committees who may lack technical expertise to judge the theoretical and methodological promises of a proposal. Their interest may be over which contract organization or individual will be able to deliver on time and "whether the award is given to a group with enough blacks, women, in an economically depressed state and so on" (Freeman 1974). In the new administration in 1974, "a reelected President has initiated a number of actions reducing federal expenditures for social programs. The rationale cited for these actions is that the programs have not been achieving their desired objectives" (Abert & Kamrass 1974). This meant a reduction in funding for evaluation, too. As a result, evaluators and funding agencies lowered their expectations for programs' potential impact. Many expected these poor findings to lead to a reduction in federal funding for the War on Poverty. Instead, even through the Vietnam War and the Nixon and Ford Republican administrations, substantial federal aid for anti-poverty programs continued.

The new administration sought two significant changes. First, social programs would be funded only if they could show purposeful rationale and realistic objectives that could be demonstrated through program evaluation. Second, as a result of the newly centralized Office of Program Planning and Evaluation, creating a more flexible accounting system, replication of models tested by the federal government, state and local

agencies would become responsible for replication and would also be provided federal funds to carry out the evaluation demands of ESEA. This would free up the Office of Education to carry out large-scale national evaluations of program effectiveness (Buchanan & Wholey 1974), since the Experimental School Project evaluation seemed to be progressing favorably. These large scale studies would emphasize experimentation. The financial effect of a successful evaluation was continued, or even expanded, funding. For instance, Buchanan & Wholey note:

> Even though the voucher concept is quite controversial, OEO is funding a demonstration in one site which will total approximately $2 million over a two-year period. If this first site proves successful, it is expected that the program will become much larger (1974).

As mentioned before, it remained clear that the evaluation requirement was not generating desired results, which led to a resurfacing of Taylor-like thinking and approaches to doing evaluations. Thus, the reauthorization of ESEA in 1974 strengthened the requirement for collecting information and reporting data by local grantees. Greater emphasis was given to "systems management" approaches that relied on econometrics, underscoring the value of education systems designed according to standardized and measurable data, which would permit transparent examinations of inputs and outcomes. The period was marked by experimentation into new sectors, but with a close adherence to "traditional" ethnographic methods. These studies were formative in nature.

In the mid-1970s, several projects and symposia brought evaluators and decision makers together. The Washington Operations Research Council held a symposium in 1972 and the MITRE Corporation held another in 1975 that each consisted of evaluators representing sociology, anthropology, psychology, economics, and operations research and content areas (transportation, police, education, welfare, health, workforce, and so on). This conference underscored the similarities in measuring social problems. A common theme in these comings together was a general dissatisfaction of evaluations from intended or prospective audiences of evaluations. The big picture, discussed widely, was that short-term programs had little or no effect. The concerns included a lack of understanding bureaucratic relationships, management idiosyncrasies, long and jargon-rich reports that came after decisions had already been made, and research answers that strayed from the original policy question. These meetings may have sparked a few major shifts in the industry. First, timeliness became a core value, as did providing truncated, simplified versions of reports. Policy implications memoranda (policy briefs) were begun in 1972 to increase the policy impact of evaluation results. Furthermore, these meetings broke down canonical boundaries and normalized interdisciplinary teams.

Before the mid-1970s, evaluators were clustered with like kin. That is, evaluators tended to work with others representing the same social science disciplines and with the same kinds of methodological and content expertise. Early studies tended to be large-scale involving, such as did Head Start, hundreds of sites at once. Sociologists, psychologists, and economists brought with them complex statistical analyses while anthropologists brought case study methods. The 1970s also marks ethnography's formal

introduction to program evaluation. Martin Stake is credited with introducing full-scale qualitative studies (House 1993). While initially, these studies reflected academic research, by the mid-1970s, evaluations used mixed method approaches and almost all used interviews and observations. Education reformers called for evaluations that would move beyond the basic input-outcome models; ones that would shed light on the process as well as the outcomes.

During the late 1960s and early 1970s, social scientists faced a changing job market. Policy research began to emerge as a new ethnographic possibility at a time when, as Coleman (1972) found, "there is no body of methods, no comprehensive methodology, for the study of the impact of public policy as an aid to future policy." At this time, too, applied social scientists turned to engineering for models to guide multi-disciplinary evaluation teams. Around the same time, the Health Services and Mental Health Administration set up the first long-term training program at the University of North Aliceina for graduate and post-doc training in program evaluation methods. Soon afterwards, courses sprang up in academic departments across the country.

As Shumar (1997) points out, this coincided with a fiscal crisis involving "inflation, declining jobs and rising taxes; economic concerns that have dominated the national stage ever since" (p. 79). Shumar links this crisis to declining enrollment in higher education, which, in turn, totally shifted the academic playing field through declining tenure-tracked positions and increased use of adjunct faculty. The over-production of social science PhDs in the late 1960s met with fewer secure academic positions and the opening of government-funded social scientific research and

development. Borrowing from Harvey's (1990, 2005) analyses of the breakdown of the economic system of capital accumulation, the Fordist regime came to rely not only on the amassing of profit, but the core belief that accumulation could continue indefinitely. As I have noted throughout this chapter, early prototypes of program evaluation borrowed Fordism's tune of scale, standardization, and process. During the days of Fordism, liberal scholarship involved a "dynamic and vital tension as well as pressure from the state to keep education a public institution instead of in the complete control of business" (Shumar 1997; 82).

But, after the Second World War, markets became increasingly saturated and petro-control increased the costs of production. One adaptive strategy that survived led corporations to opt for a less fixed modes of production. This led to what Harvey refers to as flexible accumulation. The American economy was transformed as capital became increasingly globalized and as factories closed. Service industries began to rise alongside military contracts. Meanwhile, as Shumar cogently points out (1997), America's activist coalitions were broken up and went their separate ways. What Shumar does not mention is that part of this may have been related to the federal government's enacting policy that at least gave the appearance of giving voice to those liberal interests. Many radicals had gone "inside" by joining the ranks of the federal government or contracting. Campus unrest, successes in the Civil Rights movement, and war protest opened new spaces for applied social scientific inquiry, legitimized by government spending. This space was confirmed, briefly, from 1970-72 as appropriations were made for the evaluation mandate for all Title I funded programs. Massive programs such as Head Start, Follow Through,

and Experimental Schools offered millions of dollars to local schools and for evaluators, creating the groundswell that would institutionalize the field.

As local markets became saturated, flexible accumulation brought more aspects of US society into the realm of the economic. New financial markets included those that had been previously seen as purely civic. According to Shumar (1997), "schools, healthcare, government [had to] find ways to be profitable themselves, that is, to act like businesses" (p 82). As a result, knowledge production was redefined as a marketable product. Knowledge had been for sale via contract work since well before the School Survey Movement. The differences here involved the noted changes in political economy and related hegemonic regimes of ERD knowledge.

Guba and Lincoln predicted a new generation of program evaluation that would provide a basis for evaluating contract program evaluation, itself. Missing in their elicitation of concretized ideal types, however, is both the cumulative formation of evaluation and the fact that many modern contract agreements and the contracting stakeholders (who also hold the purse strings) support non-critical, non-reflective regimes of knowledge production. Also overlooked in this teleological account is the fact that broad measures and achievement tests continue to dominate the measure of success. What has changed is who has authority to produce, read, and put into use the results of those measures. Guba and Lincoln's analysis presumes ideological-theoretical change without structural and discursive shifts in the industry. The industry has, increasingly, given rise to outsourcing of services (e.g., transcriptions of interviews, descriptive statistical analyses, and literature reviews) to the extent that a multinational knowledge

assembly line further distances evaluation activity and products from its consumers and, thereby, complicates it critique, whether internal or external.

Setting the Standards for Scientifically Biased Research

The 1980s mark the establishment of the American Evaluation Association. The organization brought together evaluators representing multiple fields (e.g., education, military, health) and, in that way, formalized networks and pathways for moving from one sphere to another. During the 1980s, the effects of Reaganomics severely reduced spending on program evaluation activity. Between 1980 and 1984, nondefense evaluation units decreased from 206 to 141 and led to layoff-induced exodus from federal evaluation units. Rist reports that while federal funding overall was reduced by 4% during this period, funding for program evaluations was reduced by 37% (1990).

At the same time, federal agencies had installed their own evaluation units—FBI, GAO, HHS, FDA—as had larger school districts and all state education agencies. Evaluation findings would be used to support the neoconservative agenda to make light of the ways in which the federal government had purportedly squandered money on social programs (Chelimsky 1987). The findings that few programs were working according to plan seemed to reinforce the perception that Black America, for instance, was not taking advantage of opportunities. Interestingly, despite cutbacks, even the Reagan administration continued to support social programs, possibly to assuage and, therefore prevent large-scale resistance or perhaps out of political interest in a growing electoral bloc consisting of the poor.

In the midst of these cutbacks, attempts to standardize processes were put into action. In 1980, for instance, the U.S. Department of Education circulated general administration regulations known as EDGAR, which established criteria for judging program evaluation components of grant applications. These changes in legislation and regulation reflected a continuing federal interest in evaluation data. Less clear, however, was how, exactly federal, state, or local level agencies were to use program evaluation results. In 1983, *A Nation at Risk*, (National Commission on Excellence in Education) brought educational reform back to the center of debate. The focus, again, was on raising performance standards for both low-performing teachers and low-performing students. Although there was little emphasis on how this would be achieved, the federal government funded content area teams to develop national educational standards. At the same time, "effective schools research" led to a reconceptualization of school personnel roles from those introduced by scientific management. Specifically, principals were encouraged to become instructional leaders rather than managerial supervisors. And, based on research in anthropology, sociology, and social psychology, teachers were encouraged to hold higher expectations for students. The system would monitor itself through locally-generated assessments and data on access to and allocation of resources to determine specific needs.

After the Reagan years, the professionalization of the evaluation field was furthered by the outcropping of new academic programs in sociology, anthropology, educational psychology, educational administration, and others with either explicit focus on evaluation or applied aims that fed graduates into the world of contract research and

evaluation. Simultaneously, some of the foundations that supported basic and academic social scientific work during the 1960s and 1970s decreased funding while federal and state government funding sources were substantially increased in "applied" sectors. With the increase came a new regime and method for improving education.

In 1988, amendments to ESEA reauthorized the Chapter 1 (formerly Title I) program, and strengthened the emphasis on program evaluation and local program improvement. The legislation required that state agencies identify programs that did not show aggregate achievement gains or which did not make substantial progress toward the goals set by the local school district. Those programs that were identified as needing improvement were required to write program improvement plans. If, after one year, improvement was not sufficient, then the state agency was required to work with the local program to develop an improvement process to raise student achievement.

Also during the mid- to late 1980s, several authors shared "how to" methods for conducting contract evaluation. For instance, Adelman (1984) provided a set of practical "guidelines for the conduct of an independent evaluation." Methods were codified into easily digestible and replicable steps. Several publications criticized methodological orientations of earlier evaluations finding most of them poorly designed and, therefore, resulting in ambiguous conclusions of limited practical decision-making value for policy or reform. The evaluation industry also began to take divergent paths with specializations in summative (outcome) and formative (process) forms of inquiry. The former was dominated by statistics experts and the latter by qualitative researchers, thereby perpetuating the qual/quant rift in this new figured world setting.

Along with the further institutionalization were shifts in the loci of evaluation's workforces. A survey of program evaluators conducted in 1990 showed that 40% were in universities, 12% in state or local governments, 5% in school systems, 11% in nonprofit organizations, and 10% in private business (Morell 1990).[6] This coincided with a call to bring together school improvement, accountability, and data. The National Education Goals were promulgated and formalized through the Educate America Act of 1994, calling for "world class" standards, assessment and accountability to challenge and, as the logic went, inspire the nation's educators, parents, and students to improve.

The 1990s were dominated by a trend toward globalization and downsizing. A few threads need to be established in their own right before I can braid them together. Although the criticisms of federal agencies had remained pretty much the same in terms of initiatives and evaluations, the 1993 Government Performance and Results Act brought the specter of evaluation to focus on the agency itself. It required federal agencies to conduct project management-related activities. With the mandate requiring strategic and performance planning, the program evaluation industry was expanded even more as organizations and agencies adopted evaluation as a management tool. The obvious implication was the opening of a new area of specialization that merged strategic planning and formative evaluation. Management consultants sprang up overnight around the Beltway. At the same time, new funding streams emerged that would reduce the effect of partisan politics associated with federal funding. Bell, for instance, contended

[6] This is most likely not a representative sample of evaluators at the time, since the AEA was predominantly oriented toward academia. Evaluators in the private sector had little to gain by sharing their inroads with agencies and the approaches to doing the work that worked well.

that moving away from federal funding and, instead, seeking foundation support, would provide evaluators a modicum of independence (1997). Large-scale foundations succumbed (or subscribed) to the logic that positive evaluation findings would be good press.

With the further expansions and a tide of new self-appointed experts, the AEA sparked a series of ethical, credentialing and methodological debates about who could be considered a program evaluator. Various scholars contributed articles and presentations attempting to define the work and the worker. Scriven, for instance, noted that professors of evaluation are not evaluators because they do not do "professionally demanding evaluation as their primary job responsibility" (1996). He goes on, in that article that defines the "types of evaluation and types of evaluators" to note:

> The bottom line here is that someone who can't competently do technically challenging evaluation tasks is lexically excluded from the professional status as an evaluator, even if they can and regularly do perform the ancillary tasks of a professional.

While others were less inclined to draw such rigid criteria, an overarching theme was a sense of impotence amidst other realms of knowledge production in that betwixt space. Scriven beseeched "real" evaluators to "take a hard look at what they are doing," "especially in early days while we're trying to get the discipline conceptualized, validated, and credible."

In the 2001 reauthorization of ESEA, the US Education Department (ED) began a new top-down approach. The NCLB Act re-formed major elements of ED by disbanding the Office of Educational Research and Improvement and establishing the Institute for Educational Sciences (IES). Again, standards were to drive instruction, assessment, and curriculum, but this time they would be accompanied by "scientific" methods of evaluation. Under the Act, accountability is based on whether schools, districts and states make adequate yearly progress (AYP) towards the goal of bringing all students to academic proficiency. Schools that do not meet AYP for two years face sanctions. If a state fails to meet any of the requirements, the secretary of education may withhold funding. IES has as its mission getting states to set standards for what students in each grade should know in reading, math and science. This meant evaluation would, in theory, move away from localized, custom formative and mixed method case studies to rigidly-structured, multi-level (e.g., HLM) designs that attempted to link discrete inputs to particular, measurable outcomes.

Some Threads in the Rise of Program Evaluation

Several threads emerge from this spotty historical overview. First, by whom and for whom evaluations were conducted shifted. Early on, educators used data to make program decisions while political advocates used larger data sets to encourage policy decisions. These separate discourses clashed and were eventually merged with ESEA rhetoric. Eventually, data collection and use became the activity of professional externally contracted technico-scientific evaluators. Citizens and educators became subjects of evaluations instead of intended audiences.

Unlike the School Survey experts, program evaluators work according to the contract, which represents a shift in political economic structure; a move towards greater workforce flexibility, constantly expanding markets, direct competition, globalization, and individual responsibility for success. In this context, neoliberalism also fosters the paradoxical generalized expert. Instead of the expert on some aspect of educational reform, companies and individuals market themselves as experts of "capacity building," "strategic planning," and "needs assessment," vague indicators that permit perpetual customizability. At the same time, market expansion leads to layers of "expert" types who provide fragmented services—survey construction, classroom observation, scope and sequence—thereby promoting both the contract and the competitive nature within organizations. This change in political economy, in addition to epistemic shifts in the shape and structure of knowledge and knowing, complicate a comparison because the increase in specialization types compounds the layers of experts.

Program evaluation results are seldom intended for the public or even for the schools or their personnel. They are intended for curriculum/program developers on the one hand and accountability experts who represent the federal or state agency on the other. Ideally, according to the design, curriculum developers, another kind of ERD expert, then use the findings to refine or scrap the program while funding organizations use findings for funding decisions or in some cases as a political lever to maintain or increase funding. Seldom, other than supplying data, is the local school system involved in this process. True, this was also the case with "expert-only" school surveys, but even the most elitist surveys involved district and school-level leaders.

Second, with the focus on individual responsibility for success, failing programs seldom reflect failure with a company's curriculum, but, instead with the target population. Thus, evaluation evidence is often used to indirectly or directly (re)produce the image of the failing minority as the "state shifts the blame for the evident inequalities in access and outcome it has promised to reduce, from itself on to schools, parents and children" (Apple, 2001).

Third, funding streams shaped American social science landscape, as well as that of program evaluations. Funding streams moved from self-administered evaluations to philanthropic-funded studies and, since 1965, from the federal government. Since funding became increasingly tied to demonstrating rigor and replicability, it gave rise to the expert external evaluator. Furthermore, the increasing link between funding for individual educational reform programs and evaluation highlights the federal government's increasing right to intervene.

Finally, this historical précis underscores that while methodological approaches, scale size, funding size, duration, etc. were adjusted, program evaluations continually failed to meet the expectations held by any stakeholders or according to any approach. Perpetuating the quest for the "one best system" was rhetoric of being on the cusp of discovering the ERD panacea.

Chapter 3: The "Figured World" of Contract Program Evaluation

As I noted in chapter 2, throughout the brief history where social science and evaluation have intersected, concerns have lingered about method, theory, and being true to the canons of social sciences while being relevant to decision makers, program developers, the recipients of educational reforms and those who fund evaluations. Beginning with industrial capitalism's efficiency movement, which linked itself to scientific ways of knowing, along with social accounting as a means of improving social services, the program evaluation industry endured several epochal shifts. These shifts were marked by changes in policy, spending, use of findings, dissemination, theory, and political economic context. Moreover, these shifts were indexed by dramatic changes in the underlying assumptions about seemingly-commonsensical educational reform objectives and how they ought to be accomplished. In other words, these are shifts in discourse. This study, therefore, revolves around a figured world that exists, primarily, because of a single mandate and the ERD that gave rise to that mandate.

The early 1960s marks the institutionalization of contract evaluation. For the first time, contract evaluators were independent, full-time knowledge workers. Journals and a professional organization materialized, providing the infrastructure and network needed to perpetuate the industry. It was, doubtless, difficult for anyone at the time to see how these trends would fit into the global political economic shifts that were manifesting themselves in the emerging figured world of program evaluation.

Jameson (1984) describes the postmodern condition, which began to take shape in the mid-1960s, as the social field in which individual identity is produced in a broad web

of disparate institutions and groups, imbued with incommensurable contradictions. With little delineation between virtual and "real" worlds, constant bombardment of media images and ideologues, and time-space compression, the fragments of self accumulate and are arranged according to a never-ending dialogic schema. This fragmented condition imposes a disunified sense of purpose and history; a multiplicitous subject-self. This sense of disunity is echoed in the political-economy through the contract. As with the adjunct faculty, the contemporary contract represents a shift towards greater workforce flexibility, ever expanding markets, and globalization.

In this context, market expansion is made possible through ephemerality. The emphasis on property in classic and market liberalism, for instance, was replaced by an emphasis on the contract in which services are split, benefits not provided, and permanence erased. This neoliberal world fosters the paradoxical generalized expert or as Hirst and Zeitlin call it, "flexible specialization" (1995). Vague marketing terms such as "capacity building" and "solution implementation" are inextricably linked to perpetual shape-shifting and customizability. This adaptive strategy most notably manifests itself in rapid conversion through an instantiation of methods and deployment into new markets. For instance, criminal justice contractors redefined themselves as experts in education within months of changes in federal funding in both sectors. At the same time, market expansion and the contract promote a layering of "expert" types who provide specialized services, creating an assembly-line of research knowledge production. The contract, at once, provides a measure of performance and a means for comparability between competing researchers.

With this new form of market capitalism's insertion into the ERD, shifts in the discursive and material backdrop provide tools and space where evaluators' identities are continually produced through everyday cultural performances—data collection, analysis, report generation—with constant reference to and use of inter-subjectivities, positionality and material conditions. A central concern here is the nature of the relationship between agency and institution in terms of agency. When agency is "explained through internalized social norms or externalized constraint, the meaning of action becomes historicist" (Somers & Gibson, 1994). Along these lines, Holland, et al, contend that Foucault was overly-deterministic. People attend to more than one discourse at a time. Furthermore, direct inscription of power and complicity undermines true agency while also generating a continual, yet fictitious, boundary between individuals and institutions.

Somers and Gibson argue that the terms *actor* and *society* are themselves part of an historically situated narrative in which the prevailing logic presented to individual actors vis-à-vis institutional domination instead of examining the interconnection between them. In fact, Somers and Gibson hope to overcome this illusory chasm through a narrative approach to action [that] assumes that social action can only be intelligible if we recognize that people are guided to act by the relationships in which they are embedded and by the stories with which they identify (1994). They replace the institution-individual binary notion of agency with one that re-defines agency not as individuation, but as constituted within institutions, power structures, and networks (1994). For instance, it now applies to both expressive and instrumental acts. Evaluators are not passive victims of a carefully controlled plot. Instead, a system logic helps

produce their sense of who they are while they contest, refine, and intentionally habituate the tenets of that logic. From this lens, the figured world of program evaluation encompasses a relationship of mutual conditioning, which is fashioned through the interplay of story about the individual's inextricable connection to the figured world. This problematizes the binary between institutions and individuals, since, in this framework, they co-develop one another in a constant dialogic. Neither is ever complete.

In this section, I delve a little deeper into the institutional types, their aims and structures, and the kinds of evaluation work they perform that constitute the figured world of program evaluation. In order to understand the transition from academically-produced social scientist to program evaluator, I turn to a body of literature that describes the ways in which people reinterpret their past experiences and develop new self-understandings. In chapter 1, I described the figured world as the collective "everyday" space where program evaluators become active and passionate about what they do (Holland et al 1998). People become active and passionate about what they do in figured worlds and, at the same time, their "identities and agency are formed dialectically and dialogically in these 'as if' worlds."

A figured world is a "socially and culturally constructed realm of interpretation in which particular characters and actors are recognized, significance is assigned to certain acts, and particular outcomes are valued over others." It provides a system of appropriate actions, rules, story lines, character and organizational types, boundary conditions, ideals, and aspirations. It is, in essence, an imaginary collective space that mediates identity,

behavior, and beliefs through everyday practice with special attention to and enforcement of a set of normative guidelines and value orientations.

Before moving into a description of the scope of program evaluation's figured world, I will delve into a few of its more salient features. First, figured worlds and their effect on individuals appear as natural and normal. The connection between individual and figured world becomes so naturalized that the figurations are experienced sensorially through participation in its everyday routines and the upholding of its rules (Holland et al 1998). As this occurs, the actor codifies the figured world through story—self story and stories about others who inhabit the figured world—which creates a space for reproducing it as it becomes adopted, adapted, and embodied by other participants. A powerful element of this story is derived from a shared sense of we-ness (based upon historically-significant intersections—whether based on first-hand experience or narrative retellings). When a person joins a figured world, she also becomes a part of its collective past. In addition, she comes into a social position within an existing relational hierarchy. While this is, in some ways, similar to Bourdieu's field—a structure-in-practice, a world of relationships and social positions defined against one another—Bourdieu favors hallmark events and prominent figures while Holland et al look to everyday reproductions of social positioning.

The Contract Evaluation Industry

In this section, I provide an overview of the figured world of contract evaluation by extending beyond its historical matrix and by examining some of the current types of organizations that perform program evaluations under contract. I begin this discussion

with a brief list of what I am not studying—policy institutes, lobbying organizations, fact finders, efficiency experts, and privately and foundation-funded think tanks (e.g., Brookings Institute). Because these organizations fit into a different figured world, they require a different set of lenses given their different agendas, funding streams, and expert pool. In addition, the General Accounting Office's Division of Program Evaluation and Methodology, which employs nearly 100 social scientists and over 4000 data collectors/auditors, works directly for Congress and conducts studies in defense, health, education, national security, and other areas. This group, too, is beyond the scope of this study.

Furthermore, this is not a study of internal state and district evaluation departments, although many contract program evaluators have worked in these spheres. These fit into a regime that might be better labeled *school* evaluation, which is more properly a function of accountability rather that *program* evaluation. School evaluation relates to and studies district and state internal operations—the gathering of accountability, demographic, personnel, curricular and financial data for day-to-day operations, placement decisions, hiring, and so on. Program evaluation, on the other hand, is more concerned with potentially innovative curricula. Internally-harvested indicators may become the basis for comparative analysis, but do not constitute a program evaluation. Program evaluations are concerned with process (usually implementation) and outcome (changes attributable to the innovation). The recommendations, therefore, are not intended to drive a school's hiring decisions, or to

target teachers or identify "low-performing" principals, but to determine whether or not a funded *program* is worth continuing.

At the core of accountability is the contract and revolving around the contract are the layers of assessment of the degree of compliance with the contract. Thus, a presumption of the logic is that the contract ensures quality. This is a fundamental assumption across layers, which, at least for a federal flow-through project, might include the following:

- Federal to state education agency (SEA)—federal monitors may use the evaluation results to ensure the reform is implemented according to plan.
- SEA to school district (LEA)—the state education agency is required to use evaluations to ensure that local districts deploy all elements of the reform.
- SEA or LEA to program developers and professional development providers—state or local agencies may use the evaluation results to hold external vendors accountable.
- SEA to program evaluators—the state may hold the external evaluators accountable through the delivery or non-delivery of the contracted program evaluation.

Thus, the fundamental question, and quite a different orientation from basic research, is not simply "does it work," but "does it work according to this system?"

That being said, it is important to point out the relationship between government and evaluations. According to Ernie House, one of the pioneers of contemporary evaluation, as well as a critic of the industry (1993), evaluation legitimates government

bureaucratic activity and educational reform program development by providing solutions (or the appearance of solutions) to social problems. Program evaluations permit agencies to say they put programs through rigorous testing and can, therefore, make informed, "unbiased" decisions about them. Since government authority can be exercised by regulating behavior or through resource allocation (Weiler 1990), evaluation is a means for both strengthening centralized regulation and allocation while also helping to ensure that "implementation" goes according to plan. Yet, as noted throughout the history section, "government" is only a single node of power, if that, behind which operate many other institutions and interest groups.

Contract program evaluation, I argue, is a hybrid figured world that occupies spaces in government, as noted above, corporate R&D, and academia. Included in the contract research and development (R&D) world are multi-billion dollar, one-stop-shopping firms that employ 15,000 or more employees, such as SAIC, which defines itself as a:

> leading systems, solutions and technical services company. We solve our customers' mission-critical problems with innovative applications of technology and expertise. In medical labs researching cancer cures, in the desert testing next-generation robotics, in the ocean deploying tsunami warning systems, SAIC people and technologies are there. In crime labs investigating new evidence, in Iraq helping protect and support our men and women in uniform, SAIC is there (http://www.saic.com/about).

R&D also boasts big name research outfits such as Battelle Memorial Institute, Research Triangle Institute, and Abt Associates, which employ thousands of social scientists to complete government contracts. Together, the U.S. R&D industry contracted an estimated $282 billion dollars in 2003 (OECD 2003) with only about $20 billion going to military contracts (Boeing financial report 2006). The U.S. employs nearly 1.3 million workers in this industry (OECD 2003) by far the largest R&D workforce in post-industrial societies (followed by Japan and China).

Like the umbrella R&D world, contract program evaluation is tied to funding organizations and consumers, program developers, vendors, technical assistance providers, schools, and evaluators (as partners, subcontractors, and competitors). Some program evaluation companies are housed within these large corporate infrastructures. In the Washington, DC area, for instance, I embarked on my journey into this world when I began working for Macro, a company deserving of the title "beltway bandit," employing some 500 social scientists to develop programs for agencies (such as NASA, FDA, HUD, ED), field test products, work as management consultants, conduct market studies, administer opinion surveys and, of course, conduct program evaluations. These corporations face the same pressures as the rest of corporate America and must often turn to survival strategies such as outsourcing, restructuring, downsizing, and moving toward dynamic work teams for each project, thereby contributing to the itinerant nature of program evaluators.

Unlike academically-oriented research, where faculty may devote a lifetime to a highly specialized topic, scientists and social scientists in R&D outfits are flexible

generalists.[7] Whereas new assistant professors must juggle teaching, bureaucratic, and community functions while beginning a research program and generating publications, a young research associate (either a masters level or a new PhD) at an R&D firm is usually hired to complete existing or recently awarded contract work, providing the luxury of a clear set of tasks and financial security for several years, but without the prospect for notoriety or intellectual freedom. The weeding out process in the R&D industry, for instance, focuses on record of serious applied projects, ability to meet timelines, to communicate politically, to work as a team player, and to analyze data. I maintain two professional summaries—an academically-oriented CV that boasts publications, presentations, awards, and community work and a corporate-targeted résumé highlighting dollars earned, project management experience and skills, technical reports, and methodological approaches with which I am most familiar. What is highly valued in one sphere may go largely unrecognized in the other.

At the same time, program evaluation is vitally-connected to the academy. Evaluation's prototypical predecessors of educational science and social sciences, of course, were developed in relative academic isolation from each other in the academy. The academy was the commonsensical designee to perform high quality, applied research on the implementation and effect of curricula being used in schools. A rift between the university's basic research and this sustained real-world call for diagnosis-based prescription, from one lens has remained relatively static for over a 100 years. From another lens, however, that perceives the university as a multifunctionary locale whose

[7] In this case, I have borrowed from Barrow's (1995) notion of "flexible specialization."

employees participate in both scientific and humanistic inquiry, as well as contract research, that rift has expanded.

The rift between academia and contract evaluation is mediated by this notion of "necessity made into a virtue" (to borrow Bourdieu's 1990 terms) and stems directly from the different modes of generation or, in this case, production (of knowledge as a service). As that expansion of functionality for academia has occurred, there is some indication that a new brand of extra-university basic research has also taken root; the post-university intellectual. The basic-applied rift is an antithetical relationship; emerging from a dialectic, but then splitting off, to a degree (they are maintained in close contact by the fact that evaluators require graduate degrees, that many contribute to academic journals, and that many teach while still others conduct offer to the university both basic and applied research funding). A new synthesis has emerged between semi-converging social scientific fields of psychology, economics, sociology, and anthropology giving rise to a hybridized institutional practice. This is evaluation.

Institutional Types

Several organizations—corporate, non-profit, quasi-government organizations—provide the kinds of services once referred to as "school surveys." Take, for instance, in Austin, the National Center for Educational Accountability, the Southwest Educational Development Laboratory, the Texas Association for School Boards, and Region XIII Regional Service Center to name a few. Each of these, either for a fee from the district or through grant funds, examine school finance, curricula, assessment practices, instructional approaches, facilities, school climate, leadership, and so on holistically and

then through guidance over a number of years help to implement the recommendations. In turn, the low-performing experimental schools supply enough of a research and "proof" base that these organizations can package comprehensive programs and further develop their "toolkits." In addition, these organizations offer do-it-yourself school reform kits. The federal initiative, Comprehensive School Reform, offers nearly the same services, but, in addition to the research and development process, requires that each reform "package" undergo a scientifically-based program evaluation.

As noted in chapter 2, ESEA ensures federal funding for these latter evaluations. Several types of evaluation companies compete for these contracts. Here, I provide a sketch of the differences and, perhaps more importantly, similarities, between these kinds of organizations.

University-based Institutes—Contributing to the discussion about program evaluations in education at conferences such as AERA, SfAA, and AAA, university-based research groups focus, instead of on basic research problems, applied problem areas such as program evaluations. Oftentimes, university institutes and centers confine themselves to a handful of contracts with enough strategic overlap to keep money flowing continuously. A key difference between these organizations and contract research organizations is that university institutes may produce data for publication and, therefore, projects are selected according to the director's area of expertise and/or interest. In addition, academics are likely to pursue research agendas with an appeal to their discipline rather than adhere to the program objectives. Beyond staying with the scope of work, a major concern with using university-based researchers is meeting

deadlines. According to Abert, the "academic community is often slow in fulfilling the terms of a contract, in preparing reports, etc" (1979).

Furthermore, university institutes make use of a cheap available labor pool, known as the *student*. My first paid research assistantship was working on "An Ethnographic Assessment of the HUD Urban Redevelopment Project" with the University of Maryland's Cultural System Analysis Group (CuSAG). CuSAG had contracted with Abt Associates as part of a HUD study to assess sentiments of residents of urban public housing complexes that were about to be razed. Methods involved assessment of public housing data, "windshield" tours of public housing units, and focus groups with long-term residents. Organizations like CuSAG are managed by professors, typically operated by a research associate (an aspiring PhD student), and data are collected, "cleaned," placed into databases, and analyzed by graduate and undergraduate students. Aside from doctoral students carving out dissertations from the evaluations, the labor is usually nameless and easily replaceable. Institutes often rely on individual subcontractors to carry out specialized tasks.

The other side of the story is that while some university-based centers rely exclusively on soft money for salaries, many do not. Mary commented: "Even though faculty feel pressure to bring in contracts and to publish, they have more leeway. Their jobs don't depend on bringing in new work. They're supported in other ways. So I think they can afford to be pickier or to specialize." In fact, funding streams are often used to support student work rather than faculty FTEs. Participants who had worked in university settings made comments similar to this one:

> We needed some money to support the students, but not to support the faculty, so in the [project] operation, we decided to take on smaller things which we could manage and maybe do something with and we would learn from, and we would publish articles about, so we were taking on projects not as a contract shop.

That being said, university professional lines are increasingly dependent on soft money. There is a shift toward greater reliance on flexible human capital. In my work for a mid-western university, for instance, most of the staff members were laid off when a contract was not renewed.

Contract Research Organizations (CROs) aka Beltway Bandits—Throughout Washington, DC and close to most technically-oriented industries in other cities, beltway bandits are highly-organized consulting firms that typically sell themselves as specialists in as many areas as there are agencies with contract work to be done. Some companies had previous specialization areas. The CNA Corporation, for instance, was, until recently, the Center for Naval Analysis, a military R&D firm. Given overhead and competitive salaries, beltway bandits typically go after larger contracts, but, as I recall one boss' mantra, "elephants eat peanuts." Small contracts provide entrée into new agencies or contract areas. Doubtless, beltway bandit work in education began with ESEA. Some of the largest including ORC Macro (established in 1966 as Resource Management Corporation) took on larger federal evaluation contracts such as Teacher Corps.

Abert refers to private companies as dronelike and depicts a widespread image of the time that they are "easily co-opted as they do not want to 'kill the goose that lays the

golden egg'" (1979). These companies, however, are the "most businesslike in their approach, formulation, timing of reports, stress on management and administration . . . and [are] more concerned with the with the formal fulfillment of the contract than are universities and non-profit organizations" (Abert 1979). Furthermore, private companies tend to have the greatest variability in quality of work produced.

Beltway bandits recruit talented social scientists from government agencies, other private sectors, and recent graduates. In these settings, academic degrees are less important than are what people are willing to learn (the CEO for Macro, for instance, held a BA in liberal arts). Once in, trainees go through extensive and ongoing, high quality professional development, especially in methodology. Some training is actually a means to field-test potential commercial training sessions (e.g., focus group moderating). Indeed, any product or service can be commoditized. A lot of energy and focus goes into writing proposals; perhaps more so than what goes into completion of awarded contracts. Beltway bandits have well-developed processes for writing contracts overnight and identifying requests for proposals (RFPs) before they are posted. But, at the same time, given name recognition, CROs are positioned well enough to be able to target large projects and to strategically refuse business opportunities.

As some participants who had worked in CROs noted, staff work across areas. Helen, for instance, said:

> I can pull in people working on a project who have backgrounds in health care or criminal justice or child welfare or housing. There's lots of opportunity to learn,

> lots of opportunity to do interdisciplinary types of projects, and it's you know, it also, you have greater capacity to go after more stuff.

Because methodologies are similar across fields of inquiry, beltway bandits warehouse talent banks of employees to write proposals. One implication is that employees do not always know when they are being bid as an expert in a certain area and, when the contract is awarded, must become quick studies. During the three years I was with Macro, while I was a project director for a qualitative study that tested instructional materials for Deaf students and for students with learning disabilities, I also worked on dozens of projects for the Food and Drug Administration, HUD, the Department of Health and Human Services, NASA, the Corporation for Public Broadcasting, the National Captioning Institute, and so on. I recall one week in which I worked for 14 different clients.

Labor conditions can be fierce in these organizations. Long hours and brutal demands to produce give the, often earned, reputation of meat-packing plants that offer rites of passage for new graduates. Labor is often organized loosely in work teams. With a flexible organizational structure, new teams are formed for each project. Two participants mentioned their experiences with such organizations:

> A lot of those companies bring in people with masters degrees to do most of the data collection, analysis, and writing and then have PhDs run departments at 5-10% FTE on each project. They work these masters level practitioners to the

> bone with the expectation that turnover will be high. Then, the ones who make it through, of course, advance.
>
> I interviewed with Abt Associates. They said during the interview that I bring more experience than they wanted. They preferred to bring on new graduates. I asked why and their line was that if we bring in people from other organizations they have trouble getting enculturated. They think that there is a culture mismatch. They think that if they bring them in new and fresh they can form them the way they want them.

Small for-Profit Operations—Any type of company operates similarly to large beltway bandit corporations, but maintain a relatively small staff (1-10). They specialize in some specific aspect of the contract process thereby providing subcontracting services—building online surveys, setting up and carrying out focus groups, or creating interventions while others carry out small-scale program evaluations.

Directorship is usually carried out by a sole company proprietor or a partnership arrangement. Many are former government or beltway bandit program managers or officers who maintained their network of contacts. Depending upon how it is situated and managed, a small private firm may offer greater intellectual freedom and higher pay. A lot of small for-profit organizations are operated out of spare bedrooms and attics, greatly reducing overhead. A lot of times labor is comprised of networks of colleagues, hired guns (described below), and graduate students in need of research experience). Several participants with whom I spoke said that smaller operations like this were often

headed and staffed by people who left smaller companies. Greater control over the sort of work sought, prospect for higher pay, and the prospect for building long-term customer relations are some of the benefits while loss of even a single major contract could spell the end of the organization.

Not-for-Profit Operations—While they tend to be more expensive than private companies, the reputation holds that non-profit organizations tend to conduct higher-quality evaluation work. Some non-profit organizations, however, have their own intellectual and political agendas that limit the usefulness of evaluations for the federal government (Abert 1979).

While day-to-day research operations may, on the surface, appear similar to those of beltway bandits, not-profit work is typically overseen and decided by a board of directors. Furthermore, since they are linked to larger organizations, there is usually less demand to write proposals. New work maintains the steady state. In times of financial hardship, the larger organization can financially maintain the program evaluation department. But, the catch is the evaluation unit usually provides in-house services of some sort, such as institution research, constituency polling, or workshops on recent policy. Furthermore, unlike academic work, which might result in publications, this work is generally repurposed for organizational policy or constituent-focused reporting (e.g., newsletters intended for district-level administrators throughout a particular state).

Regional Educational Laboratories—One particular type of non-profit organization deserves special attention. The passage of ESEA I (1965) brought into existence ten federally-funded regional education laboratories (covering 50 states and US

territories) to provide technical assistance and develop "effective" reform models for struggling schools. Each REL houses scholars and practitioners who work in teams on secure five-year research and development project contracts. Some RELs, such as AEL (in Charleston, WV and SEDL in Austin, TX) employ approximately 100 people with departments of 5-10 specialists dedicated to contract research and evaluation.

REL work is overseen by federal program officers who visit periodically, conduct audits and review research. More locally, the RELs are managed by a CEO, CFO, and an operations person (usually a vice president) and the work they do is determined in large part by a board of directors. The board is typically comprised of a combination of state education commissioners, local university and public school educators, and members of the corporate community.

REL contract evaluation work is focused on model development and internal evaluation of the model's implementation—the latter sometimes leads to animosity between evaluation and technical assistance departments. With the passage of NCLB, RELs were led to believe they would be required to conduct scientifically-based research (i.e., experimental design). Compliance with this understanding of the mandate was complicated by the passage of NCLB midway through the five-year contract. In other words, where mixed method and qualitative studies had been underway for nearly three years, overnight RELs were required to plan and put into action random trials. While proposals had been, historically, used to layout a relatively flexible scope of work and RELs were almost guaranteed to win, in 2005 guidelines became more rigid and several

organizations lost their REL contracts. Some contracts went to university-based institutes and others went to beltway bandits (e.g., CNA).

Anticipating a possible loss of the substantial federal contract that had been renewed each year since 1965, many labs diversified their bases of revenue by seeking contracts through departments of education and school districts to perform mandated program evaluations, as well as with private model developers. More than half of AEL's revenue was generated through contracts with such non-federal contracts. This former REL also established a higher education "co-venture" with 16 universities to supply "services that require additional staffing or specialized expertise." Thus, AEL was transformed, almost overnight, from an REL to a beltway bandit company while CNA moved from its R&D status to that of REL. Federal support has not been consistent for the RELs. As James noted:

> At different times, federal administrations have tried to shut 'em down. The Reagan people were never very happy with 'em and tried to shut 'em down but the labs had enough political influence in congress, they had their own lobbyists, that they managed to keep things going.

Hired guns—While all contract workers may be "hired guns" in the general sense, the term in the industry is reserved for freelancers who, rather than have their name associated with contracts or subcontracts, operate according to service agreements for specific, short-term work. These are not consultants in the sense of expert knowledge producers, but instead are people who carry out discrete tasks—collect subsections of

data, conduct a discrete analysis of data already collected, or write reports. They have little or nothing to do with project planning, creation of instruments, report writing, etc.

Hired guns are drawn from other sorts of industries—academia, retired researchers, for-profit, non-profit, etc. Hired gun work opens doors for individuals and can lead to crossover. And, sometimes, of course, hired guns are former employees. Pay is, typically, on a per diem basis. Usually, for the organization hiring these day laborers, the rationale is to fill recent vacancies temporarily. Hired guns do not have much power in decision-making. As Helen, the director of a contract research unit, noted, "I like working with hired guns. They're not embroiled in office politics." They are brought in to carry out a discrete task, not ask questions, and move on when the job is done.

Gurus—Another type of individual program evaluator is a highly-sought after expert in the field. While gurus hold academic or government positions, many hold the full-time status of well-paid guru. Like hired guns, they are brought in to provide a specific service and often work off sub-contracts.

As Helen noted:

> The people I've hired to do analysis are you know, they tend to be specialists or to be able to do very sophisticated things - things that are outside of my skill set or my team's skill set. And that's always fun because I'm annoying and I like to butt my way in and say, "Hey, show me how you did that" or "Will you explain that to me." You know, it's a learning experience for me too.

Work roles might include helping federal agencies restructure or undergo some other sort of strategic planning, assisting in the selection of external evaluators for high stakes federal or state program evaluations, or providing technical assistance for research designs of large-scale projects. And, as one more infer from Helen's comment, gurus also provide opportunity for evaluators to pick up new skill sets. It, therefore, may offer some internal capacity building and professional development.

Before moving to the next type, here is a note about consultants. Whether guru or hired gun, the niche for consultant program evaluators grew tremendously in the 1990s. As Mark, a guru evaluator who specializes in strategic planning, noted:

> Mark: It [consulting] became a recognized part of the evaluation community and that would have happened back in the mid-90s to late 90s and it's going on now. I think there's probably even a consulting evaluation TIG [topical interest group].
>
> Keith: What do you suppose were some of the impetuses for the shift?
>
> Mark: There was an insufficient number of people who had the skills to: 1) talk the languages of performance measurement and evaluation, and 2) actually play the game of performance measurement and evaluation. So it was something in government that was totally, well not totally absent, but it was infrequent.

But, that is not the whole story. Consultants were brought in because of their specific areas of knowledge, but there was also an acknowledgment that they would only be needed during the transition. Mark continued his recollection:

> You can get an evaluator that comes in and does his work and you don't have to pay him any benefits or anything like that. You don't have to give him an office or anything like that. The problem is that some of the consulting groups will double charge so you're going to be indirect anyhow.

The increase in use of consultants was tied, no doubt, to the Bush Administration's continuation of the Clinton's efforts to build an accountability system. While many social scientists, whether in academic circles or in private facets of the figured world of program evaluation, vehemently opposed NCLB's apparent legislated approach toward evidence (i.e., the gold and silver standard, which was actually a suggestion put forth by the National Research Council in 2001), this reauthorization of ESEA brought evaluation to the fore. While, ultimately, researcher's fears about the winnowing away of non-experimental frames were assuaged, NCLB forced evaluators to communicate with broader audiences, including practitioners. As Sarah noted, "NCLB made the idea of program evaluation a little bit more accessible to people that I'm talking to." She went on to describe a sense that she, overnight, needed to learn to defend findings to audiences:

> I needed to be able to defend my approach and my take on something. Before that I didn't understand the need to do so, I was just like, "That's what I'm doing, I can't really articulate why." I tend to do something intuitively and kind of in a way that I don't necessarily, can't articulate why until I've thought it through, and

> so it helped me to learn the value of that and also be able to express it in a way that still felt like I had integrity but also wasn't going to be just knocked down.

Data-based decision-making, participatory evaluation, logic modeling and formative evaluation tied to outcomes took rise. Mary described the ways in which evaluators began assuming more of a facilitative role:

> A big change has come about in terms of actually helping people thing through what they want to do and how to get there. And, then there is being able to be in on the discussion in a formative way—being a reflective person, a critical friend and being able to go into the project in more rural areas that do not have personnel and who were really inexperienced ass program people or evaluators. Though they were good at what they did, they didn't know how to run a program and be able to go out there and help identify the goals and why particular activities were chosen to achieve those goals. I guess the shift was away from thinking of evaluation as kind of summative thing and thinking of that role as being part of the project team. We provided valuable input from the beginning all the way through.

And, Laura, a hard-core econometrics type researcher, commented:

> The days really of bullshit evaluation, like "I'm just gonna do a survey and tell you what the answer was," are coming to an end. They were supposed to have come to an end 6 years ago, 7 years ago. It used to be, when you evaluated, you didn't have to say anything besides, "Hey we worked with 1000 people." And

> then you had to say, "We worked with 1000 people and helped them get jobs." And now we have to say, "We worked with 1000 people and helped them get jobs and 80 percent of them were employed 2 years later."

Thus, at least according to the participating evaluators in this project, program evaluation underwent a sort of transformation that made it more relevant to stakeholders and that required evaluators to explain to those stakeholders the technical qualities and pragmatic value of the research.

Globalized Research Services—Outsourced contracts do not only go to hired guns and small collectives of contractors. They also go to larger companies, some of which are multinational, especially for mundane services such as transcription, cleaning statistical data, literature reviews, telephone surveys, and recruitment for research studies. Others set up and administer online surveys or chatrooms for conducting online interviews and focus groups. This has fostered an institutionalization of virtual research. Returning, briefly, to the hired gun type, many projects are negotiated, contracted, and delivered entirely over email. This past summer, because an organization for which I used to work had too many contracts, but none secure enough to justify a new hire, I was offered a substantial contract. Over email, I handled business transactions, received data, and sent in a final report for an interview evaluation of a Teaching American History program. I received 21 transcripts and interviewer observational comments, an overview of the project, and copies of protocols. The interviewers then reviewed my rendering and we negotiated the final report electronically.

In addition, international efforts to adopt or adapt facets of the ERD (or the American system of ERD wholesale) have given rise to the internationalization of gurus in federal departments of education throughout the world. James, for instance, described his work in several countries in the South Pacific and Northern Europe. He was brought in to help establish evaluation departments and to help sort out various pieces of the system's units through strategic planning. I have worked in Puerto Rico for the PRDE doing similar kinds of work. While we write about the liberating aspects of this work—helping to give voice to marginalized populations that are being brought into national and global economies—often omitted from our accounts in the crowding out of local competitors. I will give one brief example.

Last year, a local university asked me to sponsor a fellow from Mongolia who is a professor of education. My role was to allow her to experience all phases of a program evaluation as she took formal classes in evaluation methods. In the interstices of her dozens of questions an hour demand for a Socratic experience, we talked about Mongolia's system. She described to me how European companies had recently been awarded government contracts to help establish an accountability system. She also told me that there was a chance the country's research and evaluation system would be outsourced, too, unless they built the capacity to do their own.

I point this out partially as a sideline issue, but also to identify another way in which the ERD's expansiveness has manifested itself. ERD's privatization permits it to enter, almost effortlessly, new markets. In turn, this creates new niches for program evaluators, myself included.

Fragmentation of Types—While these ideal types might help to conceptualize the larger field of evaluation, in reality everything is in a constant state of flux. At the individual level, many evaluators, over their careers, work for each type of organization. At the organizational level, many current beltway bandits were, at one point, non-profit organizations or even quasi-government agencies. Because of this movement, these types should not be used to predict the type, motivations, or quality of work individual organizations do.

The Logic of Program Evaluation

Regardless of the type of organization, elements of the ERD permeate the everyday marketing and mindsets of evaluators. Amidst the ERD rhetoric, which thrives on the notion that program evaluation is a fundamental aspect of the intentional transformation of schools into more equitable institutions, evaluation is depicted as apolitical and unbiased. Evaluation services are marketed as *non-partisan*, which Datta defines as "evaluation that is and is regarded by partisans of all persuasions as balanced, fair, and faithful, so that if methodological quality is high, debates focus on the implications of the findings for practice or policy, not on the credibility of the findings themselves" (2000). A quick perusal of websites of organizations that specialize in external contract program evaluations illustrates this tendency:

> Our Research and Evaluation department conducts research and evaluates education programs to determine what works in the field. We then use their

> findings to help improve education outcomes for schools, teachers, families, and children.
>
> [We] provide original, nonpartisan research and evaluation to policymakers, state agencies, nonprofit education organizations, and school districts.
>
> Our work is always objective, nonpartisan, and evidence-based.

A fundamental contention, which I must make explicit, is the fact that social scientific knowledge production is never neutral (Trouillot 1995; Popkewitz 1990; Bourdieu 1977). Evaluation fits into the ERD in which the manufacturing of practical, objective knowledge mediates existing conditions, desired short- and long-term outcomes, funding, and implementation of educational reforms with the intention of providing ideas for corrective programmatic action or to make judgments about the efficacy of the program. Because, according to convention, program evaluations are intended to affect program development, refinement, or termination—embedded with ideological and economic implications—they are enmeshed in relations of power and privilege. In order to remain unquestioned, these relations are shrouded by a convincing, commonsense, incrementalist chain of reasoning: equity and improvement come about through logical and sequential processes, program effectiveness can be measured objectively through expert practice, and the enactment of these practices is a humanitarian course of action directed at the learning needs of "at risk" students. That program evaluation activities both retain a prominent position in decision making and lurk beneath the visible surface of educational reform activity is an interesting problem,

especially in the midst of a discursive regime that seeks to expose failing children, underqualified teachers, and poorly-constructed and implemented programs in an effort to make educational processes "transparent."

At the same time, an almost automatic assumption emanating from the ivory tower targets the ways in which knowledge production in program evaluation raises doubts about the potential for the practices of this industry to contribute to equity-minded change. The argument goes: if evaluations are created by powerful interest groups to inform instrumental questions, instead of deeply rooted social inequity, does that not make funding organizations the determinant of the research agenda, which, in turn, maintains current power and meaning structures, while silencing others?

As I will demonstrate in chapters 4 and 5, this assumption might deny those who become evaluators any agency and presuppose that the conceptualization of program evaluation research is automatically devoid of findings that are critical of systems and practices. Doubtless, program evaluations are affected by the political environment's omission or "amnesia" of some findings and the magnification of others. Furthermore, outside of government-mandated program evaluation, NGOs perennially sponsor evaluation with the intention of influencing public policy (Patrizi & McMullan 1998). But, organizations that vie to maintain their credibility in the contentious environments of multiple stakeholders and competitors must produce defensible and credible findings that are not unduly biased.

In fact, credibility and objectivity are the spaces in which the supposed power of evaluation to bring about social change is carried out. House comments: "The more

objective and less ideological evaluation becomes, the more useful it is and the more it threatens established authority" (1993). In this vein, one particular passage, written by another master evaluator, stands out for me:

> Evaluation has a much greater potential for fulfilling a "subversive" mission of undermining support for bad decisions because of the associated fear of absolute power. Rather than carrying out studies assigned by a central authority with results controlled by them, we have a range of constituencies and great latitude about how we choose to do the work of evaluation (Henry 2001).

Henry's passage underscores not only the degree of agency evaluators possess, but also the "invisible" pervasiveness of the objectivist ideology within the ERD that manifests itself in the figured world of contract evaluation. Objectivity is a political tool, a choice, an ideal that, when applied with "fidelity," will purportedly contribute to positive change, as well as make strides toward a democratic ideal of participation.[8] In theory. And, in dogma, at least within the figured world of program evaluation. I will elaborate on this theme in chapters 4 and 5.

The Modern Program Evaluator

The evaluators with whom I spoke provided insight into the dispositions they felt were required of a social scientist who aspired to be a program evaluator. The characteristics echoed most often were those related to possessing a desire to make a difference, having proficiency in a wide range of methodological approaches, being

[8] Contemporary evaluation, as a neoliberal manifestation of the participatory nature of democracy gives the image of hearing all the *relevant* voices.

attentive to relationships with participants and clients, accepting the economics of the industry, and embracing a generalist social researcher stance.

Wanting to make a difference—From their perspective, program evaluation offers an opportunity to contribute to social change. One participant, for instance, who has conducted program evaluations for most of his 40-year career and who is now a guru, remarked:

> [Program evaluation] incorporates a piece of what I believe is the helping profession, which is, "Here you are, generating information to really help people make important decisions," which relates back to some of the ideals related to social justice, either at a micro level or a more macro level.

Another program evaluator, who had left academia because evaluation seemed more promising as a liberatory tool, commented that evaluation "incorporates the issues of equity, social justice, as well as questioning whose values inform our judgments." As I will demonstrate in the following chapters, *making a difference* is often contrasted with *producing academic knowledge*.

Knowing how to collect and analyze data—In order to do all this liberating knowledge production, program evaluators need a repertoire of tools. Participants also said that having a broad methodological base was important. Bonnie, for instance, said:

> I have to know how to manipulate complex SPSS files. I have to know what to look for. I have to know how to interview and do qualitative coding. I have to know how to design an evaluation and write a report.

Similarly, Emma suggested that prospective program evaluators "Get as much methodological training as they can. Qualitative and quantitative research. I just think that that's really key." Dr. Nicole, a pioneer of program evaluation in its contemporary form, provided the following list of dispositions required of program evaluators:

> Master statistics and quantitative analysis and economics because this is the language many decision-makers best understand and in which various debates are framed. Master ethnography, culturally appropriate theory and practice, and the history in your areas of interest because these form the context within which numbers live. Master the arts of journalism, of negotiating, and of listening because you'll need them all as a practicing evaluator.

Building and maintaining relationships—Furthermore, according to participants, program evaluators must be good at forging and maintaining relationships. Mark, a guru evaluator, said:

> Hone your relationship skills. That that's the number one contributor to success in our field; if you can get along with people, if you can build effective relationships, you'll be successful.

Similarly, James, another old-timer, said that:

> If you can't stand presenting people with stuff that's going to upset them then you should do something else. Do something more like pure research or something along that line because evaluation has a huge political component to it, which

> makes it far more interesting and exciting, but also far more conflict-ridden than pure research. If you want to get along with everybody, it's not the place to go.

Again, especially with James' comments, the program evaluator's characterization is constructed in contradistinction to the socially-awkward academic "pure researcher."

Being able to bring in money—The ultimate measure of a great evaluator is not in publications, but in dollars. Thus, another important facet of the figured world rests in its economic realm. One implication of this is the lack of fixed salary lines. Participants underscored the importance of having a willingness to work on soft money. Helen said:

> I wouldn't want to hide any of the warts and wrinkles that you are accountable to a different kind of master. Soft money is always an issue. I know that you have to do that some in higher ed, but there really is a press in organizations like this to bring in money. And, if that gives you the heebie jeebies, then this is not for you.

Contract researchers tend to be itinerant. The soft money aspect means that researchers must occasionally relocate and/or must reinvent their areas of specialization. One participant said:

> I would think that would be even a little more complex if you felt that you really wanted to stay in one particular geographic location, especially where there wasn't a whole lot of competition between organizations for that kind of work.

This has implications for selling services and, therefore, selling oneself. In other words, as legislative winds shift, what counts as a suitable methodology or the priorities of the

reform agenda, program evaluators must, quickly, be able to perform in those new areas. This means that being a quick study and having the capacity to master new subjects in relatively short periods of time is essential.

Mastering flexible generality—And, this leads to the fifth facet, being a generalist. According to participants of this study, an appeal of evaluation is that it "introduces you to different areas and different kinds of programs all the time, as opposed to working in one kind of thing." Katherine, a full-time professor who conducts contract work on the side, said, "If you have a low boredom threshold, this is a wonderful field to be in, because it affords you the opportunity to work on lots of different and interesting projects."

Of course, this list is neither exhaustive nor are the categories mutually exclusive. In fact, these dispositional descriptors interrelate. For instance, being good at maintaining relationships implies being able to negotiate. And, being able to negotiate work in new areas has major implications for keeping abreast of new methodological developments. I will return to these facets in more detail in the next two chapters, especially with reference to those pertaining to the educational reform discourse (e.g., social change through research) and those related to flexibility (e.g., being a generalist).

Counter-World or Hybrid Figuration?

Throughout this section, several notes have indicated a contradistinction to basic research knowledge production. Holland et al might refer to academia as a counter-world of the figured world of program evaluation (1998). They contend that:

> The movement from play world to figured world, from a world without a public to a world with communities, imagined or otherwise, is often accomplished by the figuring of 'the opposition' to this publicization . . . [O]pposition and barriers to the emergence of these worlds and the development of tools of insult and derogation for threatening or mobilizing action against those who supposedly endanger the course of emergence (250).

Counter-worlds clarify our threats and position those who inhabit them as inferior. The coming pages provide sufficient evidence to support this stance. And, from this stance, the figured world contains important cultural information about the boundaries that help maintain it vis-à-vis the *other*.

The figured world of program evaluation occupies a marginal knowledge producing territory. Like anthropology's identity crisis, program evaluation positions itself as a relatively underdefined realm of knowledge production. Because of its occupation of a space between academia, bureaucracy, and corporate industry, evaluation draws upon and contributes to existing meaning structures and systems of privilege and, concurrently, provides a method of questioning educational reforms, legislature and spending. Evaluators gain content and social scientific technical proficiency, as well as refine a sense of social engagement, from the academy, while maintaining a space that, from an academic lens, occupies a lower status than those of the ivory tower. Therefore, the realm that legitimates is also the one against which the evaluator identity is framed.

As Henry points out in his lucid critique of the relationship between democracies and evaluation, "Findings from evaluations routinely permeate the boundaries of

academic publications and technical reports and find their way to the public through the popular media" (2001). Yet, these are not the same stuff of basic research. The marginal knowledge production status is both a black mark and a status symbol. Donaldson argues that "the failure to concern ourselves with our reputation outside the 'evaluation in-group' could be detrimental to future evaluation practice and the advancement of our field." After discussing the difficulties of legitimizing evaluation for those directly affected, Donaldson turns to the "second rate" status of evaluators in academic circles for taking on research on "messy 'real world' problems and questions that often require giving up scientific controls" (2001).

The transition to that lower status in the knowledge production matrix is not an uncomplicated selection of career and resultant training. It is both agentic and purposeful in its process of repurposing. It is also not sheer acceptance of a lower caste in the world of knowledge production. Indeed, only a small percentage of people who go into program evaluation have specific training in evaluation. This trend is corroborated by the stories participants of this study tell. For instance, one commented:

> My little sort of funny line is, "No one grows up saying they want to be an evaluator." Well lo, CNN last week put out a list of the ten best unknown professions and Program Evaluator was on that list. So I thought, "Look at that! We made the list. At least we're getting recognition as being a good, but unknown, profession."

Social scientists enter the industry only partly prepared, not yet attuned to project micropolitics, funding cycles, methods of oversight, and the constellation of entities—and their interests—involved in everyday work. At the institutional level, proficiency is marked by a mastery of methodology, winning contracts, building a client base, and accumulation of profits and intellectual property accumulation for the contract organization. Yet, as I will show in the next two chapters, for the individual, the influence on policy and practice represents a kind of soft activism; a way to engage in social justice from inside the institution and, therefore, justifies the use of such "practical," palatable research approaches.

Adoption of the program evaluation figured world involves more than the unproblematic adaptation of social scientific epistemology and a simple exchange of academic-like ways of knowing for techniques and processes that meet the scope and agreed upon demands of the contract that are readily accessible to clients. Becoming a proficient external program evaluator involves the internalization and habitualization of the ERD and the rules of the figured world of evaluation. The adoption of these is, in turn, a fundamental aspect of identity production. One director of an evaluation unit responded to my question about what she would tell a budding social scientist who was considering program evaluation as a career path:

> I have to do that every time I hire somebody. I say, "Go ahead and learn your methods, but program evaluation is an apprenticeship. Learn everything you can about the methods, but you're not going to learn how to be a program evaluator without doing an apprenticeship," basically. Because you don't learn the stuff in

> grad school, you can't come out with a Ph.D. in psychology, for instance, and say, "Ok well I'm going to do program evaluation now and I'm going to be great at it." You've got a lot to learn, and it has to do with the consulting side of the role and it has to do with applying all those methods that you've learned, all those designs, those methods, those analyses, to working in the real world, where your subjects aren't getting ten points in their psych class for being your subjects. And if you don't find off-roading fun, go do something else. Because this is not clean. It's not neat, and you're going to be constantly challenged with something new and different every time you take on any project.

That being said, not everyone is permitted to enter this figured world. The perimeter of the program evaluation figured world is patrolled by upholders of the normative rules that codify credibility, rigor and ethical soundness. Well-positioned evaluators have cautioned each other and clients to be wary of fakes, “especially in these early days while we’re trying to get the discipline conceptualized, validated, and credible” (Scriven 1996).

Thus, social scientists enter the figured world of program evaluation under the auspices of socially-positioned gatekeepers. When they join, their learning about program evaluation is far from over. In the next chapter, I will examine the steps to proficiency required of social scientists when they enter the figured world of program evaluation.

Chapter 4: Becoming a Proficient Program Evaluator in Education

From the ashes of the War on Poverty, program evaluation's emergence coincided with a narrowing of career prospects in social scientific academia and a concurrent widening of job opportunities in corporate and government sectors. While they rely on government and academia for legitimacy and work, social science program evaluators do no enter this figured world as automaton-like victims of economic transformation, misguided career advice, or rejection from academia. They are motivated to enter it with a carefully guarded belief that this repurposing of social science has the potential to enhance planned educational change. As James, an old-timer program evaluator, noted:

> My whole impetus was to change the schools. I didn't like high school at all and I didn't like how they treated me and other guys I grew up. So, I wanted to change the schools doing that. I went from doing that to my primary identity now which is as an evaluator.

As I discuss in Chapter 5, evaluators share a felt need to speak on the behalf of the "underdog," a seemingly natural proclivity for using social scientifically interpreted data to contribute to social change, and an experience-based understanding that planned change does not come about as easily as some organizations might lead us to believe.

Evaluation may seem like an uncomplicated set of methodological and accounting procedures. And, perhaps, in some circles and cases, it is austerely mechanical in its assembly of evidence and unquestioningly compliant with funder wishes. It does, after all, reward generalist orientations toward research and the use of efficient methods of

amassing and making sense of data. But, the social scientists in this study who consider themselves to be proficient, excited members of the industry actively adapt social scientific orientations—both theoretical and methodological—in creative, intentional ways. Their toil is seldom a strict adherence to industry formalizations and policy. For instance, while professional organizations such as the American Evaluation Association propound loyalty to the Evaluation Standards, nearly one-third of the evaluators with whom I spoke had no idea the Standards even existed. And, some were unfamiliar with the history and policies that pertain to evaluation, including the federal ESEA mandate. Yet, the descriptions of their work and the rules by which they do it suggest a broad observance of a few ERD tenets and common practice. While two evaluators may conceptualize and carry out the same study in vastly different ways, evaluators expressed similarities in what they needed to learn in order to be proficient and how they learned it. As I have argued, becoming a proficient program evaluator involves the recursive, multi-phasic process of internalizing the ERD.

In this chapter, I explore the process of learning how to be a program evaluator from initial entry into the figured world to becoming proficient. I offer a rendering of their articulation of how they make sense of (and space for action in order to make) the leap from social scientific discipline to pragmatic and, perhaps even, liberatory knowledge work. The set of story snippets provides insight into the legitimizing principles and ritual practices associated with program evaluation, its group ethic, maintenance of its boundaries, and the implications of group work (occupied by inter-

epistemological and interdisciplinary merges, contestations, and heteroglossic shared spaces).

Getting Into Program Evaluation

Contributing to the shrouding of program evaluation is the lack of voice of those who do the work. It is not intentionally hidden, but it is also not a particularly well-advertised or well-known figured world. As I will demonstrate in the next chapter, participants hold a range of responses (roughly corresponding with class conditions of their youth) from an overt association as program evaluator to being a social scientist who happens to do evaluation work. Regardless the title they use to self-describe, people who do program evaluation work have difficulty expressing to non-evaluators, whether in personal or professional life, their title and what they do. Evaluators often agonize over how to describe themselves and their work to disciplinary colleagues. Katherine, an academic practitioner,[9] for instance, said, "I think self identification is an important question and it's part of where evaluators kind of flounder. They always hesitate like, 'Uh, well…'" Emma shared her experience with describing her work as an awkward and onerous task:

> It's interesting when you're talking to people who are outside the field because nobody ever knows what you're talking about. Like I say, 'I work in education.' 'Oh you're a teacher!' 'No..' 'Oh, well you're a school principal!' 'No..' You

[9] An academic practitioner, as I will illustrate in the next chapter, is a type of program evaluator who is situated in an academic setting, but whose primary work is the conduct of contract evaluations.

> know? 'I'm a researcher.' 'Oh, so you work at a university!' 'No..' It's this whole realm that nobody even knows exists.

Those who had worked primarily in non-academic settings and had been doing this kind of work for less than a decade, speculated that non-evaluators had no way of pre-categorizing evaluators. This certainly has implications for the study of professional identity self-authoring. Lack of narrative articulation that is easily grasped by outsiders does not signal a deficit of meaning. More important is the interplay between inner speech, interactions with other inhabitants of the figured world, and communication with those representing proximal figured worlds (e.g., clients and other potential sources of conditioning). The inability to communicate adequately to outsiders may further insulate members of this figured world, thereby enhancing a sense of community.

Academically-situated and old-timer (who had, at one point or another, worked in an academic setting) evaluators were able to bridge and make sense of the social scientific discipline and evaluation. For them, being an ambassador to the social scientific discipline from which he or she came made sense or, as in the case of old-timers, did not resonate as much as did the designation "program evaluator." Academically-oriented knowledge production (that is, work not intended for contract audiences) is often reflective and speaks to the interweaving of the unique positionality of the academic practitioner. For instance, one participant positioned himself in a hybrid space between program evaluation and anthropology. He writes for program evaluation journals bringing about an awareness of the liberatory potential of program evaluation, offering critiques of the status quo. But, most of his work is in training graduate students

qualitative methods in program evaluation and assisting students in getting their first evaluation field experiences.[10]

This is not to say that academia is automatically perceived as the high ground. As the following exchange I had with Laura illustrates, power and position change according to particular circles of influence.

> Keith: Ok so just to make sure I got it straight, being a researcher is more prestigious than being a program evaluator but perhaps to a lot of folks in the education industry, especially foundations, evaluation might carry a little bit more clout?
>
> Laura: Absolutely. Once you get out of academia, yes. They just have a better grasp of what that means.

According to Laura, program evaluators are a cut above academic researchers among program staff and foundation folks. They are responsive to client needs and they deliver project reports on time.

For most, however, evaluator professional identity is situational.[11] *Program evaluator* does not resonate with the general public or scholars as might a term such as *Psychologist* or *Sociologist*. Sarah commented, "I don't even bother describing in-depth

[10] In the truest sense, therefore, he is no longer a program evaluator himself, but is certainly ensconced in, helps give meaning to, and enforces the discursive and practicable rules of the figured world.

[11] Borrowing from Mary Waters (1990), I use the term *situational identity* to refer to the status labels such as *PhD*, *Anthropologist*, *Program Evaluator* that may be used with varying degrees of usefulness (i.e., cultural capital) according to the structured situation.

what I do until I know that there's some kind of not even just interest, but some type of context to hang it on. Otherwise, it's just frustrating." She went on to describe how she describes herself to others and her speculation as to why she describes herself that way:

> I typically introduce myself as a researcher because a lot of this has come from the experience of having people kind of stare at me blankly. I think I have been possibly conditioned to explain myself this way.

The situational aspect of program evaluator identity takes the form of an internal heteroglossic (Holland et al 1998) community. The orchestration of different voices, the self-author of a program evaluation identity does not, until perhaps several decades of practice in the figured world, conclude as a single label.

That evaluators have difficulty articulating what they do is indicative of its distance from public space, a trend that Popkewitz notes about education research in general (1990). It also highlights the difficulty of making a case for why anyone should listen to what an evaluator has to say. My intent is to draw attention to these because the source of discovery of the figured world—that which enticed them into the program evaluation world—tends to be either serendipitous or intimately familiar. That is, everyone found this work either through an unexpected job ad, a professor who needed or provided opportunities for graduate students on projects, or requests from professional networks to help on a project. Nonetheless, I do not contend that program evaluators are victims who lurk in the shadows of educational reform. I believe the uncertain or marginal status actually serves as a source of power and freedom. The narrative is ever-

shifting. In fact, to nail down an elevator speech or widespread distribution of a biosketch calcifies the position of the program evaluator to the extent that she may find it difficult to compete for subsequent contracts in slightly different areas.

Entry into the figured world is contingent upon its discovery and playing on its fringes, where thought meets action (Vygotsky 1978). While Foucault's view on habituation is a compelling bodily manifestation of activity and thinking (1979), inscription, alone, is an insufficient explanation for auto-scripting since it ignores envisioning and making other worlds by rearranging bits from familiar figured worlds. Personal agency is not the creation of a self that is always uniquely one's own, but it takes shape in a field of contest, the "space of authoring" (Lachicotte 2002, 61). As I mentioned earlier, self-authors create their identity through orchestration of a multiplicity of important voices within and against constraints that also provide space for their voice. This opens the possibility of experimenting with program evaluations; identifying spaces that move beyond the interests of the status quo and, therefore, human agency, in what appears, at first glance, to be a sharply-controlled industry.

Here, I take a look at how evaluators were enticed into the figured world, their reflections about why they approached it, and some of the boundary-maintenance and other structures that helped or hampered their entry. Several were enticed by experiences in graduate school. Bonnie recalled the moment she discovered the existence of program evaluation:

> In my first semester of graduate school, we had to take one of those one-credit "Welcome to the Psych Department" things. It was about introducing us to all the

> different professors in the psych department and the types of research they did and one of the lead professors did community-level substance abuse program evaluations. I saw the fit and I was like, "Wow, that's what I wanna be when I grow up." It was real. It was out in the community. It was dealing with real people and real problems.

The one good class theme seemed to dominate quite a few stories. Katherine, who had decided to study clinical psychology, said she discovered and "fell in love with" evaluation when she took a required methods course. Similarly, Mark described how, when in his master's program for counseling, a faculty member said, "'Hey, Mark I'm getting a big federal grant and I'm looking for a research associate for my grant. How would you like to go on for your doctorate and join the project?' That's how it happened. I said, 'Sure.'"

To get a slightly different angle on entrée, I also spoke with several evaluation company managers about what they look for when they are hiring. Aside from the expected expertise in methods and good writing skills, they also looked for people with team-work experience, willingness to, as Beatrice put it, "Take some dimension of a project and run with it so that it's really yours to focus on," and an ability to adapt to new kinds of work. The manager of a top-notch university policy research center said that he posted an ad recently for a position in which "We were looking for a quantitative researcher, but who could get along with qualitative researchers."

On this last point of what they seek, during the interviews, some managers realized that what they wanted to hire was someone who, at least academically, reflected themselves. The director of that university-based policy center, for instance, said:

> What I was always looking for was a sociologist because I am a sociologist. And, there was always a plan for a project so it was a matter of how do you fit with the existing projects but also a sense of, are you going to be able to fit with the next set of projects? That's why I think sociologists, so we'd all be looking at the same kinds of things.

While not at all universal, as I will show in coming pages when I describe inter-epistemological teaming, this practice of disciplinary propinquity was not uncommon. Although perhaps a way of expressing a kind of built-in, "Hey, I understand you," orientation, the net effect on the evaluation knowledge that is produced and disseminated may also suggest a kind of disciplinary hegemony. Whether an insider appreciation or an intentional exercise of power, hiring is one method seasoned evaluators use to maintain the boundaries of the figured world.

Social scientists enter the figured world of program evaluation under the auspices of socially-positioned gatekeepers. I recall my first boss joking during the interview when he realized that he knew a few members of my master's committee, "Remember, it's not who you know, it's who *I* know that matters." At the time, I did not realize the amount of truth in the statement. As Beth, the CEO of a contract research organization, said about new hires:

> If it's a newbie, they need to have the credentials, they need to have the book learning and practical experience and maybe have a mentor that I have respect for.

While this approach does not always pan out, it is indicative of another way in which the boundaries of evaluation are maintained. Therefore, it favors those with existing insider networks (i.e., social capital).

The Road to Program Evaluation Proficiency

Although I have depicted evaluators according to temporal points in their careers—from initial entry to resolving initial struggles to becoming proficient—a more accurate depiction would not rely on these seemingly-discrete periods. As Katherine points out, the road to proficiency in program evaluation is recursive and continuous. Specifically, she said, "I don't know that it was any one key moment, any one key turning point actually. I think it's ongoing." I do not mean to deny hallmark events of their potential for rapid change, however. Katherine, and others, described to me what they believed to be "pivotal moments."

Learning to be an evaluator is not a systematic process like graduate school. Most of the evaluators with whom I spoke described their participation in required professional development including continuing to take graduate methodology classes (which is how I ended up back in graduate school working on this dissertation, incidentally), and participating in seminars hosted by trade groups such as the American Evaluation Association. Others learned by self-directed reading. Helen recalls, "Because I was self-taught or I learned just through doing, practicing, I don't feel nearly as well-grounded in

evaluation as a separate and distinct discipline or field as I do in sociology." All learned through experience, however, most notably, working with new conditioning agents (such as interdisciplinary work teams). Getting in was just the beginning of their learning about program evaluation. Much of the reinforcement about what makes a "good evaluator" came about through the rebuke and praise of the work team. Here is a brief exchange I had with Helen:

> Helen: You only have to make a few of those mistakes before you don't make them anymore.
>
> Keith: [Sarcastically] Sure, I'll do those 187 interviews, no problem! I'll do the transcriptions too!
>
> Helen: I know this about myself and it's a little bit of a weakness, but people started praising me and saying, "You're doing good work" and "Your writing is good" and "You have a feel for this" so that sort of helped me think, "Well, maybe I can pull this off."

Potential recruits to this figured world bring with them several dispositions. As I explore in greater depth in chapter 5, prospective new evaluators who make it through the first set of hurdles want to make a difference, possess a repertoire of methodological tools, are able to build and maintain relationships, can thrive in a competitive work environment without fixed salary lines, are willing to adapt to new topic areas, and are open to new epistemological approaches. But, what do these entail, exactly, what are the major learnings, and how is this accomplished? In this section, evaluators share some of

the initial surprises and challenges they experienced during the first few years in the field and on through their careers. The interconnected threads include learning about the business mechanics of evaluation, re-routing theories of humans to theories of program action, gaining a savvy client orientation, and learning how to operate as a member of an interdisciplinary team.

The Mechanics of Program Evaluation

The contract implies a constellated structure with several interrelated points that include foraging for optimal and continuous resources (e.g., RFP), guarding against "scope creep" (doing work in addition to that which is stipulated in the contract without additional compensation), and being ready to adapt quickly to change. Emma said most of what I want to convey regarding the mechanics of this work:

> The mechanics of it is where I had the biggest learning curve, like learning how contracts work and writing proposals and what was involved with that. I mean as far as designing research, collecting data, analyzing data, I knew what I was doing there. It was more the logistics of I guess kind of the business side of it that was where I had much more a learning curve. And especially things like the difference between a contract and a grant and what you can do under those kinds of circumstances and the way that you're constantly scanning the field for RFPs [requests for proposal] and what an RFP was and how that differed from a request for applications. You're funded by your contract so that defined your scope of

> work. There's not really anything outside of that. That was all the stuff that was new for me that I had to learn a lot about.

She further noted, "It is very hard to specialize in something because you just kind of live or die by whatever contracts are available at the time." She referred to topic area rather than methodological approach.[12] The point here is that the contract delineates the focus of the work and that has implications for how research is conceptualized and designed. Helen noted:

> You don't have permission to go asking questions that the client doesn't really want you to ask. And there's no money or time to do additional analyses either. It's focused on very particular kinds of questions.

Mary also commented on the practicality of evaluation work. She said, "Knowing the practical way of applying it in a world where there's budgets and resources and people who are always interested in some of the questions you could possibly ask." The contract, both in terms of the scope of work and the budget, becomes the driving force for day-to-day decisions. I will return to the scope of work and its implications for knowledge production, team-building, and internalization of the ERD throughout the rest of the chapter. But, first I want to extend my discussion, briefly, about budgeting contract evaluations.

[12] Evaluators do specialize in the methodological approaches associated with their disciplinary origins, which provide them some opportunity for specialization, as I will illustrate later.

Evaluators learn how to commoditize their services. Anyone who conducts grant research needs to learn to do this, too, to an extent. The contract, however, demands a level of detail that often requires specific tasks. One evaluator commented:

> One of the big learning moments for me was sort of the commodification of my services. I learned how to do interviews and things like that in graduate schools, but I certainly never learned how to sell them. Putting a dollar amount like $2500 per focus group, and coming to that sort of conclusion was hard.

The slightest deviation from the contracted set of activities may signal distress from clients.

The focus on money is not always about profit margins, but it is about keeping jobs. Because most evaluators work on soft money and the contracting organizations are often tenuous and dynamic (according to available and awarded funding streams), budgets require ongoing attention. As James noted about one of his first evaluation jobs:

> The Department of Education shut down the REL. They closed it down and the Washington people closed the whole thing down. Most of the people just lost their jobs.

When this happens, especially in locations where no competitors are present, researchers are dislocated. Most evaluators have faced experiences like this one either directly or indirectly. Even with projects that are already funded, shifts occur. Mary, for instance, described a federal flow-through project that had been funded for three years, but for which the state, which was the funding agency, adjusted the scope of the overall reform:

> We had to sort of play it by ear and made adjustments each year as the budget changed. By the third year, they didn't get as much as they were anticipating. So we had to shift downward as well as they cut down the evaluation budget. That's normal.

While this might seem commonsensical, it is a formidable transition from graduate school models of tenure tracks and relative autonomy. Many organizations do not offer formal training or support systems for budget monitoring and preparation while they have tedious requirements for monthly reconciliations (as a measure to be prepared for potential audits). One evaluator who had just moved to contract evaluation from an academic context, said:

> In both places that I've done this, I end up being in charge of creating the budgets and working up contracts with partners. I don't like it. And in fact, I'm going try really hard to make changes so that I don't have to. There's nothing wrong with it, and I understand it's a reality. I don't like the administrative side of it.

Those who had their earlier experiences in non-academic settings who then transitioned to the academic world, however, brought with them a keen understanding of and level of proficiency with this role. Two of the university-based center directors, for instance, were able to bring in and manage large projects and produce for both clients and their academic roles.

These pragmatic facets of learning the business required reworking previous understandings of how social science knowledge work may be done. They also laid out a

structured set of rules and obligations that delimited and drove notions of worthy projects and the work that might be accomplished on those projects.

From Theory of *Man* to Theory of Action

In addition to the mechanics, participating evaluators told me about how they modified their thinking about the subject of study. More specifically, they felt a need to shift away from social scientific theory to localized *theories of action*.[13] Some did this creatively and on their own while others learned formally. Either way, most struggled with transforming their assumptions about the purpose of social research.

At odds with earlier convictions about social change, when program evaluators don the mask of non-partisan voice of the ERD, they begin the internal dialogue of what I refer to as *de-stancing*. They cannot afford to take a strong stance in literature or liberatory perspective because of political implications and the nature of change in education reform literature. Being overly identified with one theory of educational reform or notion of pedagogical approach may complicate adaptation to new initiatives and perspectives. But, it also goes, clearly, against the rudiments of the ethical orientation of evaluation. This does not mean that evaluators do not care about the outcomes or that they have nothing at stake, but that they adopt a persona of dispassion.

[13] Here, *theory of action* does not refer to Bourdieu's use of the term, but, instead, a widely-distributed program development and evaluation notion referring to what planners and implementers believe is going to come of a particular reform model when implemented according to plan (see, for instance, Archer 2004), which often results in a logic model or other heuristic device. In turn, the theory of action and its associated offshoots become the basis for evaluating a program's implementation and impact.

Isabella, the hired gun, who also self-identified as a practitioner (the director of a non-profit outreach program), was surprised to learn that evaluators were not intimately familiar with education content areas and that, perhaps more so, their feedback was not intended to influence practitioners who are involved in reform programs directly. Her biggest concern, however, stemmed from something I said during one of our conversations—that few people intentionally entered program evaluation. To her, this signaled a considerable distancing from practitioners:

> From talking to you, it seems that many people don't plan to go into evaluation, intend to focus on evaluation and they end up doing a lot of work in that area and that there really is this whole society of evaluators that operate a particular way. Because I still define myself as more of a practitioner. It always surprises me that a lot of times the evaluation isn't, because of the funding criteria or because of the goal behind the grant, that it's not necessarily set up to help the practitioner improve. I think that was a surprise and a disappointment.

Isabella "read" the function of evaluation as a direct feedback tool, a way to improve participants' practice immediately.

By entering as a practitioner and without experiencing full entry into the figured world, she sidestepped the *de-stancing* that is made possible by coming to understand evaluation as apolitical and unbiased knowledge work, both a moral obligation (in a liberal democratic ideal) and a political tool. This de-stancing is a crucial step in the transformation from social scientist to evaluator. To an insider evaluator, Isabella's

concerns miss the mark that goes back to the mismatch between the purpose of evaluation—both from the inside and in its mandated language—to how it gets read outside of that context (which includes the subjects of study). To Isabella, evaluation's usefulness is embedded in helping teachers, here and now, through feedback. But, to evaluators, the idea is to influence somewhat bigger-level change (i.e., the program).

The lack of consensus about what constitutes big-level change is apparent. As I have noted, evaluators become social scientists to understand patterns of human interaction and behavior. They then apply this knowledge to systemic change as evaluators. But, along the way, they exchange a modicum of the focus on a disciplinary-inspired theory of some aspect of humanness for, instead, programmatic theories. Anne, for instance, recounts this shift:

> I needed to learn to frame the research questions in relation to what a program was trying to accomplish. If I think back, I can almost recreate that feeling of like gestalt shift or something. I'm not driving the questions. The program is driving the questions. And there's a kind of a very interesting subjective shift and I think that has to happen to other people because I've worked with folks who are researchers who work on evaluation projects who can't make that shift, which is that you have to somehow balance what's compelling to you by virtue of personal or intellectual interest or whatever, or academic history of training you've been part of and what would be meaningful to the program people. That was a really important shift… It was like maybe shifting from doing algebra to doing geometry. You have to undergo a cognitive shift.

Two particular activities helped evaluators undergo this "cognitive shift" from social scientific theory to program theory: 1) learning about theories of evaluation and 2) understanding the context of evaluation. Helen, for instance, who had been trained as a sociologist, said, "I'd never read any program evaluation theory or practice. So I had to work on all these new names and theorists and schools of thought and kinds of evaluation." Beth, the CEO of a contract research organization, made a similar comment:

> I needed to be really well grounded in the theories of evaluation because I think it's like research methodology. How you proceed is always driven by the nature of the problem, but then you're looking for the model that best suits the issues on the table. You have to be really well grounded in that. I think you need to be open to doing things differently than you've ever done them before. An innovative approach might better meet the goals of the evaluation, but you need to do that in a way that lets the client know that this isn't necessarily tried and true. There're always unknown elements in program evaluation. It's not like you're just trying to find out if it works. You're trying to find out how it works, maybe why it works, maybe when it works.

As Beth noted, selecting an appropriate theory of evaluation entails addressing the research questions, but it also entails taking into account the clients' needs.

The second requisite body of knowledge is understanding the context of evaluations and the policies that both help shape educational reforms, as well as the policy-makers that might be responsive to evaluation findings. Helen, for instance, notes:

> I didn't have a really clear picture of the policy world or the federal or state and local program world, and I didn't understand how it all fit together. So, I had to learn all the different layers and requirements and competing priorities and institutions because I didn't have a map of that in my head, and without that, I really didn't know how to think about what the point of evaluation was. Who was it serving? What was the point of doing it? Who were you talking to? What change could you hope to effect?

In essence, evaluators adapt theories of humanness while they learn about micro- and macropolitics related to education reform. This crucial step is the endpoint to some evaluators' careers, since giving up the theories related to their discipline feels like (and is, to a degree) giving up disciplinary purity. Those who proceed, however, use this cultural knowledge as a way to speak to new audiences and stakeholders (i.e., clients). To speak to clients, they maintain the appearance of political and moral de-stancing, which permits evaluation to operate as business, practical ERD, and potential agent of change.

Becoming Client-Focused

The contract defines the scope of work and, ideally, the focus of study with sufficient detail and clarity that both client and contractor share expectations about the structure and organization of the content of the final report (deliverable). This increased the likelihood that clients get what they need and that evaluators are not asked to toil endlessly on a project, doing what is known on the inside as "scope creep." The contract

lays out the beginning and end dates, the questions to be answered, the methods by which they will be answered, the theoretical approach, and a description of the deliverables. It does not, however, dictate the amplitude of the individual evaluator's voice. The space and volume of that voice is negotiated dialogically between two mutually conditioning structures—the client and the contract research organization.

Based on evaluators' recollection of experiences from their first few years of work in this figured world, I explore some of the ways that new evaluators learn to amplify their voices or come to feel (or act in ways that suggests) they must be silent. Becoming client-focused does not mean breeching ethical standards. In fact, most evaluators experience, at some point, pressures to beef up findings that align with the needs of the client or to remove those that may be deleterious to its reputation.

As the next few examples illustrate, these pressures are activated and responded to in different ways. Helen, for instance, told me about how a client demanded, repeatedly, removal of an analysis that was framed in Bourdieu's language about *social capital* including statements from community members that went against the grain of the view of program staff. She said of the event:

> I felt stuck. I felt like we were compromising the perspective that the interviewees had shared with us, that we were holding fast to some idealized, pure reporting. But, mostly, I felt the hold. It signaled for me that this kind of work really has constraints around it. Particular constraints having to do with who pays you and what authority they have over you. Now, I know my place. I look for

> intellectual engagement around things that might be controversial elsewhere, outside of my work life.

While she later revealed the degree to which she contradicted herself, this post-academic researcher felt constrained by the client and the organization for which she worked. For her, succumbing meant removing conversations about power, as well as removing the voice of a sub-group that was affected by the reform.

Similar to Helen's example, Mary recalled an early contract negotiation in which she began offering input on the project's conceptualization when her supervisor, who was also present, began deflating her lofty ideas. She said, "My boss was literally downplaying what we could do. I didn't understand. I was like, 'Why are you being so discouraging?'" Mary went on to note that she learned at that pivotal moment that because of the contract, evaluation is limiting. She said:

> I remember having to come to terms with the fact that there are sometimes political reasons, sometimes sensible resource-oriented reasons limiting what you can do and what makes sense to do…Some of the best questions to ask for a project aren't the ones you actually get to evaluate or work on.

Like Helen, one of Mary's internal voices demanded that she "come to terms" with, accept, and adapt to the preexisting exchange between client and contractor.

These examples have implications for *how* to communicate as much as they do to whether or not one should or does feel capable of communicating in an agentic sense. *Coming to terms* was not a strategy shared by evaluators who had been in the field for

several decades. Nicole, for instance, described a high profile project in which the agency director did not like the results. She said:

> After adding whatever I felt I honestly could in the way of context and history and good intentions, I gave the agency a choice: don't change another word if you want my name on the report as required for a third party review or if you want to re-write the report, take my name off. The report is still "under review," a euphemism that avoided the choice, alas.

Anne recalled a more general approach to producing deliverables that were both palatable and that were grounded at the ethical juncture of liberal social science (i.e., neomarxist-influenced analyses of power) and program development according to stakeholder needs. Specifically, while she decided to comply with contract contents of deliverables, she identified spaces for the voice of anthropology. Of this decision, she said:

> There wasn't anything I felt I had to give up. Maybe initially the reason I didn't feel like I had to give anything up was because in writing up these evaluations we put in, I don't want to say I invented this but I didn't know about the kind of formulas for evaluation reports or anything, I just wrote reports and there were certain things the funder required to have, headings and stuff, but when we put in this reflections and interpretations section, that was a place where I could kind of lapse into talking about the program in a broader context. And that allowed me to

> do more ethnographic kinds of things to bear but in a way that was relevant and not just for my own edification.

She added:

> I guess when I am involved in evaluating programs that there is some facet of them that I disagree with politically, that's difficult. The way I try to deal with it, to the extent that findings challenge assumptions, is to add to the discussions.

Not being socialized into report production in a structured group setting permitted her to bypass the felt need that evaluators, such as Helen's and Mary's examples above portray, to downplay personally resonant work. I will return to the theme of group work and its effects on writing in the next section.

What I would like to emphasize here is that the willingness to enter a pre-existing discourse may be a sign of succumbing to, but it is also a method of entering the figured world in order to perform pragmatic work according to a contract. Embedded in this strategy is a declaration (and acceptance of that declaration) that evaluation is a helping industry. A major role in evaluation, for instance, is helping clients understand what can and cannot be measured and ways in which evaluations are more than required actions. For instance, Beth said:

> I needed to learn how to ask questions in a way that wasn't intimidating or threatening, but that would lead the client to think in maybe different directions than they had been thinking, or to clarify their thinking.

Beth went on to explain her role as "translator:"

> The role of a program evaluator is often in a consultation model, where it's a lot about asking questions and helping them answer their own questions. It's as much about that as my bringing a specific solution to the table.

Assuming a client focus in this helping role also requires using data collection and analysis methods and communicating findings in ways that are accessible to non-researchers. And, it implies completing work on time. This combination of accessibility and timelines translates into learning to prioritize efficiency before social scientific rigor. Laura, for instance, recalled an early realization about dumbing down her analyses while working for a non-profit contract evaluation organization:

> I do a lot of regression modeling, trying to come up with hypothesis and defensible things. We didn't do any of that at [organization]. I wasn't resentful but I was sort of surprised at it because I'd never done education evaluation. So, I felt like what we did was more just like describing programs rather than research. Through our descriptions there's like a weight of the evidence kind of thing, but I don't know that we did anything that people would go spend millions of dollars on just because we said it was good.

Methodological "bending" was problematic for most, at least at first. Ethnographers, for instance, were required to focus, psychometricians did not have clinical control, and economists had to deal with data that were not uniform across sites.

For most of these program evaluators, becoming conversant with unfamiliar methodological approaches was not only mandatory, but opened up new ways of thinking—the "cognitive switch"— about the work. Sarah, for instance, who had been trained in survey design and experimental methods in her doctoral psychology program, recalled the dramatic shift when she learned about, and then embraced, the use of qualitative analysis. She noted:

> The ability to conduct an interview and get twice as much out of it because you're picking up on a word here or there that's really, really pertinent to the study but might not be on the interview protocol or might not be a person's specific answer to the question that you put on the interview protocol that you might respond to the interview questions, so it's that flexibility and ability.

There is a double-edged sword in methodological expansiveness, eclecticism, and multiplicitous approaches. I will come back to this, but first I want to describe the influence of group work on identity production and everyday work.

Locating Agency on the Team

Most program evaluation knowledge production is completed as team work. And, usually that team work draws from methodological and conceptual skills of social scientists representing several disciplines. Participants described work groups consisting of anthropologists, sociologists, measurement statisticians, economists, and so on working on a single evaluation project. Becoming a functional multidisciplinary team player is not an easy transition for everyone (or, perhaps, anyone). It requires re-learning

or negotiating the meaning of discipline-specific terms. Consider the term "culture" and the way it is used in different ways across social sciences. And, as I demonstrate throughout this section, group work has implications for agency and conditioning. In other words, teams reinforce the rules of program evaluation's figured world. Because of their conditioning role, teams serve as semi-permeable structures. Their semi-permeability is due to the vast spaces of contestation. Disciplinary hegemony is seldom complete, even in the context of mandated epistemologies (as was attempted with the National Research Council's contribution to NCLB).[14]

Learning to work in a team environment takes time and effort. Mary, whose story is consistent with those of others, myself included, commented:

> Working as a team member on projects helped because we had the benefit of learning together and figuring out how we were going to do it together. I would say that it took me probably a year to feel proficient and confident in what I was doing. Just feeling like I had a handle and kind of surprised myself after a while when I could not be nervous and go in and have a discussion with people, even talk off the top of my head and be able to help someone. Like, how could we go about evaluating this project? About my third year of working at this job, I was also able to go to evaluation initiatives and really focus on looking at evaluation standards and being around other evaluators.

[14] This refers to the NRC's specific recommendations to the Bush Administration that effectively put into place a hierarchy of research designs. Not surprisingly, the economists on NRC staff produced language that made experimental and quasi-experimental designs optimal.

Evaluators with whom I spoke tended to fit into one of two categories: 1) those who learned to work in groups during graduate school and 2) those who learned to do so later, in corporate settings. For those participants who began their data collection and analysis experiences in program evaluation while in graduate school or right after, group work felt normal. But, some, especially those transitioning from academic settings, described a challenge in learning to work on teams.

In an early contract project, Mary came to find out the need to stay in touch with both her client and her team. She recalls:

> I had an experience where I worked with an external client and I didn't have the knowledge of how to set up things where we could regularly communicate with one another. And, the client didn't communicate with me. We got down to a deadline and they were on a crunch on their end and they wanted something that we hadn't even collected the data for. They got really nasty about it and I had to face the music on that in front of the boss who did not defend me at all and let me be dressed down in public. So, that happened my first year.

The transition, in some cases, required learning to survive. David, for instance, noted:

> When I was in graduate school, I was doing research on my own. When I was in [company] in South Dakota, I was doing research on my own, I had to keep the people in the school district happy with me, and I had to keep my bosses more or less happy with me, but they were thousands of miles away. Here [in the policy center], I had 2 or 3 bosses of various sorts and project officers and all of those

> kinds of things. That was the trick of surviving and doing well in the environment was keeping people kind of working in harness.

While Mary was forgiven (or forgave herself), David learned to make himself indispensable among coworkers and management and to build a client base. Relationship building and making oneself generalizable to shape-shift for new contract work, yet specialized enough to offer a unique set of skills, is the balancing act most evaluators internalize. Regardless of how individuals learn how to (or learn not to) work in groups, teams are selected and built. Mark remarked, "Teams are made, not born. The most difficult part of evaluation work is being able to build relationships."

While there is little content area or project specialization on evaluation teams, depending upon how they are socially positioned, individual evaluators may create niches within a specific contract or set of projects. I will return to this theme. First, however, I offer a brief discussion about the implications of teaming for epistemology and, therefore, knowledge production as a process and the ways in which that process becomes modified in program evaluation.

Mediating Eclecticism

According to participating evaluators, multidisciplinary group work is eclectic, which makes it both limiting and expansive and, in particular, a means of adequately answering a multiplicity of kinds of programmatic questions. In the earlier days (1960s and 1970s), evaluators were expected to master the gamut of methodological approaches. Scriven, for instance, wrote that in order to be considered an evaluator, one would need to

be proficient with a full set of qualitative and quantitative methodologies and research designs (1996). While, unless running a niche-oriented shop—such as a sampling statistician who works exclusively as a subcontracted consultant—most need to be conversant about multiple methods, the teaming approach fulfills the call. As one guru who became an evaluator at a time when they were expected to, and possibly could, know it all, said:

> Most of us haven't sufficiently encyclopedic substantive knowledge for policy-level, for complex, or for larger evaluations with many sub-questions. Downsides are careful delineation of the scope of work to keep costs controlled and the need for extensive and skillful communications.

Her comment brings the conversation back to concerns regarding client focus and efficiency, but it also underscores how the call for mixed methods is, in great part, a pragmatic move. The most concise rationale is twofold: it is both too much work for one person to do alone and simultaneous outcome and process questions require quite a spread of epistemological frameworks. Beatrice noted:

> At this stage in the game for me [as director of a major research and evaluation university center], having a really competent group who can reliably put forward a mixed methods approach to conducting research is key. In other words, I can't do this by myself.

Emma added to this logic:

> You have to have a team because you're usually looking at more than one kind of research question and more than one methodology. Even if it's all quantitative data, you might be looking at analyzing achievement data and analyzing survey data, and you need to know how to really deal with all those things. It's hard for one person to do that.

Having multiple disciplines and methodological approaches at the table also contributes to the ways in which evaluations are conceptualized. At the outset, sharing conceptualization space with folks who seemingly naturally think quite differently (e.g., top-down psychologists and inductively-oriented ethnographers) is not easy. After a few successful negotiations, that "cognitive switch" begins. Beth described her view:

> Different disciplines bring such a different world view to the table and there's always something to be learned from that view that can be brought to the thinking about the design for program evaluation in my own field, which would be psychology in education. I think it always makes it richer.

No one, including managers, has the complete picture of an evaluation project.

Complex constellations of specialized analyses and studies within studies complicate directors' likelihood of having a complete understanding of the whole project. Beatrice, for instance, described to me a foundation-based evaluation her shop had been awarded recently that will combine ethnographic methods with Hierarchical Linear Modeling (HLM). She said, "I know we're engaged in one of our studies in propensity score matching and in other kinds of analyses that really I'm not skilled at." While

evaluators engage in what appears to be struggles for methodological supremacy and enter hierarchies based on an organization's attitude toward and use for particular disciplines, once functional teams are formed, individuals identify the spaces in which they can frame and protect their specializations. That is done, in part, by creating space for each approach in the methods of data collection, analysis, and write up. Another is within-organization training of evaluators with differing epistemological repertoires. Either way, as I mentioned, everyone gains at least an appreciation of other epistemologies (or chooses to move on).

For instance, Sarah, who was trained in experimental methods, described the tremendous difference in the way she approached project work when she learned qualitative methods. She said:

> In quantitative there's not necessarily an answer but there's a close approximation to the answer. Getting comfortable in that space like, "Well we're not quite sure yet and there's huge amounts to look at and we really don't know what's going to come out of it." Getting more comfortable with that has been a challenge but that's another good thing that I found working with [the staff anthropologist] in particular because she's one to bring up a lot of different questions from the start and sometimes I get overwhelmed by that. But it's to the point now that I recognize when that starts coming up for me and I get kind of anxious when that starts to come up. I'm like, "You know what, it is ok. It will become clear and the answers will emerge."

Alice, a cultural anthropologist, offered a similar example of how she learned to appreciate and conduct quantitative research.

> My original orientation was very qualitative. I was much more interested in description and understanding and I was very what you referred to earlier as formative sorts of approaches. Apt Associates was our employer and their design for this project was to have us do the qualitative and they would plan the quantitative cross-site assessment. And so I learned a lot about quantitative methods that I hadn't known before and I also came to value the idea of triangulation more in that context. I did some research using a questionnaire, of which I was originally rather suspicious. I got some data that I never could have gotten another way really. And it also made me more open-minded about some of these other approaches.

In order to analyze the multiple inputs and outcomes and to be able to address a wide variety of evaluation questions, evaluators self-author themselves onto functional teams. The benefits of doing this include increasing the potential for specializing on facets of projects and learning about other domains of inquiry. But, there are also costs, as I have implied throughout this section. Methodological purity may be an ideal that is out of reach because of the pragmatic and client-focused aspects of the work. Many beginning evaluators, who have been taught to respect contributions to their social scientific discipline, reel at the end of their first few contracts when they discover that it

is also the end of meaning making for the particular project. It may also mean end of access to these data, as Emma illustrates:

> We worked on this one big project and that was it. And we couldn't use any of our time to work on anything because there was no funding outside of our [company] contract to do anything else. And, that was really a shock for all of us and ultimately it became very disappointing because it was like, "Wow," you're not really going to have an opportunity to publish any of your own work, to develop any of your own projects. Everything that you do is just defined by the contract that you have to be working on at that time and when that contract is over, your work in that area is pretty much over. You move on to the next thing.

In fact, few teams have the intention of publishing the work beyond the final report. The decision about whether or not to publish or present at conferences is determined in great part by the client (who, in most cases, has intellectual property rights to data and all findings). Some clients, such as the National Science Foundation (NSF), define greater expectations for publication and dissemination than might, for instance, a public school district undergoing a mandatory program evaluation of its High School Redesign grant.

In addition to the client, different organizations have slightly different foci that may hamper or facilitate publication. For instance, academically-oriented shop staff members carve out time to produce publications and presentations. In that context, it might be of value to pursue an agreement regarding intellectual property rights. But, CROs and non-profits have neither the luxury of time nor the reward structures that

promote or support publication beyond the final report. Perhaps more importantly, because of their need to remain de-stanced in order to survive changing administrations, reform fads, and competing clients, CROs are left with fewer options for publication, aside from methodological innovations or personal (group) reflections.

That being said, CROs vary in their orientation toward publishing. This next example explores the way one evaluator dealt with writing chores as he confronted maturation of the figured world. When David came into program evaluation, evaluation-specific knowledge production was relatively new. Several evaluators used academic ways of reporting as the model for writing. It made sense and publication of findings seemed natural. But, after an earlier major federally-funded evaluation, David opened an academic center and gave himself the title *Academic Entrepreneur*.[15] In that center, he worked on several contracts in research and development. He hired external evaluators who, at first, surprised him with their lack of interest in publishing. He recalls a turning point in the writing process:

> I ran a math-science partnership for 4 years and we hired an outside contractor from Metis Associates in New York City and they were a contract shop and nobody there was really interested in doing anything academic. They were very sensitive to what the client needed and very sensitive to coming up with good measurements of what the effects were and very quantitatively rigorous in a sense of understanding what was going on in this project at this time in a more

[15] *Academic Entrepreneur* is a brand of what I have called *academic practitioner*, as I will describe in the next chapter.

> pragmatic sort of way. She really doesn't have any academic publication type interests at all. It's more about doing good work for clients.

This was eye-opening for David, who decided that, ultimately, it would benefit him and his venture to continue publishing the work. I will return to some of the reporting lessons in the next section, in which I explore becoming a proficient evaluator.

David recalls how at some of the organizations for which he worked "We did publish for journals but it was sort of not part of the normal culture. [The company] didn't care much what we did as long as we got our reports in on time and they did a certain level of quality control." He went on to describe another company for which he worked that included a formal review process. When he got to the university-based REL, he noticed:

> There was always an understanding that everybody was tenure track or going for academic promotions. [The REL] put out a lot of publications and was very aggressive about getting things into educators' hands and policy makers' hands. But they knew that people needed to publish things in journals and were very happy about that. And in fact, some of our products were like edited books. The books probably were in between and then you could do what you wanted to do.

For most evaluators, the technical report or similar deliverable is the only publication they produce and that form of report usually involves pieces written by various co-authors. Thus, the script that must be internalized is one that, unlike academic work,

results in little individual recognition beyond the work done with clients and participants directly.

Meaning Making and the Team

Most central to everyday work as a social scientist and a program evaluator is the tangible result of meaning-making—instrumentation, writing and reporting. The evaluators with whom I spoke provided accounts that situated the knowledge production process between making a peaceful and constructive space for all disciplinary voices at the table and a hotly-debated or even hegemonically-contorted harnessing of those voices. This is at great odds with the approach most novice evaluators used on their first evaluation—treating it as an academic exercise—a style of writing that did not translate well to program evaluation. Regardless of the micro-political context, new evaluators experience the conceptualization of new work with group members who bring disparate, and sometimes incommensurable, preconceptions, assumptions and ideas about what can and cannot be measured or described. One strategy for sorting out this diversity is divvying up the work while orchestrating it so that the various streams flow together. Emma provides a great example:

> We had this case study. We had a whole chapter in our report focusing just on that, and that was our anthropologist's lead on that. And then we had another chapter on just looking at student achievement outcomes, and the assessment guy took the lead on that. And then we had this survey and these interviews and I ended up taking those and kind of combining those and looking at the surveys and

> interviews together because it was on kind of a similar topic area so it was like the mixed methods person doing the mixed methods chapter. And it worked out really well and we all got along real well and understood each other and had a lot of respect for each other and we did a good job on our components and so we were able to kind of look across that and make everything flow together. Because a lot of times when you do that, it can read like it's just completely different like there's no continuity.

Of course, the process does not always go so smoothly. Sometimes what is considered measurable and important enough for inclusion in the report is carefully controlled. This reminds me of Trouillot's analysis of power in the production of knowledge. Specifically, he contends that:

> The play of power in the production of alternative narratives begins with the joint creation of facts and sources for a least two reasons: … facts are never meaningless … and … facts are not created equal: the production of traces is also the creation of silences (1995, 29).

What I wish to emphasize here is that no matter the functionality of the group dynamic, power is always present in the production of knowledge and facts. Just as important as what is said is what is not, or cannot, be said. Thus, learning to speak and learning to be silent are both imbued with power. In the next section, I will explore what is described as and appears to be a commonsensical component of knowledge production in the presence of the contract.

Efficiency of Knowledge Production

In order to stay in business, companies stay within budget on contracts. To do this requires pretty accurate estimates and monitoring of each task. One of my employers, for instance, created a pretty sophisticated budgeting system that required estimating the number of pages for the final report. Based on the number of pages, an algorithm generated a dollar amount that included researcher and research assistant time, quality assurance review process, management oversight, budget office time, allocations for the folks who would work on the report's layout and printing, overhead costs, and materials. The difference between a 50- and a 75-page report could be thousands of dollars. New evaluators learn soon that budget monitoring and project management are critical. Furthermore, they learn that specified contents of the contract must be delivered, including the deadline. Mary described, in brief, her early experiences:

> Some of the early projects took it out of my hide. It comes down to, by the time you do 2 reports, you're beyond budget and hours and so I'm working at night and all this kind of stuff to get it done. Anyways, you just did what you had to do. So it was pretty rough.

To compensate for this, some mid-level evaluators picked up short cuts they leaned along the way to proficiency, including recycling and using a standard reporting format. Terms, such as "economy of scale," are repurposed to describe acquisition of similar projects so that the same conceptual features, instrumentation, and analysis procedures (e.g., SPSS syntax and graphs) may be used for multiple projects. Helen

described one of her first project experiences, "I remember looking around the office for protocols for documenting the telephone interview and surveys." And, here, Mary described the "standard" method of reporting she had devised after doing evaluations for a few years:

> The data analysis was pretty standard. We had the survey, I believe at that time, where I had them complete the survey online. I transferred it into something that I could use…It's really nice when you have your own surveys and know how the data is laid out and to put it right into SPSS. Then, I just did basic statistics, frequency and standard deviations and so forth. I usually take out the open-endeds and put those all together and kind of look for some themes or commonalities and how they're responding to the questions that were there. And, then I would write that up as part of the report and then looking back at the report, it would have just covered the summary of what the project was, what the evaluation methods were, all the basic kind of things that you would put in to kind of tell the difference between the years. I would summarize the prior year's findings and the issues that came up in that current year, and then present the data. So usually I would do survey data and then the interview summary qualitative kind of things would come after that. And then just the summary, which would kind of reflect back to the evaluation questions and try to come up with some recommendations that would lead to the following year and kind of just looking at the lessons learned or the things that were unexpected as well. And, then attach in the appendix, samples of the surveys.

The frankness of the production and assembly of facts contributed to the attainment of a state of de-stancing. The knowledge produced for clients is manufactured in the examples above in an assembly line fashion.[16] It is matter-of-fact and devoid of too much interpretation. Or, so it seems. The group and the figured world of program evaluation reinforce ERD's tenets through this appearance of commonsensicality. Evaluators, from their initial entrance through the first several years in the field, become proficient with approaches and learn to speak within the existing ERD conversation. But, learning to speak beyond that is a wholly different realm.

Learning to Speak

Not everyone on a team has the same voice. Status indicators serve as structural boundaries that define ontological positioning and, therefore, determine who is granted the right to speak. Evaluators sometimes work with teachers and other practitioners who are content-area specialists, curriculum developers, outreach workers and foundation staff. This gives them an insider perspective. Such relationships, however, are seldom permanent; these key informants are brought on to serve instrumental purposes, as organic consultants, but do not share in decision- or meaning-making. The evaluators with whom I spoke saw this necessary connection a difficult one to navigate and tended to reveal a somewhat vigilant stance with regard to evaluation's boundaries. Helen, for instance, said:

[16] Other examples provided throughout this chapter illustrate a counter position in which each report is fashioned uniquely. From my limited observations, recycling of textual material and use of "standard" approaches prevail in these cases, as well.

> I've worked with educators before, like they've been data collectors. And they're interesting because they can do more expert review because they knew, for example, literacy in a way that I definitely didn't. So they brought a lot of content expertise to the development of an instrument, for example. But, they weren't really trained as evaluators either, so there was a bunch of monitoring that I had to do to make sure the data that they collected or the things that they wrote were evaluative, were based in data. That the evaluative things that they said were clearly linked to data, rather than their own kind of intuitive, "Well, I've been around for years and I know X." Their content expertise was both a strength and a challenge.

As I pointed out with the reference to Trouillot's work, the existence of a hierarchy of knowledge and, therefore, the right to speak, was ubiquitous.

Here, I offer a case of a practitioner hired gun. The example illustrates varying perspectives on positionality and notions regarding who may speak and at what times. From the perspective of the CRO director and staff, Isabella had been hired as a subcontractor to collect data, but her understanding was that she could, based on her experience, offer feedback on instruments and data collection procedures. She recalled an instance, at which I was present, in which she and the director had dissimilar opinions about the most appropriate rating for what they observed in a classroom. Isabella recalled:

> I gave something a 1 and she gave it a 4. For inter-rater reliability, it was definitely bad. And we debated about it for a while until we both just descended into silence. We were both completely arguing our case about it.

Ultimately, the director made clear to her that the higher rating would stand. Isabella reflected:

> From my perspective as a contractor, all the decisions were already made. I was going to use my experience in the classroom and doing evaluation and observing. I was going to bring my perception the table, but everything felt like it was laid out and it did really seem like all the stuff that you all were doing was, there was no way to even get into it as a contractor.

Laura, who worked at the CRO full time, also recalled the event. She recalled that there were what she believed to be two types of hired guns: 1) "those whose livelihood depended on having that job, so they would acquiesce more readily just like people that work for the company do" and 2) "those who didn't actually need that job and came on and challenged the status quo." Of the latter, she recalled how Isabella "ultimately realized that she wasn't going to be met successfully, and just shut down." She added:

> That was interesting to me see that dynamic happen. But do you remember seeing that, Isabella came in and said, "Hey maybe we should do this or that," and she came in thinking it was open-ended I mean this is an active environment where we can start coming up with a new way to do things, or the best way to do things? And it's like, "Oh hell no. Who are you?" But how often do we just acquiesce to

> whoever has the prevailing wisdom. I guess consultants acquiesce more often because if you work for someone, you feel like there's a process to getting barred so maybe you won't get barred the first time you say something that's not to the liking of whoever you're talking to.

Soon after, Laura was asked to resign when she debated with the director about how a final report section would be developed. At that point, she also resigned from program evaluation work. This set of events, combined with her earlier experiences with having to "dumb-down" the rigor of her work, was more than she could resolve. Both Isabella and Laura identified spaces of agency. While Laura redoubled her ethical stance on epistemological purity, Isabella used the incident to reflect on the importance of her stance on helping practitioners and adjusting her criteria for selecting project partners. Specifically, she said:

> The results focus, looking into more from the practitioner side, was what influenced a lot of my thinking about how to do evaluation. If the project or the program is such that I can relate to it in that way and if I have enough expertise in the area then I'm really comfortable working backwards and structuring the evaluation questions or the design. It I have a partnering opportunity where that can't happen again, I either politely decline or suggest a colleague.

I have presented a continuum of experiences and, more importantly, responses to them with varying degrees of agentic action and reflection. With the examples I have provided, however, it is difficult to disentangle idiosyncrasies, serendipity of team

"chemistry," and larger-scale structural forces that are imposed onto the figured world. Yet, those who became career-oriented program evaluators, crafted self-stories that contained reference to purpose.

Sense of Purpose

The evaluators who participated in this study did not follow a teleological trajectory that would lead to proficient and fully-constructed program evaluator. It was a gradual process that required, as I have noted throughout this chapter, action and reflection in a social world (both external and the orchestration of important internalized heteroglossic voices). The initial transition into this figured world was not unproblematic. It was a series of confrontations with and subsequent reworkings of earlier conceptions about what is entailed in social research. When these evaluators entered the figured world, they felt the presence of the hierarchy that placed them onto a lower rung and felt out of place—both vis-à-vis academia and as inchoate recruits into program evaluation. But, as they became increasingly comfortable with and skilled at the everyday work, they also became passionate, creative, and increasingly gained a sense of agency. As Sarah noted:

> I was relatively detached from [evaluation] in the beginning. I was more interested in applying the methodologies. I didn't think about the programs a whole lot, but the more I kind of did the work more and more, I was drawn to work in a place like we are now, which is to say it's relatively immersed in programs and in schools and so I feel like I get more autonomy.

She added:

> I feel like I am helping to give a voice to the little guy, whereas before it was doing a lot of analysis of test scores and that wasn't really a service. In the end it was a service of the little guy, but not overtly so.

Mary described a similar sense of increased self-efficacy and proficiency over time. In fact, to her the work she was doing at the time of this writing was closer to the liberatory ideal she imagined herself doing. She said:

> There's the part where I'm listening, asking questions, helping them think through things, pointing out things about their decisions they were currently making. All that was something I didn't know about in the beginning and it is something I really enjoy now—doing almost, but not quite, participatory evaluation.

Therefore, and expectedly, an increased sense of self-efficacy or agentic control appears to emerge as evaluators become proficient in their everyday evaluation work. Status as an established program evaluator gives the appearance of opening up the potential for soft activism and planned change at the institutional level.

Protecting the Boundaries of Program Evaluation

In this chapter, I have pinpointed some of the components, methods, rules, and valuations that become habituated through everyday practice. More importantly, they serve as reference points in the narratives crafted to connect oneself to the cultural world of program evaluation. Evaluators internalize unwritten ethics that, in turn, respond to

the juncture of moral obligation and de-stanced, data-driven fact creation. Telling the truth, for instance, is a central and naturalized theme. It is a resistance to client pressures, yet, for the sake of client and the targets of reform (i.e., teachers and students). James told a horror story of sorts about what he sees as an emerging threat to evaluation.

> Over the years clients have gone from being sponsors and clients, whose programs being evaluated, to being pretty sophisticated about it and taking over the evaluations themselves. So the drug companies I use as the starkest example of this because people think of these drug evaluations as being the best evaluations. They're randomized and they have a double blind experiments and all that stuff but the fact is that a lot of these studies are highly biased. If we don't get this thing under control, then evaluation will lose all the credibility.

While drug company bias is nothing new, what James points out is the enhanced endorsement of internal evaluation. His story serves as a clear boundary-enhancing tale. It is a call for the protection of the sanctity of de-stanced external evaluators.

Rules of the figured world, couched in ethical terms, help to highlight and maintain the frontier boundaries that separate evaluation from other institutions permeated by the ERD. The resonant calling, the personal fit with program evaluation, aims not only for program improvement, but to activate deeply held beliefs about its liberating potential. The boundaries are maintained, sometimes vigilantly, through strategies including comparing the social contribution of each type of work.

Program evaluators defined “evaluator” in contradistinction to the “other” social science knowledge producer—the academic researcher. Whether posted in an academic setting or in a corporation, the contract evaluator, when doing evaluation, is not an academic researcher. He or she is a social scientist who conducts research, as a service, with the intention of improving particular education programming. An important implication of this betwixt status is that evaluators produce knowledge in a context distinct from academic knowledge. Contract evaluators are both connected to and distanced from academic research.

As Beth said, “You kind of have one foot in the academic world and the other foot in the real world and that’s expected of you so you don't get criticized for moving too far away from the discipline.” In several ways, the academic connection is important to evaluators. Resumes highlight and boast the ivy league status and the publications in scholarly journals of particular staff members (as a selling point). And, accolades from scholarly circles are used to self-promote organizations. For instance, organizations might add to their organizational capacity statements awards received at AERA or social scientific conferences. Recently, I saw a statement that included a passage about how an organization’s work had been recognized by Harvard University’s Top 50 Innovations in American Government Award committee.

This juxtaposition between the contract world and academic research goes beyond the scope, social usefulness, and generalist orientation to social interaction. The comparison between the contract world and academic research also builds on the personal and relational skill sets. In the workplace, through a combination of collaboration on

work teams and depersonalization, boundaries are maintained through group determination of who may speak and at what times and how, ultimately, knowledge is produced and positioned. Helen said:

> One of the things that you don't get in an organization like this is the incredible egos of some higher ed folks. Most of the program evaluators that I know and work with really get that they are kind of in between worlds. They are using social science methods, but for really practical purposes. You cannot be an arrogant asshole. It doesn't work very well. These become interesting places to work because you see the real world, but get to play with data.

Boundaries also take the form hiring from insider networks, which may be expansive. One pivotal moment in the industry I have not yet mentioned is the creation of lists of acceptable evaluators. The following exchange with Mark described the "makings of a professional community." But, the passage has major implications for who gets in and who goes on to become a professional evaluator.

> Mark: In late 90s, a natural step was setting up standards and certifications et cetera, lists of acceptable evaluators. People who now make money off of, including AEA but people who make money putting on evaluation training sessions. So those are all markers of the developing professional community.
>
> Keith: Who, typically, would create and maintain those lists and set the criteria?

> Mark: The professional associations, like AEA, have a locator list. Foundations have their list. I think NSF and the Department of Ed. has a list.

Another boundary-preserving strategy is the comparison of the social contribution of each type of work. Evaluation work is not done with the purpose of contributing to theoretical and empirical knowledge bases. To be an evaluator requires distancing from both the source (social science) and the subject (education). This extension of de-stancing is a gradual accomplishment that is managed through regular maintenance. One example of this kind of boundary maintenance is the trivialization of academically produced research. Bonnie noted:

> Program evaluation is research with a consultative twist. It's not research with a capital "R." It's not all about just knowledge generation. It's about answering people's questions. So it's very applied and I love that about program evaluation - that it can't be conducted in a vacuum. I can't go into a little laboratory and just sit there and do program evaluation. I've got to be out there with the clients. I have to understand the clients' needs and work with them. That's why I say it's more like off-roading because it's messy! It's rare when you can have these nice, very pretty, pristine little protocols and you follow step A, and then step B, and then step C, and you run your little statistical analyses and then you write up your little report or manuscript and you contribute this tiny little bit to the knowledge base, which all of 30 other people in the world care about.

The valuing orientation of this non-academic work is defined clearly through the intermittent diminutives (i.e., use of *little*) used to describe the basic research process.

Underlying the presumption of usefulness is the contention that it is easier to do academically-oriented research than it is to do evaluation research. I heard several comments along the lines of "It's easier to be a basic researcher because you can specialize in one thing and one type of research but with evaluation you have to know a lot about a lot of things." At once, participants pointed out the demand for flexible generalists and they also underscored the importance of being a subject-specific expert. Katherine, for instance, said evaluators should "Get training in some substantive area, because it really does help." Emma commented:

> A lot of firms just kind of evaluate whatever, and so they don't always have a lot that they can bring to the table in terms of, "We know a lot about this topic," or "We know a lot about this policy area or this kind of program." So that they can really inform their funders. And I think funders are really looking for contractors who specialize, too.

An apparent contradiction emerges between the logic of these statements and those pertaining to being a generalist. The underlying message: To be an evaluator, you must be able to do what everyone else does. To be an evaluator with promise, you must possess some skill or knowledge set that no one else possesses.

The contradistinction with academia, then, is not exactly with academia, but with academic forms of and purposes for conducting research. It points to the politics

embedded in all social research and to the degree and nature of intellectual freedom to which a researcher has access. Evaluation, as a genre, relies on maintenance of proximity to the "other" for legitimacy, rapport, and relevance to a bigger context. Moreover, unlike basic research, which, at least in theory, provides a modicum of intellectual freedom, program evaluation is conducted as propriety intellectual property under signed contracts. This relation between client and consumer, especially in its contemporary global economic context, means that knowledge is produced in a process of exchange. Thus, data and process of the development of that intellectual property is commoditized.

These forms of boundary maintenance have both explicit and subtle effects. While there is a practical side to quality assurance (ensuring that only members of the figured world may perform evaluation work), of particular importance for this discussion is the prospect of institutional effects on which cross-sections of socially positioned knowledge workers make their way into evaluation.

In the following section, I explore some of the interconnected themes in an attempt to reframe my précis on knowledge work as the space and source of discursive tools that may be used in the identity production process. Specifically, I describe, in brief, implications that extend beyond each of the compartments explored in this chapter.

Concluding Thoughts

Power and access to the realm of pragmatic ERD knowledge production may also be apparent through social scientific disciplinary struggles. Throughout the chapter, I have identified points of contention between evaluators of differing disciplinary roots and different levels of entry into the figured world. Collectively, these examples reveal what

might appear as a struggle for ideological control of what is said and in what ways it is said in the course of evaluation work. From project conceptualization to reporting, multiple contestations co-occur. But, perhaps what appears as multiple hegemonic struggles may be, instead, or in addition to, an indicator of the level of an evaluator's cultural production as an impassioned member of the evaluation community. The canons of big social science seldom breach the inner world of evaluation as does that of the client. Resistance to the group norms may be, at once, a holding on to sacred ground and source of useful ontological orientation. It is both agentic expression and the result of succumbing to the perceived constraints of a naturalized figured world and with reference to its alternatives.

Indeed, a common thread throughout this chapter has revolved around the balance of what appears to be agentic action and acquiescence to the compulsory aspects of evaluation work. Based on what they said in their identity narratives, it appears that these participating evaluators learned to identify the spaces in which they have voice. For instance, during the proposal-writing process, particular methodological shifts might be negotiated successfully with clients. In this case, the evaluator, as soft activist, might, as a result, have a sense of bringing about institutional change. But, at the same time, the identity narratives make clear that some aspects of the work, such as the focal points of the program evaluation, are non-negotiable. Helen's comment "Great, now I get to hate myself all over again" is one, of many, that make me think of Taussig's discussion on learning to be silent (1999). Evaluators learn to be silent about the non-negotiables, although, as the stories note, they are deeply known tidbits.

In the identity production process, conditioning qualities enforce this learning to be silent at the appropriate times and with reference to the appropriate topics and adherence to the tenets of the ERD. The everyday cultural struggles of program evaluation, such as methodological and discipline-centered vies, which are eventually assuaged by *learning to bend*, permit an effective sidestepping of the bigger, more enduring problem of reproduction. The micro-drama of these everyday struggles mask the prevalence and depth of privatized neoliberal knowledge production.

It is difficult to distinguish agentic silence from passive acquiescence or retreat. I argue that, in the process of re-authoring from social scientist to program evaluator, individuals reconceptualize the notion of agency. It is through the very strategy that makes doing the work of program evaluation possible, *de-stancing*, that evaluators come to accept, and perhaps even protect, the spaces of silence. Avoiding strong theoretical stances may, ironically, open cracks in which the neoliberal political economic mechanisms are able to seep into the entire knowledge production process. In fact, there was little evidence that evaluators could produce, at least in the final reporting, much more than the expected (read contracted). I will return to this topic in the concluding chapter. What I wish to emphasize here is the likely relationship between identity production and the macro-political facet of the ERD.

With the roots of agentic expression in mind, in the following chapter, I take a closer look at the cultural production of program evaluators in education with reference to formative years of the participants of this study. More specifically, I attempt to connect self-authored elements of class consciousness, desire to bring about social

change, a seemingly-natural affinity for data use, and a sense of self-efficacy and voice in doing this work.

Chapter 5: The Naturalized Path to Program Evaluation

In this chapter, I consider narrativized snippets of self-understanding that naturalize the transition from social scientist to program evaluator. According to the self stories about the journey to proficiency in the figured world of program evaluation, as these participants ventured along various career paths that would minimize threats to their principles and beliefs, program evaluation was scripted as a natural fit. This figured world offered the appearance of a merging of identifying patterned human behavior and planned, meaningful social change. Sarah described her entry into a psychology doctoral program that she believed would help her work in the "real world" only to find that it was less applied than she had hoped:

> I really did want to apply the work because I felt it was important work we were finding about human nature and how to make the world better. And, so I was actually somewhat disappointed when I got to the program because I thought it was going to be more applied. There was still a lot of basic research going on, which is important, but I was wishing there was more of a focus of taking it from the ivory tower into the world.

As I discussed in Chapter 4, she left the program only partly primed for evaluation. The cumulative belief about the role of social science in social change had produced a somewhat clear notion about that change that would survive not only graduate school, but career experiences, as well. For instance, Emma commented about her notion of change through research:

> I've worked for the Texas Elementary Principals and Supervisors Association, and that was the time of my life when I was introduced into the field of education and decided to go into that route rather than going, to graduate school for something like developmental psychology. I saw how useful good research and good evaluation really could be, especially in the policy-making world.

The belief that applied research could contribute to change was shared widely among participants. Mark, who is a founding member of the American Evaluation Association and who began his career as a clinical psychologist, said:

> I had an advocacy background and I'm also a child of the 60s so one of the things that I believed was that clients shouldn't be exposed to programs that don't work or that are not guided by information-based decision making. At the same time, I realized that nobody gets it right the first time out. So program staff need the kind of information that they can use to understand what's working and not working and improve it.

Therefore, these participants felt evaluation was important work. The calling, the personal fit with program evaluation, went beyond program improvement to deeply held beliefs about its liberating potential. Sarah, for instance, said, "It is great, important work. It has always been about giving a broader, wider voice to those who don't necessarily get represented. I would say it is a great path." This proclivity was expanded by and, at the same time, deepened by the "fit" with the labor needs of the evaluation industry that is oriented around flexible specialization.

Before drilling into the personal histories of participating evaluators with reference to their sense of purpose and how this purpose interweaves with personal narrative, I will refine my theoretical premises. As I demonstrated in the last chapter, the principal artifact and the target of daily toil is the deliverable. This knowledge production is a cultural-political act that takes place in an ever-shifting figured world. As a cultural-political act, the identity of the producers of knowledge is continually forged through everyday cultural performances of work, against the backdrops of tenets of the widely-broadcast ERD, and filtered through the individual and collective lenses of local history and social position. Evaluation knowledge is produced in an entanglement of structures, artifacts, and personal positioning. In this entanglement, this knowledge becomes a naturalized, reified carrier of cultural messages, including those pertaining to the ERD. Thus, I look beyond scientific questions to multilevel, sometimes hidden, and sometimes contradictory, perspectives that identity and position are not foretold, but, instead, are negotiated, constructed, and reconstructed.

Contemporary work on identity production is in line with the Foucauldian contention that identity is neither coherent nor a terminable product (Holland et al 1998; Sfard & Prusak 2005). Instead, people are "composites of many, often contradictory, self-understandings and identities, whose loci are often not confined to the body but 'spread over the material and social environment,' and few of which are completely durable" (Holland et al 1998, 8). Identity is also the "process by which social actors recognize themselves—and are recognized by other social actors—as part of broader groupings" (Della Porta & Viani 2006, 91). It gives a sense of boundary and purpose

and, thereby, greatly influences what a person does in everyday life. In order to appreciate identity it in its rich complexity, identity research addresses social position, power, agency and improvisation, the interplay between internal processing and interactions with the physical and symbolic world, and the ways people are situated historically and in culturally constructed roles and activities. Before using this lens to focus on evaluators' identity production, I attempt to describe how I conceptualized facets of this framework.

Discourses are not only limits and boundaries, but are also cultural tools. These discursive implements allow cultural actors to develop a self-understanding about and for themselves. Identity is produced through the *mediation* of powerful discourses; it is not, as Foucaultian constructivists would argue, the direct outcome of Discourse. Drawing from Bakhtin, this mediation occurs at the crossroads of social and internal "dialogue" (1981). And, drawing on the work of Voloshinov (1929) and Leont'ev (1977), identity production fixes itself on the ways in which external objects and ideas become internalized and how internalizations get expressed. Holland et al state, "People tell others who they are, but even more important, they tell themselves and then try to act as though they are who they say they are" (1998, 3).

In a sense, there are two selves: one that reflects and one that acts (Holland et al 1998). A dialogic relation mediates the two. At once, self-understandings are intermittently "figured" in improvised self-communication about the past and present. And, at the same time, in order to be recognized or known, a person must be categorized;

a process that reduces, essentializes, reifies and gives coherence (Wortham 2003). Thus, identity is, at once, both a complex, ongoing process and a narrativized snapshot.

This brings us to the theoretical point I wish to emphasize most. Self-authoring is the orchestration of internal and external voices, which include, among others, key, influential community voices, or what Sfard and Prusak (2005) refer to as *significant narrators* who transmit important "cultural messages" (i.e., Discourse). Evaluation heteroglossia draws on a diverse repertoire of methodological tools and idealizations. There is the ERD and the, ironically, different language of equity—not as crude discourses but internalized, living and embellished "recordings" of influential figures. In such an orchestration of voices, different narratives assert unbalanced degrees of authority. For instance, during a data analysis phase of a project, two methods experts—former professors who had different ways of conceptualizing thematic analysis—might vie for hegemonic primacy while drowning out the whisper of the former CEO who wanted it to sound good so we would be sure to win the next contract. But, beyond the internalized voices in my head, I also react to a continual struggle between improvisation and perceived obligation.

While actors' improvisations constantly alter the conceptual and material landscapes of figured worlds, which opens the space for change, discourses are imposed on people "through recurrent institutional treatments and within interaction, to the point that they become self-administered" (Holland et al 1998, 62). In other words, those discourses, themselves, become internalized and help to normalize self talk. The processes of evaluation research and the production of artifacts, such as evaluation

reports, have the potential to contribute to the inscription of discourses and impositions of discipline while they may also increase self-control and action (through innovation and improvisation). Program evaluators' identity is continually produced with constant reference to these powerful discourses and perceived struggles.

Since perspectives on cultural worlds differ by social position, cultural study must include historically and socially situated forms and the "processes through which they are negotiated, resisted, institutionalized, and internalized" (Holland et al 1998, 26). Overlooking social position is tantamount to silencing (Trouillot 1995; Shumar 1997). This is of particular import here because the real and felt ability to improvise, to assert oneself, to speak with authority, can be hampered or facilitated by social position, historical dialogues within communities, and a sense of proficiency. People position themselves and are positioned in relation to a series of discourses and social differences (Dole & Csordas 2003; Lachicotte 2002), which results in acts and actors assuming and being assigned to rank and status within structures of power in a relational hierarchy. An internalized social position may guide conduct, which in turn, may reinforce or weaken a sense of belonging (Della Porta & Viani 2006) to a particular group. This, of course, co-occurs with other factors including, as I illustrated in chapter 4, agreement with the ethos of everyday work.

This has important implications for identity change. That sense of belonging holds the potential for identity transformation. While change for Bourdieu comes about through a change in structure rather than achievement (de Certeau 1984), improvisation vis-à-vis subject positionings holds the potential for changing identity (Holland et al

1998). If the social and material conditions change, "old 'answers' about who one is may be undone" (Holland et al 1998, 189). In other words, as the material and discursive structures shift, so, too do the social positions or level of authority of internalized voices and, therefore, understandings brought to and interpreted from them (Sfard & Prusak 2005). Even institutional narratives, such as policies associated with the ERD, have the potential to alter identity narratives.

In this chapter, I used this emergent framework to explore personal histories of evaluators to identify each participants' cast of significant narrators, the discursive frames that resonated (or were so natural that they went largely unnoticed), and the ways in which power, social position and symbolic capital interacted to contribute to self-scripting stories. The chapter begins with by peeling back layers of self-understanding that underscore the calling to program evaluation and then moves into an analysis that places oral histories into formative class conditions. This analysis leads to the final section, which includes a set of ideal types of program evaluators that fits neatly with the social class extremes.

The Formative Production of the Contract Evaluator

In this section, I offer an interpretive, cross-case comparative description of the formative years of the evaluators who participated in this study based on patterns in social position, socialization, aspirations, and educational pathways. Following the understanding that emergent versions of self-group understandings would greatly influence later versions, I discuss Bourdieu's critical analysis on class reproduction as cultural process (1977) with reference to these data. While I did not begin this study with

a class-oriented typology, I found Bourdieu's *habitus* construct a useful tool for beginning to understand information pertaining to upbringing, recollections of important activities and characterizations of self during youth, descriptions of values, and depictions of families and communities. As I pored over these personal histories, it became clear to me that class conditions, as an indicator of social position, would help explain perceived and experienced academic opportunities and pathways, limitations and choices that preceded becoming a social scientist, and budding aspirations to become a contract program evaluator.

Most of the data used in this chapter were generated through the first interview. In that interview, I posed to all participants a series of semi-structured questions including:

> Tell me about your childhood. Give me a broad description of yourself from your earliest recollections through grade school. Also tell me about things that really captured your interest, especially bothered you, and any skills or talents you recall.
>
> What activities were especially important for your family? What values or everyday practices were most important to your family? Please provide a few examples. What was most important to you in terms of values and activities? Tell me about high school days.

> Tell me about your postsecondary education. Tell me about your college selection process and how you chose the academic path you took. At the time, what did you hope to get out of your college experiences?
>
> Tell me about graduate school. Why did you choose to study what you did? What did you hope to get out of your university experience? Compare yourself to others in your cohort or program.
>
> Tell me about the path to your current position in relation to graduate school. What inspired you to consider program evaluation? In what ways does it relate to your graduate degree?

Some of the more salient themes to which I would like to draw attention are: (1) that many evaluators were the first generation in their families to go to college; (2) that all endeavored to do work that was socially meaningful, if not activist; and (3) that, as mentioned in the opening section of this dissertation, few of them went to college with the intention of becoming program evaluators. It is to this third theme that I would like to turn first with reference to the other two themes.

The iterative, multi-phasic process of becoming a contract evaluator involves an ontological transitioning from a deeply-felt need to do good to becoming an adept social science researcher to becoming a proficient contract evaluator. The first step in the process involves playing on the fringes of program evaluation. Drawing on Vygotsky's work on play, Holland et al (1998) contend that playful imaginings allow for the emergence of new figured worlds, "or refigured worlds that come eventually to reshape

selves and lives in all seriousness" (236). They contend that discourses, such as the ERD, are "disseminated within." As they are internalized, their practices are habituated. While habituation creates a naturalized set of thoughts and actions (Sullivan 2006), which closes off our ability to envision alternative figured worlds, play also permits figurations to move closer to consciousness. This relationship between habituation and play is dialogic.

There is no unproblematic and quick shift. Bourdieu sought to move beyond "the realism of the structure, which hypostatizes systems of objective relations by converting them into totalities already constituted outside of individual history and group history" (1977, 72) by adding a sort of group agency or *habitus*, as well as individual agency to his understanding of change. The concept of habitus would, purportedly, resolve the inconsistency of the human sciences: objectifying the subjective. This system of dispositions—an enduring learned scheme of perception, thought and action—is a form of semi-political economic determinism mediated by agency. In other words, dispositions in relation to the objective conditions are reproduced in the individual. But, the objective conditions cannot predict the habitus. Habitus is detectable only by relating the objective structure; it is "history turned into nature." This is passed on through a homogenization of agents' experiences, which makes self history, group history, material condition, and commonsense appear coherent. This homogenization of class enables the harmonization of practices.

Thus, habitus, which is a product of both history and "objective conditions," purportedly drives, if not compels, individual and collective practices, and hence reproduces localized history. If the habitus is reproduced largely intact, how does

Bourdieu account for change? Is it possible to move reflexively outside of the clutches of habitus prior to changing the material conditions? Or is a materialist intervention required? For Bourdieu, as long as there is stasis within a system and no specialization, the modus operandi goes unnoticed and, therefore, continues to operate as is. This presents a considerable challenge for an identity framework that posits identity making (albeit eclectically) somewhere in the middle of compulsory, economically-determined practice, individual and local group agency, and multiple dialogic interplays. This is the space in which "history in person" or narrativized sediment of past experiences—including those experienced first hand and those offered by significant narrators—brought to bear on the present.

A question I attempt to answer in this chapter is: *What inspired these people to become so passionate about and come to describe themselves and be described by others as Program Evaluators?* The self stories I heard competed, in some ways, with Bourdieu's work on habitus, while, in other ways, they bolstered his understanding of social position. One theme I would like to emphasize in this chapter is that while class conditions, writ as habitus, produce class-specific aspirations and dispositions, changes in discourse, material conditions, and individual status also shift the amplitude of internalized voices and, thereby, open new spaces for agency. From a distance, the life stories were similar and similarly scripted through a mediation of the ERD, previous self-understandings, and new opportunities. When social class was put closer to the forefront of the analysis, however, a few distinct patterns emerged. Before getting to these, I

describe the seemingly natural proclivity to be a social scientist that was ubiquitous among the participants of this study.

The Natural(ized) Applied Researcher: Social Sciencing the Self

Although I did not ask about voluntary service, the aspirations to apply social research to addressing "real world" problems and to contribute to social change became a dominating theme related to values and guiding principles, especially in the context of participants' descriptions of their socialization and their characterizations of themselves from early childhood through early adulthood. These core beliefs provided some of the connective tissue at transitional moments such as decisions about educational and occupational pathways to pursue. According to the characterizations of youthful versions of themselves, while few had clearly defined academic or career plans, they did have a deep-rooted drive to expose inequity, a felt need to speak for and protect the "underdog," and a seemingly natural proclivity for using social scientific research methods to contribute to social change. These deep-rooted drives inspired and directed decisions, according to participants, related to higher education majors that would enable them to contribute to social change.

I begin this discussion with a peek into the process of becoming social scientists. Participants' descriptions of this process occasionally contained the lexicon of social science. At first glance, terms like *social capital*, *projecting*, and *social networking* seemed like curious additions to reflections of identity from the distant past. But, those terms may serve as reminders that the distant past is a production of a distanced present. They also indicate the naturalized state of being a social scientist, since people define

themselves with the linguistic "tools" that are both available and personally resonant to them (Holland, et al 1998). To the extent that self stories are neat reconstructions that mediate memory, power, and a partial understanding of current position, it is only the me of now who can comprehend and describe a previous version of me. In doing so, my hermeneutic rendering of formative iterations of myself may seem to lead up to and, therefore, justify my current self. In other words, in the telling of my self story, contradictions and alternative interpretations are ironed out of the narrative, leaving the impression that the researcher I am now is the result of a sort of retroductive teleology.

I am not the first person to attempt to understand the identity of the people who participated in this study. While students pursue degrees in the social sciences for a variety of reasons in addition to working to bring about social change, among those is an attempt to understand the world, oneself, and the ways in which the two are interrelated. For most participants, the decision to become a social scientist involved more than an offhand selection from an array of academic choices; it seemed to be the *only* fit. While identity stories blur the line between the "naturally curious" adolescent and the seasoned social scientist, participants made this connection explicit. For instance, Beatrice, who is the director of an academia-based program evaluation center that relies on large-scale grants, said that her interest in "understanding stratification and issues related to equity" was present "from the beginning of time. That was very early in my childhood." Similarly, Sarah reflected:

> I always had a real interest in social dynamics; the way people interact and the non-verbal communication that people put out. And, that always led me to

> psychology. I always had a real interest in finding out what made people tick. I always knew I was going in that direction. I was born to be a psychologist.

Bonnie reflected that she had "always been an observer. I'm a curious person. And, so research fit really well because I was curious about people and fascinated by people without actually having to help them in any way." Mark said that when he was in high school, "I'd get on a bus and sit down with somebody and I might ask 100 questions," as an example of how doing research came to him so naturally. In each of these cases, the seeds for becoming a social science researcher had been present from childhood or "always," it was natural, and it was meant to be. Becoming a researcher was just short of a destiny. Emma situated her self-understanding into the realm of ways of knowing by situating the naturalized researcher thread into her epistemological outlook:

> I have a natural affinity towards making decisions in a more systematic way instead of how some people know things in a more emotional or spiritual way. I think that you combine that with natural curiosity and you've got somebody who is primed to be a researcher.

Anne perceived her penchant for research as residing at the core of her being. She described her fascination with anthropological ways of knowing as "expressive of my spirituality." In the logic of these recollective accounts, there was little room for

consideration of other pursuits.[17] Building on this, some participants, who cognizantly connected the attraction to social science research to their upbringing, saw social science inquiry as a potential solution to life mystery questions. Helen, for instance, described her upbringing in a family of "hippies" who lived a utopian lifestyle and who greatly valued education as a contributor to her becoming an "oddball." This sense of being different led her to seek out the social sciences in a sort of spiritual way:

> I guess the first big, and maybe most formative, lesson of my life was that I didn't have the background, experiences, and perspectives that other kids in my classes seemed to have. I couldn't relate to their conversations about TV shows; I didn't eat the same kind of food; and I came to understand that my family's disbelief in god was an enormous no-no. I felt that everything about me was different from, and more, against, everyone else. It was a huge struggle to incorporate that understanding into my sense of self in a constructive way. I went through all kinds of approaches—hating myself, hating everyone else, trying to convince everyone else that I wasn't a bad person, trying to convince people of my viewpoints, and acting out.

[17] While in this frame, participants described a very "closed" account. Yet, in later discussions, some talked about being primed for a wide array of possible majors if they could have helped with doing socially meaningful work. This does not necessarily reflect an identity narrative teleology; some speculated that had they found another way to make a difference, they might have followed other career paths. For instance, Katherine, who is a professor in a program evaluation graduate program and who works on contract evaluations at the same time, noted, "If I was an undergraduate student and you told me I could have done all that as a writer, then I would have said I wanted to be a writer." Of course, this contradicts the internal "logic" that seemed ubiquitously offered by PEs—that they were destined to be evaluators.

Helen's self characterization of being out of place was not unique. Almost all participants said that during their childhood and early adulthood they felt socially awkward and shy. Bonnie commented that she "felt like a fish out of water. I mean I fit in terms of I could succeed and that was very important to my family, but socially I was awkward. I was shy." What I would like to point out here is the connection of this professed social distance and curiosity to the natural fit with research. In Helen's case, social science theory and content offered a solution to her "identity crisis." She felt that "Sociology gave me all these words and ideas that made me think about all the weird things in my life around me." She elaborated:

> One of the things we learned really quickly in our county was that while we might have all kinds of cultural capital, it didn't mean shit because we didn't have the social capital, the relationships, the networks. Other people could say, "My family has been around for 100 years and we have all this land and my family knows your family." We didn't have any of that and it [sociology] helped me put things in broader perspective.

Later in the interview, she reflected:

> It took a very long time, but eventually those experiences led me to see myself as an observer, an outsider, on the margins. This was both enabling and constraining (nod to Anthony Giddens)—enabling because it ultimately led me to social science and constraining because I lived a lot of my early adulthood not participating in my life so much as observing it.

Helen came to understand her childhood experiences through the lens of sociology. Although she mentioned observing more than participating, she, and other participants, used their social science understanding, mediated by their sense of the reform discourse's value in promoting equity to a sense that they needed to participate in social change. As I will describe in more detail later, this constellation of activity and belief, along with the challenge of being a single mother of two, encouraged Helen to seek out a career that would pay well, make ample use of her skills and interests in social research, and hold promise for equity-minded social change. It is this lattermost point that I turn to in the next section.

Activating the Educational Reform Discourse

Common across participants' personal histories was a description of their active involvement in addressing social issues, most especially those focused on disrupting or placating injustice that targeted marginalized populations. Depictions included involvement with social service and volunteer work intended to assist HIV positive, homeless, mentally ill, recent immigrant, gay and lesbian, economically disadvantaged, pregnant teen, battered women, and other marginalized populations.

In this section, I consider the ways in which participants depict the conception of their interest in and were drawn to social issues. This is not meant to serve as a quasi objectivist analysis of how the Educational Reform Discourse becomes enacted and naturalized, but, instead, as a filter to capture traces of evaluators' core self-understanding. It is upon this core self-understanding, which resides at the ontological level, that subsequent experiences are interpreted, assessed and treated. The core self-

understanding is the result of layers of sedimentation that have settled and hardened at the dialogic base of self-talk, everyday practice, and personal history. The "matter" that settles is derived from these internalized voices, the ERD (among other competing discourses), socialized understandings of place and propriety. In other words, it is commonsense.

The predisposition to be a social scientist is presented as if it was "always there." It is a natural curiosity in knowing "what makes people tick." With a little elaboration, however, participants pinpointed the experiences that helped shape this perspective. Beatrice, for instance, added to her depiction of her natural interest in equity that she had witnessed how students who came from less affluent families were not treated the same in school as were students from the community in which she resided. Isabella, the "hired gun" evaluator on the side, said that her liberal leanings were fueled into action when she was a teacher in inner city Washington, DC. She said of her reflections on the experience, "It was out of dumb idealism where I mean, I hate it, but you think, 'Why are things so bad in DC? What could be done about this? What's James going to do when he graduates?'"

Embedded in the accounts of social action were speculative traces of the story about how social conditions and early socialization undergirded the production of participants' core values related to equity. The accounts depicted how they became passionate about social engagement, which, in turn, influenced their decisions to attempt to solve problems related to inequity. A few participants connected childhood social conditions to the impetus to bring about change. James, for instance, a pioneer of

contemporary evaluation, grew up during the Depression "on the wrong side of the tracks." He told me that because of this, he was placed in a vocational track with other boys from his neighborhood. During summer school, an interim administrator noted that his academic scores were higher than those of most of his peers and he was placed in the academic track and, with it, on the road to college. Once he became aware of these trajectories that had operated invisibly, he recollected sensing a need to "do something socially engaging."

Nick, who grew up with modest means in a Southeastern city in the 1970s, recalled derogatory and racist comments some of his teachers made. He recalled feeling a deeply-resonant symbolic violence directed at African Americans living in poverty. Beth, the CEO of a contract research and evaluation firm, tied her "sense of social responsibility" and belief that it was "really important to do something in the world that helps other people" to the poverty that surrounded her. She noted, "I really believed that things could be better for other kids than maybe they were for me." She tied her poverty conditions to a personal dimension:

> I grew up in an alcoholic home. And, so I think there was a sense that there is something wrong with dad and it makes me really really sad and I wish I didn't have to feel this way and maybe there is something that can happen so that I can make other people feel better. And, I think it was also a growing connection to what I would now call God. I really didn't want people going to hell, Keith. I thought maybe I could do something.

These participants felt that socio-economic circumstances helped create a tension that incited an enduring concentration on social justice in which they would help prevent others from sharing those experiences. As Beth's example makes clear, their experiences had the potential to "save" people.

While it might be implicit with the stories so far, others described family as the primary source of the ERD. Emma, for example, who worked, briefly, in a non-profit evaluation consulting firm and then as a state department of education associate who was responsible for selecting contractors to perform education reform initiative program evaluations for that state, believed that her liberal family, who often discussed issues openly, was the foundation of her leanings. She explained:

> I remember in my elementary class going, 'ERA, all the way!' My mother's always been kind of interested in politics. And, I had an uncle who's a psychology professor and whenever they would get together, they would get into these long conversations and I would kind of sit and listen and it was just really interesting. So, I think I developed that interest in politics and policy. I think that's one of the things that led me toward being interested in research because I thought, 'This is a way to get good policy and good decisions in play.'

Whether based on direct experience or socializing family conversation, these pinpointed sources that epitomize personally resonant convictions about needing to make a difference became sedimented self-understanding matter upon which a more elaborate set of principles, actions, and trajectories were developed and enacted. With enough

practice and habituation, they form an ontological-axiological foundation. The convictions revolved around such notions as advocating on the behalf of those without voice, exposing the truth, volunteerism, and activism to make a difference. Participation in knowledge work without an aim to bring about change was described as "senseless," "tedious," lacking contribution. This theme, therefore, underscores a tension between basic and what some social science practitioners might call "pragmatic inquiry."

Participating evaluators' descriptions of the kinds of social change activities in which they were involved during young adulthood was distributed on a range of conceptualizations about social change and social engagement. To a degree, and as I clarify later in this chapter, the approaches to change are linked to the kinds of career choices evaluators made, including, in some cases, the choice to steer back out of evaluation after experimenting with it. A handful of participants described working on school newspapers and otherwise brokering information with the explicit purpose of "exposing the truth." Sarah, for instance, who worked as a college newspaper columnist, wrote about hypocrisy. She told me that as she did this work, she avoided facts that made the stories less extreme, a confession that later solidified her decisions to do contract program evaluations.

A subset of participants described activity in leading change as campus activists. Bonnie, for instance, who is the director of the evaluation department at a not-for-profit contract research organization, was the founding member of a local chapter of an Abortion Rights Action League. She recalled how, at the time, she believed "If you didn't have issues that you cared deeply about and did something toward living your

values, then you were just a hypocrite." Helen, who worked for a similar company and in a similar capacity to that of Bonnie and who recently moved on to a Washington-DC based for-profit firm, was raised by activist parents. She noted:

> I got involved in two liberal campus activist groups. One was a sort of catch-all social justice group. The other was the campus gay and lesbian rights group, which I helped start up with a friend. He wanted a lesbian co-president, but no one volunteered, so I said "sure, why not?" It was embarrassing and uncomfortable for me because I wasn't a lesbian and wanted everyone to know it, yet there I was proclaiming how important it was to be open minded and supportive of gay people.

This later point highlights a theme in this analysis—a realization that certain facts or perspectives were often overlooked or contradicted their work as activist. Camila, who worked for a non-profit evaluation company less than two years before deciding that evaluation was not the career path that best suited her, did a lot of volunteer work in college, including tutoring children and ESL learners and working for a domestic violence shelter. She commented that while the experiences seemed to help some people, the system could only do so much. She noted:

> I learned a lot about how complicated these problems can be. It gave me a lot of skepticism about these non-profits' ability to help people in the context of some larger social safety net. It was an education for me.

This experience underscored for her a need to understand issues with facts and stories. As an undergraduate, she went to an Ivy League college that "had a lot of very passionately engaged idealistic people who wanted to change the world and were very into organizing protests and living the activist lifestyle and taking very extreme stances on issues." Her skepticism was defined by her sense that the polemic stances that she saw activists taking forced them to sidestep important details and facts. She said:

> I cared a lot about social policy and I wanted to learn to make some kind of positive contribution to the world in some way that I didn't really figure out. But I didn't want to be an activist. I didn't want to be one of those people who is just yelling about stuff all the time. I did care a lot about the world and about domestic social policy issues like health care and education.

Social action, therefore, complicated seemingly simple answers to real world problems related to equity. It also summoned a felt need for systematic research to better understand the problems and how best to produce and gauge the effect of policies and programs. Socially active experiences also brought to life the consequences of bad programming and policy. The retrospective self-logic—that is, the superimposition of culturally produced and, therefore, seemingly natural responses to social inequities with a story of how they became the social scientists they are today—is made possible. The logic or flow of these self stories, taken as a group, provides insight into both the rules/obligations of the work of the figured world and the ritual practices associated with these rules/obligations. This gives credibility to the notion that discourses are inscribed,

unreflectively, in institutional codes (Foucault 1977). Together, they appear as institutional logic—the package of practices, thoughts, and code that operate from the ERD.

In summary, early adulthood experiences with organizations intended to work with underserved and marginalized populations were eye opening to participants. They discovered that problem solving in the face of inequity was more complex than is seemed at first glance and that organizations did not have all the answers and, in some cases organizations presented their own barriers to services or perpetuated shortcomings. As these participants began their journey into social sciences, the measure of impact and the logic of monitoring seemed to compensate for much of what was lacking in organizations touted as equity-minded. Soon after embarking, the personal calling became one of linking measurement to social justice.

Social Position, Socialization, and Educational Pathways

Before continuing the story of how these evaluators became activated within the ERD, I take a detour into accounts of the class conditions they experienced during their formative years. While participants felt compelled to do something they considered to be socially meaningful, their identity narratives, which included a description of how they conceived of that meaningfulness, made common reference to childhood experiences. Although I did not invite participants according to indicators of their socioeconomic status or social capital, childhood *class conditions* intersected with educational pathways, professional aspirations, and, not surprisingly, therefore, their selection of professional titles that index how they conceive of their relations to the everyday activity of program

evaluation. In fact, I presupposed before beginning this project that evaluators came from pretty firmly rooted middle class families and I did not, therefore, anticipate the range of experiences, much less how neatly class would be predictive[18] of program evaluation type and status. I use *class conditions* as a simplistic composite category consisting of participants' recollections about their upbringing, parents' education and occupation, depictions of the community, higher education expectations and planning, and access to and use of tangible and symbolic resources that influenced decisions about higher education.

I organize the class conditions discussions around an exploratory interplay between what participants said they had wanted to do professionally and what they had chosen to study in college. After drawing out descriptions of career plans (i.e., those for which participants set their sights and with which they experimented when entering college) and their continuing education in graduate school, I indexed reflections about college and major selection and whether or not higher education, in general, and the specific academic pathway followed was a planned event. In addition, I indexed the conditions, perceived opportunities, and significant narrators of internalized voices that influenced higher education and career decisions, including scholarships and targeted recruiting efforts, advice from faculty and significant others, research experiences, materialistic constraints (e.g., proximity from family), and just plain serendipity.

[18] I frame these words loosely, nearly glibly, since, I must caution that these self-stories are self-contained and give the appearance of direct paths. Using a word such as "predictive" while looking back is probably a gross misapplication in a positivist research sense, but, here, I use it to convey the sense of conviction in the voices of program evaluators as they proclaimed their professional status.

In this section, I focus on these personal trajectories not to suggest that these people unproblematically adapted to their circumstances, but to highlight an aspect of the dialogic interplay between improvisation and compulsory action, which, along with socialized expectations and values to do socially meaningful work, provide durable filters for accumulating prospective professional paths. The accumulated stories of participants led me to develop and assign participants to four categories—Professional knowledge workers (n=5), Information servicers (n=4), Information borrowers (n=6), and Knowledge "drifters" (n=5)—which I detail below. At the end of this section, I reconfigure some of these categories in a meta-thematic analysis that reveals two types of program evaluator—the Post-Academic Intellectuals and Academic Practitioners, both of which, at first glance, appear somewhat out of place. But, upon closer inspection, both show signs of innovative adaptation to the political-economic conditions I described in Chapter 2.

Professional Knowledge Workers

Five participants were brought up in households in which their parents completed graduate school and became doctors, lawyers, and tenured university faculty. As expected, this influenced education planning and the everyday treatment of formal learning. Helen, whose parents were social science professors, reflected:

> My family cared a lot about ideas and thinking, but were also Romantics in the sense that they believed in love and human connection. So, they weren't cold and remote thinkers and we got the sense that the reason for thinking and reading and

> learning was for some greater good, for other people, rather than for the exercise itself.

There was never a question about whether or not these participants would go to college. Plans for higher education had been set into motion early on. Furthermore, this set of situated professional families had access to assets and social networks that would make those plans tenable. These future evaluators attended selective liberal arts colleges such as Brown, Wesley, Colby, and Antioch. Most said they went into college thinking they would become teachers or counselors so they could do something "socially meaningful." Camila, for instance, noted, "I wanted to be a teacher. I wanted to be more of a help to people in a very individual way." They discovered and then switched to social sciences as undergraduates. And, as noted in the previous section, they decided to study social sciences as both a way to understand themselves and to serve a greater good. Anne said:

> I had no idea what I wanted to do except that I wanted it to be politically relevant and contributing to meaningful social change but on a small scale, so small scale is a big issue for me. Well, because of my whole philosophy about how change happens.

These participants received ongoing guidance from their immediate families as they transitioned to social sciences. Helen, for instance, recalled a conversation with her parents in which she announced to them that she had elected to study social work instead of sociology because it did not have a statistics requirement. She said, "I can remember

talking to my folks about it and them just rolling their eyes and saying 'Give me a break. It might be hard but you can do this. You got through algebra.'" As the other class condition categories show, this kind of encouragement was not ubiquitous.

These social scientists noted the closeness of their families and the importance of maintaining deep, enduring friendships with others. Quite a difference from those in other social class conditions categories, these evaluators grew up in relatively stable community settings; their families did not move around and most still reside in those communities. While none recognized him- or herself as wealthy or upper middle class, they did have, as might be imagined, considerable accumulation of symbolic and tangible assets. Helen, for instance, recalls:

> I didn't go to high school. I finished 9th grade and then started as a freshman at [a private university] where my mom was then teaching. I'd been taking college courses in the summers starting after my 7th grade year, so I kind of knew what to expect. After 9th grade, I decided, with my parents' encouragement, that I'd had enough and wanted to go to college full time. Mom helped [the private university] establish an early entry program.

Beatrice noted that she survived graduate school because of friends she had made, including her husband and a few key people who worked in academic departments:

> When I began my doctoral work [in the early 1970s], I am not sure there was even a push to admit women at that time. I don't think that had happened yet. But, I had a really great advocate. She took an openly feminist stance in supporting me.

Advanced degrees were also expected in these families. None of the participants experienced difficulty funding advanced degrees or convincing their parents about the purpose for going after degrees in sociology, anthropology, or psychology. And, as with undergraduate education, families and significant others had tremendous influence on their decision making about graduate school. Furthermore, as I described in the previous section, the decision to pursue advanced degrees in these disciplines was also connected to convictions about the potential usefulness of social science in solving real world problems.

Furthermore, while considerations for graduate school were less exploratory than were undergraduate education pathways, graduate school plans were defined iteratively with career plans and research experiences. While most in this category began graduate school with the intention of becoming university faculty, these participants wanted to be generalist social science researchers. In fact, two who had planned on becoming professors steered away from that career path because of the likelihood of needing to focus narrowly or devote their careers to a methodological specialization. Camila, for instance, said:

> I thought I wanted to be a history professor. But, I felt like it was a tough road to travel. If you are going to be a historian, you have to choose your own particular area. And, I wanted to be a generalist and learn about all sorts of things.

Some participants did not want to limit their professional futures. For instance, Paul, who was an evaluator for less than two years before moving into a government policy role, wanted to keep his professional options open. He reflected:

> I started taking classes toward my Ph.D., initially in ed psych but I realized that I really wasn't into the educational landscape in terms of a permanent career, so I opted for industrial organization.

At the time they were accepted into PhD programs, they were professional knowledge workers who had reached the ceiling within their organizations. For instance, four of these participants who worked for contract research organizations as master's-level practitioners, returned to universities for their PhDs because, while they could manage projects, they could not serve as principal investigators. And, because they had started families, none was in a financial position to stop working while pursuing a degree. Few were situated in close proximity to PhD programs they considered ideal and only one felt able to relocate. Because of that, most selected from what was available at local, yet selective, universities or worked on their degrees from afar.

Historical themes from Chapter 2 come into play here. These evaluators, whose liberal parents had been involved in academically-inspired social change and the pursuit of knowledge for the sake of knowledge, experienced a different professional landscape than that of their parents when they were ready to enter the workforce. The two participants who completed their doctoral degrees before 1970 found few solid positions available in academia (see Shumar's College for Sale: A Critique of the Commodification

of Higher Education (1997) for a detailed analysis of this economic transition). With the opening of new markets for social science researchers in the contract world, they found themselves productively employed on long-term contract projects, doing the kind of research, more or less, that they believe they might have been doing in an academic environment. Beatrice, however, a participant who earned her PhD in the latter 1970s, found academic positions to be flourishing and soon became an active researcher who combined grants and contracts in a variety of school reform related evaluations. Anne, the second participants entered academia briefly, but said she felt more secure economically outside of academia. Similarly, those who completed their degrees in the 1990s felt that they could either opt to become transient and insecure part-time faculty or keep their non-academic, soft money research positions that paid significantly more than did most tenured faculty with the same years of service.

Some participants were concerned, almost apologetic, about "selling out" because they had steered away from academic spheres of their respective social science discipline. Interestingly, although not rewarded by their organizations for doing so, several who work in non-academic environments study and contribute to critical literature. They also reflect critically on their work. After an interview, for instance, Helen wrote in an email to me (emphases hers):

> An issue I struggle with has to do with how evaluation is used. I read this book called Seeing Like a State that sharpened my critique of evaluation, and I haven't been able to get the book's insights out of my head. To the extent that evaluation is a tool to help government make complicated local dynamics/programs "legible"

> for monitoring and accountability, evaluation is made even more reductive than it already is, and tends to help the funder more than the program staff, clients, or stakeholders. I worry about being implicated in evaluations that render the pictures of programs flat or stereotyped, or that pave the way for less helpful, more intrusive, or even harmful policies or programs. I'm sure the early evaluations of the [state program] showed increases in power production, for instance, but the [state program] also displaced a bunch of people. I'm not paranoid—I just want to be careful (and as little a tool of the state as possible).

Important here was a tentative acceptance or validation in post-academia and not being affiliated with knowledge production solely to serve the agendas of power groups. Another participant, Alice, who completed her dissertation in 1969 at what she characterizes as having been a radical and activist-oriented campus, described how she transitioned to work inside a federally-funded institution. She was concerned that her doctoral cohort and subsequent classes would see her as a "sell out" for going to work for a for-profit contract research organization. After she started working for the company, earning double the salary of what most beginning faculty earned, she was invited back to her university department to speak to students to find other students interested in what she had to say and wanting to know how they could enter that world. Anne said:

> One of my personal frustrations in the situation we're in, because we have to spend a lot of time generating work, is that I feel like I've been starved of time to read and write and I have to get back into that social theoretical realm, which I got

> a lot of training on as an undergraduate and graduate student that I find really rewarding and meaningful and important. Now I just kind of churn out the evaluations.

While all have worked on program evaluations, none identified him or herself as a *program evaluator*. Instead, their professional identities were more closely aligned with the social science discipline in which they earned their graduate degrees. They were anthropologists, psychologists and sociologists who happened to work on contract evaluations. And, as I discussed in Chapter 4, they maintained the connection to social scientific discipline in the conceptualization and production of knowledge for clients. By and large, they defined themselves as social scientists who were either dislocated (or who chose to dislocate themselves) from academia because of the trends toward commodification described by Shumar, a theme to which I will return later in this chapter.

Information Servicers

Four evaluators-to-be were raised by parents who had, themselves, been raised in middle-class households. Their parents worked in what I am referring to as the knowledge servicing industry in fields such as computer programming and communications. While at least one member of their extended families had gone to college, that person was neither a parent nor member of the nuclear family. Several participants described being raised by single mothers who were told that "women don't go to college" and who were forbade from doing so (one mother earned her college degree later in life). Instead, grandfathers and uncles earned BS degrees in fields such as

engineering and then worked their entire careers in those fields. In at least two of these households, the experience of gender barriers to higher education incited mothers to make sure their daughters became college-bound.

Unlike the more firmly rooted families in the previous category, these families faced insecure jobs, which meant occasional movement from one community to another to follow jobs. Bonnie, for instance, noted, "We moved around a lot. So, I couldn't rely on friends, but I had the adults [her mother and grandparents]." As her example suggests, while these participants did not see themselves as being connected to peers, they did spend more time with adults in their families. These participants were also part of the 1980s latch-key generation—fewer rules, less parental guidance, and greater experimentation. Bonnie reflected, "There were times I asked [my mother] for rules because all my friends had rules. They had rules to follow; I didn't have any rules. It was disconcerting." With a suburban form of self-education and "street" temperance forged in the world of lower-middle class experimentation, these participants went to decent public high schools. Unlike the knowledge worker participants, these participants' reviews of secondary experiences varied.

Participants raised by parents who were information servicers went to selective state universities. Unlike those in the previous category who went to college to explore ideas, the driving force for these participants was career preparation, which was oriented around notions of personal success and achievement. College served a purpose: preparation for professional work. For instance, Laura commented, "College allowed you to do what you wanted to do, and in my case I wanted to be an engineer. There is

only one way to be an engineer, which is to go to college." Likewise, Emma said that she went to college because she "wanted a good credential that would be good in the job market."

Furthermore, while participants who were raised by professional knowledge workers had entered college with the intention of becoming a university professor, evaluators in this category attended higher education to complete bachelor's level degrees in fields such as engineering, business and teaching. Isabella, for instance, wanted to be a teacher and taught in Washington, DC for three years before she got into policy work to improve education.

As I noted above, for a variety of reasons, including salary, security, and gender equity, these families had fully intended for their kids to go to college and had saved for it. Bonnie, for instance, noted, "There was money to go to college. Even more so than just the money, it was the middle class values of personal achievement." Symbolically, college would mark entry into a new status. Similarly, Emma said:

> When I was in undergraduate school, my parents were paying for it and my father lost his job while I was at school and I mean they really struggled for a few years. My grandparents stepped in and paid my tuition and expenses while I was in school.

As Emma's example illustrates, although these families had access to funds for college, it came at a worthwhile sacrifice.

Their decision to become social scientists was linked to the belief that social change could be brought about by figuring out how to fix particular social problems. Social science would provide the methods for identifying and understanding problems and for creating solutions. While undergraduate degrees were expected in these participants' families, graduate school was an alien notion that was discovered and dealt with in different ways. Laura, for instance, said:

> I got my PhD because this professor said, "Well, anyone can get a master's but you have to be really dedicated to get a PhD. So, I was like, "Fuck you." And, that's why I got mine.

Bonnie recalled a conversation with her mother about the implications of switching undergraduate majors from business to psychology:

> I sit down with my mother one day and I unveil the great news that I am changing my major to psychology. And she looked at me and she paused and she said, "Are you sure?" And I said "Yes! I like psychology" and she said, "Then know right now that you are at least going for a master's degree" and I said, "Well, I don't know how to do anything other than go to school, so that sounds ok!"

Emma recollected deciding to go to graduate school because she was working with two professors as a research assistant who told her that she had talent. What is important to note here is that whereas in the previous group parents helped their children navigate graduate school, this group relied more heavily on personal experiences and the guidance of faculty (that is, faculty who were not their parents). And, unlike the professional knowledge workers who ventured into the working world before continuing

their advanced degrees, participants raised by information servicers went straight into graduate programs after completing their undergraduate degrees. The first paid, professional experience, for two of them, came after completion of their PhDs.

As I noted in the previous section, a continuing theme is the tension between graduate school's demand to focus deeply in one particular strand of knowledge and the participants' desire to be highly skilled generalists. These evaluators told me they wanted to do research in many areas. Emma, for instance, said:

> When you're an evaluator, you evaluate a lot of different kinds of things and you can't be just one type of researcher. You have to be kind of a jack of all trades. So I'm kind of interested in that, in policy areas and stuff.

This, I believe, illustrates one effect of degree inflation that Wes Shumar describes in College for Sale: A Critique of the Commodification of Higher Education (1997). Shumar described the expansion of higher education systems as leading to a greater production of college graduates entering the workforce, which, especially in times of recession, means that the payoff for degrees decreases over time. Thus, "more advanced, more prestigious degree [a]re needed to get the same job that a college degree once guaranteed." In the figured world of evaluation, I contend that a PhD also no longer means specialization in a particular area. It is a license to do business. Thus, it is also indicative of the kind of knowledge producing field like evaluation that demands a doctor (or PhD) of all trades, principal investigators of federally funded contracts must have doctoral degrees, although the bulk of the work, in most organizations, is done by

masters-level practitioners. I will return to this later in the context of learning to reside comfortably in the betwixt space of generalist researcher and contributor to specialized academic knowledge (i.e., dissertation).

A common theme for these participants was the connection between work experience and the decision to go into program evaluation. Emma's work experience with a state education commission influenced her decision to steer away from her social science, developmental psychology. She noted:

> I saw how useful good research and good evaluation really could be, especially in the policy-making world. When I went to graduate school I thought, "What kind of job am I going to have after I get out of here? I don't really want to be a professor. I'm certainly not going to be a superintendent or school principal. I'm not insane. So what am I going to do?" And I thought, "There was a lot of work in program evaluation and if I have good background and knowledge and credentials in that area then, I've got a good job track."

Thus, unlike social scientists in the professional knowledge worker group, these participants, as their higher education pathways might have predicted, were determined to do evaluation work that was quite disparate from that of academically-oriented research.

Only one of the four participants in this category not only stayed in full-time evaluation but become passionate about it. The others entered evaluation, tried it out, and then sought out work in arenas they believed to be more accurate depictions of *true* research. Laura, for instance, who in her training to become an economic sociologist

learned panel modeling and theory, felt that the evaluations on which she worked were "dumbed down" to the point of being "speculation." To her, they were devoid of social scientific theory and used simpler forms of analysis. Furthermore, while some evaluators have described the notion that program evaluation work entails fewer politics than does educational research, these participants were vexed when they discovered the politics (especially micro-politics) of knowledge production such as those associated with working on teams, with clients, and according to contracts.

Information Borrowers

Six evaluators were raised in relatively modest middle class households in which their fathers were primary breadwinners as electricians, finish carpenters, police officers, and enlisted military personnel. While there is a marked difference in the kinds of professions, some similarities, with regard to education and higher education in particular, were apparent. These families were slightly less settled than were the former group of sometimes struggling middle class families. Indeed, major and frequent moves marked moments of considerable change for these participants. While the first group emphasized exploration of ideas and the second group emphasized college for career preparation, this group seemed to be dominated by order. Mary said:

> We were a military family, so what was really important to my parents was to raise really respectful kids. They were very authoritative with us. So, we learned to do our yes sirs and no ma'ams. And, paying attention and doing what they say. They always had the final word. There was no negotiating with children.

For families of participants in this category, college was not an expectation. In two cases, participants were dissuaded from going to college. One recalled her father telling her that she thought too much. Beth said, "I remember when I told my parents that I wanted to go to college because my friends were going to college. They didn't understand college and no one had ever saved money for that." College was the choice of the individuals. As importantly, preparation for it, deciding what and where to study, and paying for higher education would be up to most of them, as well.

Participants in this category selected colleges that were less selective than were those attended by participants in the previous two groups. Sarah, for instance, who went to a local state university said:

> It would've never occurred to me to shoot higher. I mean, certainly not trying to be derogatory to where I went to school. They were wonderful. It never occurred to me that there was something else. So, that was definitely affected by socioeconomics. In terms of more competitive, more highly selective institutions, it wasn't really more in terms of where I came from.

The stories sounded like—and, indeed, for the participants, were—journeys through uncharted territory. The tellers of these accounts tried out different fields and interests as they contemplated what to study. Like those in the previous two categories, participants in this one, as Sarah said, "had a sense that it was really important to do something in the world that helped other people." Sarah elaborated:

> I wanted to do research on issues that I find to be of the utmost importance and try to integrate that for the public good. In terms of looking at that, I started to see the research that people were doing in social policy, public policy and program evaluation especially in terms of social programs.

But, participants in this category selected academic paths that were more erratic than were those of participants in the previous two categories. Mary, for instance, noted that she wanted to be—and studied to become—a minister, a geologist, a journalist, a teacher, and a psychologist. While exploration contradicts the *telos* of being born to research, it also indicates a less structured and less supported search for preparation to do work in areas their families had not ventured, as might be expected (of course, they *were* doing research on their own interests as a form of self-discovery).

While these participants had access to less financial support and lacked advice from their families, their examples illustrate other kinds of support avenues—some of which are raced. Mary, for instance, made the decision to switch majors although she did not have a financial support in the second area, which, of course, complicated her decision. She noted, "I was in a geology major with funding that was for minorities, specifically. Being a journalism major, I didn't have a scholarship for that."

Other participants associated closely with other people who did have access to knowledge about higher education. Beth, for instance, was not sure what she wanted to do until she met her roommate's sister who was a teacher. The teaching shortage in California opened up new spaces and Beth soon found herself teaching music. She said:

> I had several really good teachers. One also worked for central office [of the school district] and sort of oversaw the arts program. And, he helped me get a music scholarship. And, that was great. So, I learned not only that I liked to play music and listen to music, but that I could write it. That was great fun.

Beth continued this way of making decisions through graduate school, as did most of the other participants in this category. Some participants linked their entry into higher education to the existence of social networks and, more specifically, social position. Katherine, for instance, said:

> As a white person, I have access to resources and social capital that I didn't even know I had at the time. I was afforded opportunities because I'm white because my parents were genuinely middle class. I had opportunities because I was connected to others who had more money.

For participants in this category, advanced social science degrees were pursued after entering the workforce as research assistants and junior researchers. Three had employers who funded, or helped fund, the degrees and the others relied on a combination of fellowships, work, and loans.

The calling to be a generalist researcher echoed in this group, too. Katherine, for example, wanted to be a counselor and studied counseling psychology. One professor told her to switch to experimental because he thought clinical "is not going to get you anywhere." Katherine eventually found herself in love with research methods. Her move

to graduate school was a careful balance between funding, methodology, social focus and the degree to which the prospective program would limit her interests. She noted:

> There was something about the Stanford program that seemed a bit too focused on the methodology and that it was missing what I would consider more of the evaluation piece that considers context, politics and social justice. Stanford was not an option. I was choosing between Berkely and UCLA. And although the Berkeley program was a good program, it was focused on STEM and I didn't want to limit my focus to increasing the numbers of underrepresented folks in STEM areas. I wanted to be able to open that up a little bit. The Spencer fellowship left it wide open and said, you just have to study something that relates to underrepresented groups, however you define them. I felt much more comfortable with that.

She went on to tell me that, at the time, she was not certain whether she wanted to go into an academic career. She said:

> I knew I loved to do theory stuff, but there was a position available and I thought, "You know what, I'll give it a try." I never went into it thinking, "I want to be an academic" and maybe that's why I spend half of my time doing the practical work."

A theme that resonated for participants in this category was the apparent lack of agency and, instead, reliance on serendipity. That apparent trend, however, veils the fact

that, with no family experience and no college savings, these participants successfully negotiated their way through PhDs and are now among the top program evaluators.

For these participants, *program evaluator* was and continued to be a major facet of professional identity. It was the label they applied to themselves and, more than "social scientist" or particular subdiscipline, it was the term most embraced. This was unlike participants representing the category of professional knowledge workers who were, foremost, social scientists or participants representing the second category whose labels were situational. It is interesting to note that the participants in this third group work or have worked in academic settings for at least part of their careers. Three became, and were at the time of the interviews, tenured faculty members and another was the CEO of a contract research organization after serving much of her career as a tenured faculty member.

Unlike the Post-Academics who finished graduate school to find scant positions in their fields, these flexible searchers entered the university at a time in which funding was changing from that of basic research grants to contract work. At the same time, the infusion of academic liberalism provided a focus, a cause, a historically-situated symbolic struggle. Pragmatically-oriented career tracks in college gave way to joining the quiet ranks of institutionalized social change.

Before continuing to the next, and last, category, I amplify my own voice for a moment, since my experiences fit neatly with my rendering of the current category. Raised in a lower middle-class military family, college was not a likely option for me or any of my peers. The punk rock movement supplanted the placated immobility of post-

hippie wannabes. While Grateful Dead inactivists dropped acid, new movements against the stark, ubiquitous Reagan-era establishment turned to symbolically violent methods of speaking out. School was just another institution set up to coerce our minds to accept an inequitable system and a nuclear age vie for global supremacy. After barely making it through high school, I joined the Army to "see the world" and, ironically, escape the violence associated with the late 1980s War on Drugs. Post-Army, although it was never part of the plan and no money had been set aside for it, I went to college, entering through the back door (aka community college) and soon found that I could succeed and that knowledge exploration was quite different from compliance with high school's standard curriculum. Awards, honor role status, and, most importantly, the voices of passionate academics brought me into a wholly-new figured world.

Throughout my college experience, I maintained varying degrees of connectedness to the marginalized worlds of working class Washington, DC and that of my then wife's Central America. Theories pertaining to social change held little meaning to me unless they could be put into practice. My road to graduate school was paved by internships with organizations serving undocumented Salvadorans, drug traffickers in Southeast DC, and returning seasonal workers in rural Panama. Increasingly, however, as I entered the figured world of academic work, I distanced myself from the places and people that mattered most. With a young family, I soon found myself working for a beltway bandit firm. From there, I went where I was needed or could find work.

Knowledge "Drifters"

Thus far, little in my depiction has contended with Bourdieu's thinking about the relationships between *habitus* and social capital (a theme that most participants have studied or even taught and terminology they sometimes used to describe their own experiences). Class reproduction in Bourdieu's sense (1977), however, might have a less satisfying fit with this category. In fact, there is little plausible correlation between where the first three class condition categories of participants began and where they were, in terms of their degrees or careers, at the time of the interviews. What does differ is the set of labels they apply to themselves and the way they were socialized to think about research as work. I believe the fourth class condition type complicates Bourdieu's analysis further. In these cases, participants were also the first generation in their families to go to college.

James noted, "I was the first one in my family to finish high school and the only one to go to college. I was the only one from my neighborhood who went to university." They were raised in families that did not hold steady work. Most lived in deindustrializing towns that once thrived on coal mining and factories. James, for instance, said:

> I had grew up in this kind of lumpenproletariat kind of a family. This was at the end of the Great Depression, so my family was all unemployed. My uncles and my father worked on the WPA. I lived in a pretty rough neighborhood. And, we

> all lived together because no one had a job or anything. There was a lot of criminal activity among some of my family. A very colorful family.[19]

These families moved often, which meant that the participants, in order to survive socially, learned how to sell themselves and make friends quickly.

Importantly for this category and unlike the previous category, while only modest means were available to the families of these participants, in the five cases, their families had planned for their children to go to college. As with the knowledge servicer category, some parents, who were unable to go to college, had saved money. James, for instance, recalls, "My mom always wanted to be a teacher. So, she would set five bucks a week aside so I could go to college." Participants in this group went to local state colleges, except one who went a somewhat selective liberal arts college on a scholarship. They participated in formal or informal mentoring programs. In fact, higher education offered the equalizing reform mechanism that would enable them to attend and complete college.

Evaluators in this category seemed to follow the ebb and flow of serendipity while those with more affluent backgrounds followed more or less careful trajectories. These participants depicted a ravenous appetite for knowledge and exploration of ideas. In certain respects, they were similar to those in the knowledge professional category. But, of course, the trajectories differed quite a bit. James recalled how he was tracked for the vocational world:

---

[19] He code switched as he reflected. At the beginning of our conversation, he spoke with the polished tongue of an academic, but as he reflected on his childhood, his diction and syntax reflected his upbringing. While I do not want to overanalyze this, perhaps there is a sort of claim to authenticity at work; an insider status that, with the degree and experience, creates a certain propinquity to those being studied in evaluation research.

> Because of where I was from and the kind of area it was, they put me into shop classes in high school instead of the college track. I didn't think anything of it. It was just where most of the guys I grew up with were. And, then I decided I would try vocational electricity. So I went in during the summer to switch my program and the dean of boys was in there and said, "Well, let me look at your record here." He said, "God, you got high test scores. What are you doing going into a vocational track?" And, I said, "Well, I don't know. They just sort of put me in there." And, he had me take courses like advanced algebra and other stuff. So, I just set off on another track.

Indeed, James' path was altered considerably. The Sputnik era identified a space for Cold War international competition, which inspired a federal focus on curricular materials and funding directed at mathematics and sciences. James' school was no exception. He said that a science teacher was "pushing everybody from the college track into science and engineering. And, a lot of the guys got into engineering. And, I was actually admitted into the engineering school." Once he was admitted into college, he promptly switched to a literature major because:

> A guy in the literature program—he had grown up working class and was a prize fighter—found me in his class and told me I had talent in the area. He said, "Stick with me and I can turn you into a good English scholar." So, I got off into the literature direction. I did that a few years and he went to Italy to write a book. While he was there, I lost track again.

James, like others in this category, gradually worked his way through several majors until he found history and earned his degree. Missing from the self stories were the assertions of agency, including the most overarching one of swimming upriver into higher education in the first place. None of these participants had planned on studying social sciences, but had explored, broadly, many academic fields.

College was a completely new experience for these participants. Since there was no map for what it might entail, there was no set limitation on the number of years or pre-set expectations for what to study. Perhaps the classed expectations of appropriate number of years in college, these participants did not experience that limiting effect. In other words, if they could be successful in college and learn to survive with alternative sources of funding, why not continue through graduate school?

After college, James promptly returned to the town where he grew up and began teaching at his old high school where teachers commented, "We didn't know you were that smart." James reflected:

> I think my world is really shaped by that, my sensitivity to social justice issues. We could see from our working class perspective how you have these middle class people interpreting event, they're unaware of it because they don't know how other classes think. If you bridge two different classes or two different culture groups, you can see how the other culture group thinks, but you don't see it if you don't notice the sea you're swimming in unless you happen to come from someplace else.

For James, the clash with middle-class values about knowledge exploration and knowledge handling was not easy. It went beyond the system that had, previously, put him on the fast track to manual labor. Even with the hard-earned credentials, classing was automatic when he returned home.

As with James, key players in the education system helped to put these participants onto more appropriate academic paths. For instance, Nick said he had wanted to be in an academic setting from the time he was in college. He had several mentors who helped him make the transition to academia. "I was ambitious and interested in program evaluation. I did not know what kind of job that was going to lead to."

Unlike their counterparts representing other class condition categories, the central identity and core professional work for these participants, these wanderers, is as *program evaluator*. Each described him or herself as an evaluator first who happens to have an advanced degree in a social science. Few of them attend conferences or receive journals in their social science disciplines, but, instead, are linked to this figured world of evaluation. Even those who have full-time teaching loads make sure that they are involved in contract work.

This category of participants represents a wider range of evaluator generations than do participants representing other categories, from recent recruits to those who became full-time evaluators well before the 1965 mandate, which is indexed as the industry's institutionalization. A distinguishing characteristic of this group of evaluators is their persistence in codifying much of the methodological language and tempering the

ethical debates of program evaluation. Through these efforts, they, collectively, contributed significantly to the institutional foundations of evaluation (e.g., the three who have been evaluators several decades were founders of evaluation's core professional organization, the American Evaluation Association).

While these evaluators were, in some respects, marginalized from traditional academic social sciences, each holds or held esteemed tenured faculty positions. They have taught and written as academicians about program evaluation. Three have contributed and continue to contribute to critical studies of evaluation, as well. The two newer-comers have also made major contributions to the academic side of evaluation. More specifically, they have created or contributed to the creation of graduate programs with particular foci in program evaluation. This category of participants dominates in the creation and maintenance of the figured world of evaluation, as increasingly, institutes of higher education offer field experience and practicum-based work, as well as coursework in theory, methodology, and business.

In Bourdieuian terms of social class, elites control marketing and the dissemination of knowledge through the media and printing houses (1977). Elites go to the best schools and have control of sources of knowledge dissemination. According to Bourdieu, working and middle classes are may successfully challenge the elite's control of knowledge production through two methods. They can support the status quo and become integrated, to a degree, into the official political decision making system. Or, they can improvise and take an avant garde or agentic stance. I believe the evaluators represented in this subgroup have both created a new space and supported the objectivist

schemata. They have actively produced a new subfield from the social sciences that is taken seriously by foundations, businesses, and government agencies. The emergent genre also has become legitimized by university systems (as evidenced through the creation and expansion of graduate programs). Yet, these innovators are somewhat constrained, in terms of the knowledge they produce, by mandate and contract, both of which give shape and meaning to the research conceptualization and activity. These constraining frames, in turn, ensure the intact delivery of the ERD through dissemination. In other words, as with knowledge production in other fields, limits restrict what can knowledge can be addressed, by what means, and by whom it may be addressed.

Academic Practitioners and Post-Academic Intellectuals: An Emerging Typology

Regardless of social position, everyone who participated in this study went to college and completed advanced degrees. I could detect little or no difference in the perceived prestige of the positions they held or average salary range once they were established. Furthermore, the kinds of clients and contracts did not vary by class background. And, except for the two situated extremes, identified in the title of this section, there was little difference in specific contract evaluation employer type. In fact, two sets of at least three participants representing three different class condition categories worked at the same place at similar levels.

How they approached the evaluation work and the choices they made along the way, as well as the kinds of influences (e.g., significant narrators, initial research experiences, etc.) that affected those choices, varied by social position and sense of agency. In other words, while class conditions did not seem to influence the overall

success of evaluators once they became proficient in the figured world of program evaluation, it appears to be linked to how they got there. Perhaps more importantly, narratives pertaining to class conditions, parents' education, and professional background seem to be connected to a professional identity. What might be important about the fact that while participants representing more affluent backgrounds tended to avoid the title evaluator, those from less affluent backgrounds gravitated toward it?

While it might be a stretch to use self-applied labels in isolation to say much about class-based differences in evaluator identity production, their descriptions of the knowledge work they do and its implications for this figured world, their accounts depicting how they became versed veteran evaluators, and their historicized identity narratives that seem to naturalize that practice and those accounts provide compelling evidence for the construction of a metathematic typology based, loosely, on social class extreme experiences. What I wish to emphasize here is the difference in identity narratives based on social position in the seemingly-natural ways of conceptualizing evaluation work.

Unlike the class conditions categories that emerged effortlessly (at least in my mind) from the data, the typology I offer here is somewhat abstract and drawn from less systematic assembly of evidence. Furthermore, these categories are not as rigidly bound. Most participants probably represent a middle-ground category between the two extremes of *post-academic intellectual* and *academic practitioner* and possess varying attributes of both extreme types. Nonetheless, this suggestive analysis offers potential for linking identity narrative, ERD, and the contemporary political-economic context.

Post-Academic Intellectuals

The first category, based on the experiences of people doing evaluations concentrated in the more affluent categories described above, is the Post-Academic Intellectual. Distinguished by a guarded sense that they do evaluations to make a living, but that they do not refer to themselves as *evaluators*, most of these contracted social scientist knowledge workers were clustered around families where knowledge work was part of everyday life. This translated, easily, across generations. These participants were raised in families in which at least one parent was a tenured academic. Most described utopian lifestyles as a part of their upbringing combined with a sense that education was valued not so much for the purpose of earning a particular living, but for the sake of personal and civic edification.

Post-academics have left academia—or abandoned the ideal of entering—as tenured faculty because, according to them, there was a trend indicating that the academic world no longer afforded ample opportunity to produce meaningful work or continue to delve deeper into their area of specialization. Those who had abandoned the idea of going into higher education as a career path before entering that world described changing conditions that did not fit with their expectations, which were based, in large part, on the experiences of their parents. One participant described this as such:

> Some people say, "I left academia to do scholarly work," and it's so sad to hear. Or like 90% of academics I know say, "I don't have any time to read," and you just think there's something wrong with this picture! For me [working in a small, non-profit contract organization], I like the freedom to be engaged in the research.

Anne, a dislocated social scientist who left the academic world, commented on the draw to program evaluation:

> One thing I actually find truly compelling about program evaluation, and it's not that I only want to do program evaluation, I really like balancing the more open-ended ethnographic thinking, there's something free to me about the constraint of program evaluation. It's that axiom about your career when you have limits. So, it's like what the task is at hand and your job is to figure it out as best you can. I can relate it personally. I'm a person who loves Dada and I love resolving math problems. It's two sides. That's the personal side of it I guess.

Between the constraints of the obligatory, Anne located spaces for exploring intellectual freedom that she could not find in an academic setting. Next, I offer several examples that illustrate some of the ways post-academics infuse intellectual freedom into their evaluation work.

While the post-academics I spoke with continue to publish in scholarly journals, doing so does not contribute, directly, to their professional standing or the compensation they receive. On the contrary, sometimes, because of the time taken away from contract work and because of intellectual property issues, scholarly knowledge production competes with their contract work. For example, Helen, who I described in Chapter 4 as having encountered a tremendous challenge when she attempted to include the voice of community members in a final report, also attempted to write a reflective, critical article for publication with a group of co-workers. When a draft of the manuscript went through

the company's quality assurance process, remarks that reflected on the company were removed or written in a positive light. Ultimately, she withdrew herself from the task and was satisfied that none of her coworkers continued with the manuscript. In some cases, work on publications that is critical of either clients or the field of evaluation may jeopardize everyday contract evaluations, since doing so requires taking a stance.

A commonality among these knowledge workers was a tendency toward a *soft activism* that took the form of bringing equalizing perspectives, based on their social science training, to their everyday work. One example is the intention of "educating" clients. Final deliverables contain all the agreed-upon sections, but, in addition, include references or allusions to additional work that may provide feedback that pertains not only to the implementers and targets of the educational reform, but the funding organization and developers, as well. For instance, although perhaps not more than a readily available example, in one of my ongoing projects with a large-urban school district, a recurring theme was the lack of implementation, which was tied to the lack of professional development at the school level. As the politics of "turnaround" models made their way to the central office, principals and "recalcitrant" staff became the pinpointed culprits of low test scores. The district asked me to help them identify practices in these struggling schools that might aid the district's application of the turnaround model.[20]

[20] While there is little consensus and even less research about what constitutes an effective school Turn Around, one federally-accepted model is to relieve and replace the principal and at least 50% of the staff.

I decided that my task would become not so much the protection of principal's jobs, but helping the district and state department of education, as well as the schools and their communities, understand that the lack of professional development was tied, directly, to an un-navigable finance system and lack of district and state-level support, which resulted in no funds making their way to any of the schools. Without funds, PD was not possible. The formative evaluation led to a series of reflective sessions with district leadership and, later, school-level leadership, that helped identify weaknesses in the existing system.

Another strategy is to conduct strategic planning with decision makers. In particular, I have observed (and been fortunate enough to participate in) this kind of work as a kind of action research with senior staff through logic modeling and blue-printing. For instance, I worked with a mentor evaluator who happens to fit the post-academic category well. She gradually brought decision makers of another large urban school district to question their assumptions about expectations pertaining to low-income struggling students, communities, and teachers. After a year of conversations around the development of a logic model that depicted how teachers and students would be best served by the district, district leadership had completely altered the program's blueprints and the assumptions they held. This was accomplished, in part, by collecting data and suggestions for program improvement, along the way, from teachers and students that helped naturalize their voices at the table, at least vicariously. These voices were combined with terms such as "funds of knowledge" (González, et al 2005).

And, in addition, post-academics work closely with students Socratically. Non-profit, university-based centers, and for-profit organizations offer graduate student mentoring programs. For instance, this past year, as I mentioned earlier, I worked with a visiting professor from Mongolia who wanted to learn about program evaluation. Furthermore, new recruits and even evaluators who have worked in the field for a decade, work with mentors to expand their methodological toolbags (including my example about logic modeling above). Recall the example of the experimental psychologist who learned from a post-academic how to collect and analyze interview data both as formative indicator and reflective device.

The effect of this kind of work on the figured world of evaluation includes opening up spaces for questioning the status quo while describing to social science disciplines the existence of this figured world, helping clients understand their responsibility and likely effect on program implementation, providing research intern experiences for graduate students and professional development for evaluators wanting to expand their repertoire, orchestrating multiple voices with varying levels of social position and decision making ability, and infusing, even if subtly, social theory into client work.

Academic Practitioners

The second type, and based mostly on the narrative accounts of experiences of less affluent contract evaluators, is the Academic Practitioner. This designation applies to a subgroup of people who conduct and coordinate contract program evaluation work in academic settings. At the same time, they have found ways to lead successful tenure-

tracked academic careers that include teaching, publication, community service in academic departments that include coursework in program evaluation or that focus, squarely, on masters- and doctoral-level preparation of program evaluators. By doing so, they bridge the theoretical and practical. In some ways, they remind me of the non-commissioned officer status of Gramsci's *organic intellectuals* (1971) as they define their everyday role as recruiting and providing applied research opportunities on funded projects to non-traditional college students to this figured world. Yet, this interpretation denies academic practitioners agency since they are not simply conveying a message from en high.

Academic practitioners' identity narratives locate their upbringing at the lower SES extreme of the class conditions continuum I described above. With one exception, that I describe in the closing portion of this section, these participants pursued higher education with limited support or guidance. At this lower social class conditions extreme, the evaluators refer to themselves, publicly and professionally, as *program evaluator* more so than as representing a specific social scientific discipline. Most of them no longer attend conferences in social sciences, but are almost always present and actively participate in American Evaluation Association and American Education Research Association Division H (a subgroup of AERA that is dedicated to interests in Research, Evaluation, and Assessment in Schools) annual conferences.

Academic practitioners see evaluation as their calling and put that calling into action as a translation of theory into practice and the production of the steps of that practice, according to the ERD. Their aim is to bring about change in education

systematically and gradually. Academic practitioners can be divided into two broad groups: 1) center-based researchers and 2) full-time teaching faculty.

Several of the evaluators who define this category, described their work as directing policy and evaluation centers or institutes at universities. The talent for bringing in external funding is a distinguishing feature between academic practitioners and other faculty. They have become successful at sensing prospective client needs and negotiating contracts. Some of them bring in millions of dollars each year and do not have a teaching load, except to oversee graduate students. When I asked one participant, who directs a large policy center, to elaborate on the term he used to describe himself, *academic entrepreneur*, he said, "Somebody who is always looking for funding and always trying to do sponsored research, I guess is what that means. "

Academic practitioners also may serve as faculty who have full-time teaching loads in departmentally-required methods courses or as members of newly-formed program evaluation graduate programs. As gatekeepers of the figured world of program evaluation, they teach prospective future evaluators and provide substantive work experiences. Katherine described her transition away from the private consulting world to her emphasis on securing work experiences that provide her summer salary, keep her connected to contract work and clients, and provide research opportunities for her graduate students:

> The last few years I've been prioritizing projects where I can have graduate students work with me on those projects because that way I'm offering them

support, mentorship, and those are things that I think are important to people who are pursuing an education.

Regardless of whether working in a non-classroom evaluation unit or whether dedicated to teaching evaluation and guiding students, the game for tenure-track academically-located practitioners is to maintain a balance between remaining connected to the contract world, being able to produce academically-oriented knowledge, and preparing students for pragmatic work. Contract evaluations keep them connected to the practical world while they also provide access to research for their graduate students. The knowledge they produce, fitting with their position, is somewhat de-stanced, however.

The academically-oriented knowledge produced by academic practitioners is often geared toward audiences of program evaluation journals or books about the field of evaluation. In these, evaluation is a topic area (as opposed to the subject of the evaluation). Specifically, these evaluators contribute to heuristics and theories intended to depict and measure program logic, efficacy, and feasibility. They also produce and contribute to epistemological and ethical foundations for the field. In essence, therefore, they bridge social sciences and evaluation to codify this figured world that is still very much in development.

Unlike post-academics, theirs is not so much a battle with institutions, but a pragmatic undertaking with the intent of identifying and sharing how those client and funding institutions work and ways to work effectively with them. By doing so, academic practitioners build up rules of how to do evaluation. While artifacts of their

work may indicate less inclination for challenging their foundations, they attempt to work with those organizations to improve program efficacy gradually. The business side is expected, appreciated. For instance, while it might raise a red flag among post-academics, intellectual property rights is not perceived as particularly problematic. One practitioner commented:

> I'm not so much adverse to the notion of intellectual property as much as how it's used and how people need to play it. My stand is that this is clearly something we're going to have to identify, there's going to be some code of conduct around these matters and so, I'm comfortable as one of the writers on the recent Program Evaluation Standards.

I do not intend to suggest that academic practitioners are simply rule-guided adherents to a totalitarian-like system. In their everyday work, they do encounter varying qualities of evaluation work and they are charged with ensuring their students show potential in the new generation of non-academic evaluation. The context contains some of the obligatory, sense of personal accountability and reflective matter for expectations. Nonetheless, the contact is not *that* deterministic. Still at work, in their heads as personal history and orchestration of voices, is a self-regulation of agency.

One participant, the son of a doctor, complicates the simple dichotomy I suggest, at least for the first 25 years of his career. While working in for- and non-profit organizations, he published articles and several books. When he moved to a prominent university position to direct a policy center, he brought with him his quarter century of

experience of negotiating successful grants and contracts. He also taught sociology theory courses and courses geared toward program evaluation methodology. Throughout the interviews, he defined himself situationally. Sometimes, he referred to himself as a sociologist and at other times, he called himself an academic entrepreneur. He may represent a late-career version of an in between category, as I discuss next.

Bringing the Types Together

As I mentioned at the beginning of this section, most participants probably locate themselves somewhere between the two categories I describe, exhibiting behaviors and taking approaches to their work that borrows from some combination of the two. In that betwixt space are contract evaluators, most of which have a decade or less of experience and who do not define themselves professionally in black or white terms. From my limited knowledge of their proclivities and after poring over stories they told me about themselves, I would guess that some newer entrants to the figured world of program evaluation are headed toward one pole or the other. Other participants have remained, for the most part, in this space for 15-20 years, however.

They define their professional selves situationally at conferences. I have called myself a program evaluator at the annual conferences of the Society for Applied Anthropology and an anthropologist at the American Evaluation Association annual meetings. I do not do this to be pernicious or mischievous, but to accomplish two personal objectives. First, it accurately depicts my incomplete and mercurial professional identity. Without a PhD, I cannot serve as a PI and, therefore, have less say about the specific projects on which I work or how, exactly, I work on them. That is about to

change and when it does, I suspect I will have a few years of a new kind of playing on the fringes to determine what resonates most with me. Taking this situational stance underscores my specialization within that realm. I am an anthropologist who works on contract evaluations and I am an evaluator who brings an anthropological lens. I go to both kinds of conferences. Yet, I am complete in neither space.

I believe the two most seemingly disparate types of program evaluators represent effects of the same systemic political-economic shift and recourse to the ERD for solutions that might address challenges associated with the shift. Both the search for post-academic intellectual freedom and the tenure track success of contract evaluation in academia correspond to localized adaptations to the shift in academia toward increased privatization in the broader context. The differences in reaction are tied to expectations that are, perhaps, ossified in a combination of personal narratives pertaining to childhood socialization and educational pathway preparation for the uses and production of knowledge, as well as the system of sanctions and personal calling associated with the particular work context.

Survival of the two extreme types of program evaluation—dislocated intellectuals and pragmatic academics—is brought into close proximity when compared to academically-oriented social scientific research because of what appears to be a strategically chosen generalist orientation, but, which may more accurately represent a calling for flexible specialization. As I argued in Chapter 2, the history of professional program evaluation has revolved around the ability to manufacture knowledge and to be able to repurpose knowledge workers, in terms of particular content areas, clientele, and

methodologies quickly to adapt to changing economic conditions according to the lens of the ERD. The question remains: How much of this is agentic and how much is response to a changing political-economic system?

In Chapter 6, I provide brief case studies of social scientists who do evaluations. The chapter is arranged according to the three ideal types I described in this section—post-academic intellectual, contract evaluator, and Academic practitioner.

Chapter 6: Vignettes of Three Evaluator Types

In this chapter, I offer three vignettes to exemplify key points I have made throughout. They are oriented around the three types of external program evaluators depicted in Chapter 5. These types are ideal and few people fit them neatly or completely. Given the lack of reliable information about the scale of evaluation and, therefore, generalizability of my suggestive findings to the whole of program evaluation, while it is impossible to gauge what percent of evaluators fit these types, my guess, based on experience and conversations with participating program evaluators, is that most represent the second type of evaluator I describe with the less glorifying title, *professional evaluator*. At the far extreme, and considered in the third case presented in this chapter, is the post-academic, which probably represents a relatively tiny fraction of program evaluators. What strikes me most is the existence of this category much more than its expansiveness in relation to the other types. The first category I describe, the academic practitioner, is probably numerically, in the middle ground between the other two, particularly as academic departments expand preparation programs for future evaluators and as academia comes to rely more heavily on contracts in sponsored programming.

The vignettes also provide a more holistic illustration of the thematic findings I have offered in this study. I have organized the vignettes into three broad sections. In the first, I describe each person's connection to program evaluation and how she makes sense of that connection. In the second section, I depict each evaluator's personal historical pathway into program evaluation, with lots of familiar tidbits. And, in the last

section, I consider type-specific struggles and beliefs that enhance and refine both the figured world of program evaluation and the identity narrative of the evaluators that inhabit it.

Category 1: The Academic Practitioner

Connection to program evaluation

At the time of the interviews, Kathy was beginning her ninth year as a full-time tenure tracked professor. She had been promoted the year before we spoke. During our second conversation, she told me she had accepted a more prestigious faculty position in a department of program evaluation at another university. Because of the switch, she planned to continue doing contract work, but it would be for the purpose of giving students research experience. She explained:

> I'll retain those sorts of things, but I think in terms of employing students. It would be focusing more on the larger contracts. I need to learn to look at folks and say, "I'd love to do this but I can't" and learn to say "No" more easily.

She added, “The new position is probably a better fit and, as much as I love social science theories to inform what we do in evaluation, a lot of my substantive work fits in education.[21]”

[21] Note the shift in her identity narrative. During the first interview, she identified the link to social science as imperative, especially for informing theory. I do not consider this to be a major contradiction, but a switch in self portrayal that meshes more accurately with her revised connection to the figured world.

Kathy described the three facets of her work as a practitioner evaluator—theory, teaching, and contract work. Of theory, she was careful to note: “I use that term loosely because these are not really theories. These are, at best, models for practice. Thinking about how to promote these models and advance them.” She clarified her thoughts on theory by saying that theories of evaluation “should be connected more closely to some of our other truly social science theoretical work. Evaluation is not grounded firmly enough in that theory. Often, we dilute them from the more theoretical into practical application of them.” An example of her theoretical work is posing questions about the potential impact of using the wrong assessment instrument on policy created for underrepresented populations.

Kathy also carried a full-time teaching load with methods courses, an evaluation theory course, and an introductory course on program evaluation. She said she attempts to connect these to her theory work by emphasizing:

> The theoretical literature we have on evaluations because it speaks best to this idea of these mental models when you have the paradigms and they become diluted and translated into, "What does it mean for me as a technician”? So it's making sure that the technician understands the theoretical premise behind this.

In other words, she said she attempts to make clear to students that evaluation involves more than mechanical application of research methods.

The third facet of her work involves staying connected to contract work, which pays her summer salary and that of her graduate students. She confessed that it “doesn't

really contribute back to a theoretical literature in the way that someone who was evaluating a program from within that discipline might truly be interested in pursuing those sorts of questions." She went on to say, "I come in as a practitioner, as a service provider and I conduct a study. I'm not somebody who evaluates a program on child maltreatment and then is interested in churning out side publications from that evaluation about the impact of child maltreatment." From her angle, while some evaluators situated in academic positions use those evaluation derived data for publications, she made clear that she steers clear of this for reasons of rigor and ethics.

She described two kinds of contract work. First, her university-based sponsored projects included multi-million dollar projects and "and little things like twenty thousand dollars a year as a sub on somebody's prime, because they need an external evaluator." She added, "I've been prioritizing projects where I can have graduate students work with me on those projects because that way I'm offering them support, mentorship." In addition, she described contract work on the side, outside of the university as a hired gun. This presents her with a need to "shift my thinking away from seeing myself as an evaluator and shift it to seeing myself as someone who is in a consulting business." And, this has implications for the way she made sense of her specific connection to program evaluation and, therefore, her professional identity.

Kathy described her view on evaluators' professional identity:

> At AEA, in terms of identity, there's the practitioner and academic divide. I'm asked whether I'm an academic or practicing evaluator. The question then moves into the next level. I've taught school, I've been academic, what do you do? And

maybe further broken down into, "Do you consider yourself someone who theorizes or are you a practitioner?" When I go to other conferences like AERA, I usually describe myself as a methodologist.

For Kathy, the question is not whether or not she is an evaluator or a social scientist, but what kind of evaluator she is. Her situational selection of titles is dependent upon how she reads each context. She told me that her sense of situationality is generated in a betwixt status. She explained, "I don't want to stake claim to the substantive area that I'm not part of, I mean legitimately. Yet, when you describe yourself as an evaluator and you don't have a sense of focus, they don't necessarily know how to make sense of you."

Path into program evaluation

Kathy's story of how she wandered into program evaluation, rather than intentionally preparing for this career, is similar to those of most of the other evaluators with whom I spoke. In fact, for her, going to college was not a big question. The bigger question pertained to what she might study. Like many others represented in Chapter 4, she stressed that going to college was "about finding something that I could do where I felt like I was making a difference." Her story illustrates the activation of a seemingly-natural interest in social science, an eventual discovery of program evaluation along the way, and gradual shaping of her proficiency in that figured world.

Kathy grew up in a working class family in New York City. While her father was a handyman and her mother taught piano, she noted, "There were lots of people richer than me, but I had opportunities because I was connected to others who had more money,

who had different opportunities as a result of that, and had folks like that in my extended family." Kathy was well aware of white privilege. She noted, "White people are afforded greater opportunities in this country. As a white person, I have access to resources and social capital that I didn't even know I had at the time. I was afforded opportunities because I'm white." The racialized awareness served for Kathy as both an element of social position and a calling to action, which influenced her career decisions.

She entered college with the intention of becoming a teacher. Her first classroom experience, however, redirected that interest and she switched to psychology. Right away, that field piqued what seemed to her a natural proclivity to do research. She reflected:

> I can think back to the class that I was most interested in was the one that everybody said they were dreading. Research methods. When I took it, I thought, "This isn't a bad class!" So constructs of internal and external validity weren't the most fascinating part of it, but going out and trying to do a little study was really interesting."

Although interested in research, she did not connect it, immediately, to her work.

After college, Kathy resisted the urgings of friends and family to consider law school. She recalled coming to the conclusion that "The only way I understood that I could do something was doing some sort of counseling work. I didn't have any other clear framework in terms of how to move forward with a career." Tying her interest in making a difference to her psychology training, she worked at a psychiatric hospital as a mental

health worker. The experience led her to pursue a masters in clinical psychology, which led to another pivotal moment. During her first year in the program, an influential figure redirected her. She recalled:

> A gentleman on the faculty who was an experimentalist said to me, "I don't know why you're wasting your time with counseling. It's stupid. It's not going to get you anywhere. Do you want to be a counselor your whole life?" He says, "If you want to do something that's going to employ you after you leave here, other than some crappy job conducting one-on-one counseling sessions, then do a one-year masters in experimental design." Oddly enough, I listened to him.

Although she continued with counseling, she focused much of her training on research methods.

With her MS, she worked as a counselor in a managed-care facility. After a few years of patients who she said, "really rocked my world," she called a former professor to inquire about possible positions. He hired her part-time on an evaluation study, although Kathy was unfamiliar with evaluation at the time. She said, "He didn't have that much money and it was towards the tail end of the project, but afterwards he hooked me up with a woman who was working at a well-funded institute."

The social service institute, funded by "a ton of grants," hired a new executive director who had worked with similar organizations that were striving to become more accountable to their clients. Kathy characterized the new director as someone who had recently helped a social service agency "disentangle a scandal." The new director

worked with Kathy to revise the institute's client reporting system. After six months of reading literature on her own and "doing evaluation blindly," Kathy realized that there was a lot to evaluation. She said, "I ended up spending a few years developing an infrastructure for them to churn out the output data that they needed for their grants, but also to begin looking at some of their outcome." She recalls:

> I realized evaluation wasn't just experiments or randomized control trials. It was something else and I had to figure out what that something else was. It was responsive to clients, it had to be user-focused and it had to be sensitive to the political context.

After self-directing as an internal evaluator for a few years, Kathy decided to formalize her evaluation training by pursuing a PhD. She was accepted by three prominent universities, all of which offered substantial financial enticements. She removed one from consideration immediately because of its over-emphasis on methodology and lack of attention to "what I would consider more of the evaluation piece that considers context, politics, and social justice." Of the two remaining, one was STEM-centered and the other was intended to support studies of her choosing. Ultimately, she accepted the open fellowship because she wanted to be able to do research in different areas, illustrating the tendency among many evaluators to be what I have referred to as flexible generalists.

Kathy did not enter her PhD program with the intention of becoming a faculty member. But, afterwards, when she saw a position announcement, she recalled, "There

was a position available and I thought, 'I'll give it a try.' I never went into it thinking, "I want to be an academic' and maybe that's why I spend half of my time doing the practical work." Perhaps as much as the opportunity, Kathy has positioned herself, through experience, social networking, and knowledge of contracting, to be connected to that figured world.

Enduring struggles

Kathy's thinking about how she made sense of her day-to-day work as an evaluator helped in identifying points of contention. In these contentious spaces, Kathy described the confining and enabling structures, as well as the moments in which she felt equipped to assert her agency.

Of considerable importance were Kathy's thoughts on evaluation's role and function in contributing to social justice. She commented, "If you buy into what Ernie House [an old-timer program evaluation guru] writes, which is, 'Evaluation helps to determine who gets what,' then you play an important role in ensuring that people have equity and access through this thing called evaluation." From her perspective, evaluators accomplish this by "generating information to really help people make important decisions," which relates back to some of the ideals related to social justice, either at a micro level or a more macro level. This is an important part of Kathy's story because to be an authentic evaluator, to really bring about change, means staying connected to the industry in terms of contract work, but it also means contributing to the figured world's continual refinement and boundary maintenance. This multifaceted status has implications for how she positions herself among and described herself to colleagues.

As she was preparing to parlay from one particular space in the figured world of program evaluation to another, she rethought how she interacted with it, what her priorities were, and began asserting her agency in new ways. Kathy told me that she had difficulty drawing boundaries around her time and work "and it may be because I love what I do." To improve this, Kathy described some of the resolutions she had made that would, ostensibly, help her make her identity narrative more consistent with the transition:

> I've been limiting the smaller projects that I do that I don't do through the university. I'm constantly reflecting and saying, "If you want to shift the work/life balance, it's not going to happen using the strategy that I've been using: 'Yeah that sounds great!'" because it all sounds great. I have to think about what I'm going to be doing, with whom am I going to work, and how much of my time will it realistically take to do it." Whatever I think now, I double it and I stick by it. I need a half a day to prepare. I need another day after I've done whatever it is I've agreed to do to make sense of it.

In other words, in order to thrive in her new context, she foresaw a need to shift priorities. These shifts in priority, for instance, adjusted her sense of practice from helping where she could to a more formalized set of contractual agreements.

Less in terms of personal choice, flexibility amidst change is a fundamental part of the evaluation game. Kathy told me that recent changes in the field from "an environment that honored different methodological approaches to some real stringent

guidelines" forced her to think about how to communicate her role to others. Increasingly, her cumulative experience made her more than a technician-helper. She described how she is called upon for her knowledge in particular areas.

> By virtue of having to explain what you do conceptually. I'm constantly reflecting on what it is that I do. I never see myself as the person who's teaching something to the person with the substantive expertise. I usually see myself as the person that helps to shape a study, that helps argue for the inclusion of particular groups and I can argue why. But, I don't really see myself as the one who is bringing the particular expertise. This is an example of an evolving identity to being more than just a methodological expert. Perhaps, there is a broader knowledge base that we bring and recognizing that that broader knowledge base influences how colleagues think about what it is they do as well.

One of the most noteworthy exogenous changes Kathy noted over the course of her career was a move toward capacity building in evaluation. She commented that, unlike when she began working as an evaluator, it became imperative that evaluators teach clients about how evaluation works. One major repercussion of this is on the degree to which clients are both informed about evaluation and how there is an increased tendency for them to partake in the development, conduct, interpretation, and use of the evaluation. She noted:

> As external evaluators were doing their work, it became evident to them that it is advantageous to have people who are part of the program who understand what it

> is that you're doing. I don't know of too many evaluations now that actually get done well without the support and the assistance of people who are related to the program. There's a self-interest to build some capacity because you can do your job better, the emphasis has helped support some of those capacity building activities, as well. This idea that if you want people to use it then you've got to answer questions that they care about and in some ways to get them to come to a place where they can identify questions that they care about, you have to do some teaching.

According to Kathy, this sharing of evaluatory work has positioned program staff to be able to respond more effectively to evaluators who might influence funding. But, it has also made her more reflective about the work she does. Her work goes beyond the collection and presentation of an array of data.

As an academic practitioner, one of her concerns is with the expansion of the field, not the least important aspect of which is addressing the social justice calling in terms of diversifying both students and faculty. She told me that the "field reflects that huge cleavage that exists between the academy and the practical world." She continued, "When you meet evaluation practitioners, they're not exactly the most diverse group of folks. We need to work to diversify practitioners, but when you look at them relative to the academics in the field, it is. Practitioners are by far more diverse."

She then described "efforts" at the private university to attract greater diversity.

> It is not easy to attract people to this field. I don't know exactly why. I think we have failed miserably at [college] at hiring diverse faculty. Every time we open a search, I say, "Let's not interview anybody who isn't from an underrepresented group. Let's just leave it at that." That's how we diversify. Don't bring anybody in here who's white. They look at me like I've lost my mind.

Kathy represents a refinement of the move of program evaluation to the university setting as a discrete departmental unit, rather than as a specialty area within an education program or a faculty focus for a social scientist. In this context, program evaluation expands into areas of specialization. This locates her at the forefront of codifying the rules of the ERD and taking on her own mission to diversify the field, enhance social justice, both internally and as an aim of evaluation work, and make program evaluation more than the operation of social science methods.

Category 2: The Professional Evaluator

Connection to program evaluation

Sarah came to external program evaluation after she had worked as a basic researcher in a university setting, a market researcher at a private company, and an internal evaluator in for a large urban school district. At the time of the interviews, she was beginning her second year as a research associate at a university-based qualitative program evaluation center. She said, “I feel like a consultant to some degree. I still get to kind of be the outsider looking in a bit, which I have always been drawn to.” One part of the job that she found particularly satisfying was the large urban school district’s

apparent interest in hearing what teachers and students have to say and using that information for continuous improvement. This was the space for social justice. Sarah reflected, "Have you ever read the book the Lorax? He spoke for the trees. You have to listen to the little voices that are making up the base of what you are trying to do. I still see that as our role. I feel like we get to be the voice on the behalf of people who don't have one."

As with other program evaluators with whom I spoke, Sarah described difficulty articulating to others *what* she is. I'm doing work that is so far from the traditional notion of psychology that sometimes I don't even remember that I am a psychologist. Like Kathy, Sarah told me that she selects a professional label according to the situation in which she finds herself. As she said:

> It's different depending on who I'm speaking with. Like if I am talking to somebody who is either in a social science field or an education field, I'm usually a little bit more detailed, but in general, I introduce myself as a contract researcher. I go straight to researcher and then if they ask any further, I'll say education research."

She added, "I think I'm classically conditioned by it. I don't even bother describing in-depth what I do until somebody has a context to hang it on, otherwise it's just frustrating." Sarah has moved out of her specific social science discipline into a more generic "social research" zone. She commented, "I don't like calling myself an evaluator. I don't think I would ever say, 'I am a program evaluator.' It sounds

enforcement-oriented." As she talked about this, she maintained a distance from program evaluator.[22] The term she used to self-describe, education researcher, fit with how she conceptualized her work. For instance, she said:

> I always thought the most meaningful way to approach evaluation is to use the knowledge that you have about research methodology, and I guess often in my case in psychology, once you use those to look at the research questions and design what you think would be a feasible and useful way to best answer the questions. That might be a self-evident kind of thing but that's kind of how I approach it.

This does not automatically mean she mechanistically applied social research methods to her evaluation work. As she remarked:

> In order to be really good at the work, you need to be able to have vision. You have to be able to look at something and see it and not just take for granted that somebody might be saying yes or no to a question that you asked, but they might give other cues to consider while telling you what they think you want to know."

Therefore, as Sarah began to take on more responsibility as a project lead, she revised how she made sense of and, thus, described her connection to program evaluation.

Path into program evaluation

[22] Throughout the interviews, however, both before and after our discussion about titles, Sarah offhandedly referred to herself as a program evaluator.

Sarah entered program evaluation as part of a personal quest for meaningful work and what she described as "truth." Like Kathy's story, Sarah's illustrates how a *natural* connection to social sciences took hold and manifested itself in her and the ways in which she made choices that refined that proclivity.

Sarah grew up in a tightly-knit, "very white" Western Pennsylvania coal mining town. She said, "It was like the Simpsons. There was maybe only one family of each ethnic group. Everyone else was incredibly white." Her parents and much of her extended family still live in the community. Sarah was the second generation to go to college (the first to attend graduate school). Both her father and mother studied career-specific fields (law enforcement and nursing). Like Kathy, Sarah recognized how her social position influenced her choices. She recalled, "I had a ton of advantages. I came from a family who could help with my paying for college. I paid for a good deal of it myself. I feel like I come from a place of privilege." This awareness became an impetus for her desire to contribute to planned social change. For Sarah, an important element of this work is using data to inform stakeholders about the "truth," a value she developed in juxtaposition to her father's focus on "justice."

Sarah attended a large, local state college straight out of high school. She characterized it as, "the most amazing experience because I had come from this really small town." She recalls, "I was like 'Oh my god there is a whole other way of thinking about this.' Especially the feminist literature and philosophy and perspective. Again, it was the truth emphasis that resonated with me… It was heavenly coming from a tiny little town and going to a big-thinking environment." She chose that college because, for

her town, “it was kind of a big deal. It wasn’t the best, but it has decent standards. In terms of more competitive, more highly selective institutions, it wasn’t really a norm where I came from.”

When I asked her about how she selected her major, she said, “I always knew I was going in that direction. I was born to be a psychologist and now I am a program evaluator.” She recalled, “I always had a real interest in social dynamics; the way people interact and the non-verbal communication that people put out.” She reflected on how, according to the way she self-authored, being a researcher was a *natural* part of her, “Yesterday, I was at a meeting and I could barely focus on what was going on because I was paying so much attention to the social dynamics.” In this case, social science got in the way of program evaluation. It was the default internalized voice. According to her autoscript, it had *always* been there.

Unlike Kathy and unlike Sarah’s parents, Sarah did not link her program of study directly to a career objective. When I asked what she wanted to get out of college, Sarah told me it was about exploration. More specifically, she said, “I didn’t think a whole lot about what was going to happen after that four years was up.” She added, “I might have had a hunch that I would continue on and get a law degree or some kind of advanced degree.”

One of the ways that Sarah explored and sought truth was through her involvement in volunteer activities. She recalled:

> I volunteered for various programs. I wrote for the school paper; an opinion column. I pointed out hypocrisy. It was around the time everything was

> happening with Bill Clinton. I liked writing about hypocrisy although meanwhile I was living hypocritically.

The interest in exposing the hidden truth and in reflecting on how she was "living hypocritically" became enduring themes.

Sarah went directly from her undergraduate program to graduate school. When she got to considering graduate schools, as she had with her selection of undergraduate programs, Sarah reflected on her social class:

> It would've never occurred to me to shoot higher. I am certainly not trying to be derogatory to where I went to school, but it never occurred to me that there was something else. That was definitely affected by socioeconomics.

As for her consideration of prospective majors, she said, "To do anything in social psychology, you need a doctorate. I grappled with clinical versus the social. I think I was incredibly intimidated by the clinical component. I didn't know if I wanted to work one on one with people." As had happened with Kathy, Sarah opted for research.

The interest in figuring out what makes people "tick" persisted. During graduate school, Sarah worked as a lab assistant. The experience helped her both confirm her career orientation toward research and contemplate the practitioner-academic divide:

> I didn't want to do just academic research that is going to go into a journal. I wanted to do research on issues that I find to be of the utmost important and try to integrate that into a more mass, for the public, greater good. I saw research that

> people were doing in social policy and program evaluation, especially in terms of social programs. It really interested me.

Armed with this new knowledge, she began gearing herself up for a career outside of academia that would permit her to use the exposure of underlying truths for the betterment of the society, especially those segments that were less empowered. Sarah said that graduate school also continued her quest for personal truth:

> I am a big believer that patterns keep happening until you learn what you need to learn from it and then you move on. The big changes that I got from it I became comfortable and explored my own truth. At that point, I started thinking I wanted to do program evaluation to get social ideals out into the world.

In other words, she was attempting to solidify her social science proclivity by making her personal life commensurable with that of her prospective career.

Sarah told me that she wanted to go into an applied social science "because I felt it was important work we were finding about human nature and how to make the world better." As a part of her graduate training, she did an internship in public relations crisis management. She recalled, "If somebody was getting bad press, we would put out the fire. I would search to make sure there was no bad press like a food safety crisis for a client like a food manufacturer." She said that the experience was "horrible," but that, at the time, she felt it was better than waitressing. "Looking back, I should have waitressed."

The next year, she began conducting educational research for a non-profit organization as a policy analyst. She recalled, "I read the job description and the required skills was me to a T. I was going to do literature searches and I was going to do data analysis and try to make this program better and, thereby, make the lives of students better." After a year, however, she had grown "leery about the quantitative nature of it" and decided to try out market research. She said, "It felt glamorous to me at that point." She added, "The owners of the company were PhDs in social psychology. We spoke the same language and they saw the hypocrisy of what was happening with the clients, but we discussed it and we named the elephant in the room and then continued to do the work anyway." The position also paid well and that enticed her to make the switch. In fact, as she recalled, "That was one of the biggest driving points of taking that job. He was offering me money like I would never make my current work. I was so fed up. You know how you feel when you are working on your dissertation."

As I feigned ignorance at her comment about dissertating, she went on to describe how the position tested her ethical foundation and sense of internal hypocrisy: "My values didn't really match up. We did a lot of work for [an industrial equipment supplier]. The industry was unsavory in terms of the mining and the stuff they were doing environmentally. The job was eating my soul." During her second year, the company began working with a pharmaceutical company, which marked a pivotal moment for her:

> That was the final straw. There was a relatively benign condition that they were trying to treat. They asked us to do a prevalence and incidence kind of study to

> see what their market was going to be like X number of years out. And, they were basically going to treat themselves out of a market. They decided not to make the therapy. That was it for me. I left."

She contacted a client who was an external evaluator. "She said she had an opening and asked if I could come in the next week. So, boom. I took a massive pay cut. At that point, I knew the alternative. I knew what I needed to do to make the big money. And, I realized I can live off a whole lot less."

Enduring struggles

Sarah told me that, initially, her work in evaluation was a methodological practice. She said that, "I was carrying out the tasks rather than figuring out what should be done. I had some major frustrations in the beginning of that because I didn't feel proficient at being able to look at a question and be able to think of various ways to answer it." But, gradually over time, evaluation work came to be, "much more about the end result of what we're actually doing, which is help kids get a better education." She further reflected:

> I was relatively detached from it in the beginning. I was more interested in applying the methodologies. I didn't think about the programs a whole lot. It was a big shift in thinking. I also feel as though I'm in a place now where at least for the fulfillment of it, where I feel like I am helping to give a voice to the little guy, whereas before it was doing a lot of analysis of test scores and that wasn't an essential service.

She went on to explain:

> I started to feel much more proficient at the work when I started working in a much more evaluation-oriented situation whereby I could start looking at evaluation plans from the beginning. It made it easier to come out at the beginning and think about it in respect to the questions and a good way to answer the questions versus coming into it when it was already maybe somewhat done. I feel like I got more proficient with it as I had more freedom. I was the person in charge and had more responsibility.

Of particular concern to Sarah's self-story is her move from quantitative outcomes research to qualitative formative evaluation. She made sense of this and interwove her experiences with her values about social science in program evaluation. Part of this involved racing and classing program evaluation. From her perspective, largely quantitative research represents a "white, middle class field with a downward view of others in some situations. I think it worships at the altar of data, assuming that data are objective and assuming that objectivity is a desired value." Meanwhile, qualitative and formative evaluation represents, for her, speaking on the behalf of those with muted voices.

Sarah told me that the social scientific perspective is important to her in doing evaluation work. When I asked her to discuss that perspective, she said:

> I have a beef with social science trying to be hard science. We are imposing the science method onto this, but we get that it is not science like observable

> molecules. As soon as you start to make the assumption that this can be treated like any business or science model that is out there and that is more concrete, you lose a lot of the art. I think a lot of thought has gone into transposing the social scientific methodology onto these real applicable problems, which I also have issues with the kind of appropriation of program evaluation by big organizations or the government. It is wonderful that they want to evaluate programs. That is what people have been striving for all along. But, it gets simplified and mechanized. It loses the kind of purity that the academics do bring to it. They are so steeped in theory. Sometimes the theory itself gets lost in the mechanics, that then end up so far from where it began.

This passage indicates a major turn from the training she received as a psychology researcher. The practice of formative evaluation, combined with a feminist perspective on research, helped her revise her beliefs about the field and, therefore, her identity narrative.

When Sarah became an external formative program evaluator, one experience that surprised her and from which she grew, was the way in which clients and internal evaluators challenged findings. She said:

> It felt like maybe some people who should have stayed in academia and for some reason were not able to do so brought it over with them to evaluation because that's where they could work and flourish. It was surprising to me to see some of the signs of academia showing up in evaluation.

Sarah added:

> It helped me to internalize the need to be able to defend my approach and my take on something. Before that I didn't understand the need to do so, I was just like, "That's what I'm doing, I can't really articulate why." I tend to do something intuitively and kind of in a way that I don't necessarily, can't articulate why until I've thought it through, and so it helped me to learn the value of that and also be able to express it in a way that still felt like I had integrity but also wasn't going to be just knocked down.

One implication of a trend in which clients begin going inside and, to an extent, becoming partners, is that Sarah could not operate as a simple and automatically respected operant of social scientific methodology. She would be required to defend the approaches and findings, as well as make very clear the relevance and usefulness of findings to particular program facets to a wide range of program staff, program recipients and targets.

Sarah represents a case of a professional evaluator who certainly has her own agenda and certainly acts to refine the meaning of program evaluation. She does not, however, go out of her way, particularly, to contribute to codification of these through the induction of new program evaluators. And, in juxtaposition of the next vignette, does not strive to publish the results of her work. Instead, she pours vast amounts of toil into her day-to-day work as a professional program evaluator.

Category 3: The Post-Academic

Connection to program evaluation

At the time of the interviews, Helen had been working as a program evaluator for just over a decade. After having one professional position at a non-profit research and development company for that entire period, in which she rose through the ranks from research assistant to department manager, she had just accepted a position at a competing for-profit company. The interview occurred, therefore, at a major transitional moment for her.

Helen had several roles. She was associate director of a regional center that provides technical assistance to state departments of education, was co-director of the company's research and evaluation unit ("which has a lot to do with making sure people have professional development and the resources they need, helping to develop the strategic plan with other units in the organization, and making sure managers get the support they need"), and was a key members of her company's corporate development team ("which means I respond to RFPs, write proposals and do outreach, which means I contact clients and ask what they need and how we can help").

> Helen told me that, at the time of the interviews, she actually devoted about 10 percent of her time to conducting external program evaluations. In the recent past, however, nearly 100% of her time was devoted to doing evaluations. Like Sarah, Helen did not refer to herself as an evaluator. She made it very clear that while she <u>does</u> evaluation, she <u>is</u> a sociologist. Like the other two social scientists described in this chapter, Helen said she struggles with articulating her roles:

> I say I do education research and evaluation. Sometimes I'll add that I was trained as a sociologist and if I'm feeling really chatty, I'll confess that I really didn't ever mean to be in education or evaluation. Sometimes if I'm at a conference where it's technical assistance that's the focus, I'll say something like, "I'm a cross-dresser," make light of the fact that I play these two different roles.

Thus, again, the theme of situational identity emerges. For Helen, however, the two facets of her professional identity are more or less discrete and less moving than were the roles described in the other two vignettes.

Helen described herself as having two distinct sociology faces—one in which she conducts contract program evaluations and another, outside of work, that contributes to academically-oriented sociological literature. Helen juxtaposed her two worlds:

> The education research I do is outside of work. Evaluation is pragmatic, delimited by what the client wants and by what the funders require in terms of accountability. Evaluation is not about major social change or advocating for radical reform; it is incremental. It is more about accountability. So, I understand it in terms of its delimited role that it can play. Clients, in addition to accountability, have questions that they want to ask of their programs or of their data.

When I asked her how she related her work as a program evaluator to sociology, she described a division based on a conscious decision. That decision was based, in large part, on several experiences with coworkers and clients that rejected or felt threatened by

her use of theory in contract work. Thereafter, she offered her social scientific perspective sparingly. Specifically, she said:

> Part of me has made this deal with myself that I am not going to bring much of theory to evaluation work because it frustrates me. I am a frustrated researcher. And, so much of the program evaluation we do is prescribed. You have to answer these questions. You have to fill out this boring report form for the feds. Sometimes I am able to help projects that think about who they are serving and what characteristics of their client base having implications for what they offer. I enjoy logic modeling and helping clients do that. And, sometimes I can bring in something that I know of from sociology to their thinking. If you do x, is y really going to happen? And, helping them think about the dynamics there. The theoretical stuff gives me a meta-analytic view of what I am doing because I can put it in bigger context. I feel like a craftsperson and then sometimes a technocrat.

Helen further exemplifies the move beyond a logical-operational sequence of methodological events to work that has consultative and creative twists. But, unlike the other two vignettes represented in this chapter, Helen's post-academic work gives her a sense of remaining true to her academic discipline. On the one hand, she helps maintain (and, at the same time, expand) sociology of education while on the other hand, she brings to evaluation a sociological perspective.

When I asked her to give me an example of how her contract work differed from her work in sociologically-oriented education research, Helen said:

> The research stuff feels to me that it is more about understanding dynamics. The questions are not tied to programs. They are more engaged with a literature, theoretical or empirical or both. It makes contributions to larger conversations. The thing we have out for review now is case studies of six rural school districts in [Midwestern state]. We went back and looked at the data to explore how, in these various diverse rural districts, the educators see their role as saving the poor from themselves. Enculturating them in the middle class. A lot of social caretaking. Teaching them to have good table manners. The things that educators believe are important for kids to achieve middle class status.

As she described her work, she also compared the road she took with academia. She recalled:

> I was thinking about this the other day. I was comparing my career with those of people from my doctorate program. Almost all of them are at universities now. And, I am the only one who is not. I am pretty sure I make more money than they do. I feel sort of opportunistic. Yet, when I think about other organizations out there that provide the kinds of services that I do, it may as well be me. I am not evil. I am not a money grubber. I do good work. I can sometimes find opportunities to say things that need to be said even if they are not popular or wanted. I feel more corporate than I ever intended to be, but (long pause) I am

> not hooking on the corners. There is a lot of good stuff associated with this. I feel pretty empowered about my career. But, I am clearly a little ambivalent.

When she said that, I asked her to elaborate on that ambivalence. She added:

> I wish I could ask my own questions. I wish I could have that intellectual freedom that people in universities have. But, I like knowing that the work that I do sometimes has immediate, practical application. That its meeting somebody's needs. That it gets read. So, it's a mixed bag.

For Helen, an unavoidable tension is wedged between prefigured contract work and the addition of her post-academic participation in the literature. Like Kathy, she maintains a vital connection to the "other" world, but unlike Kathy, who "keeps it real" by continuing to do contract work that benefits students, Helen opts to do scholarly work that does not tie to or derive from her contract work. In fact, from an institutional perspective, it is almost completely unrewarded.

Path into program evaluation

Like the other program evaluators described, Helen did not plan on becoming a program evaluator when she was a graduate student. She had to learn much of it on-the-job. She commented:

> I had to learn about all the different layers, requirements, competing priorities and institutions because I didn't have a map of that in my head, and without that, I didn't know how to think about what the point of evaluation was. Who was it serving? What was the point of doing it? What change could you hope to effect?

> When I look back, I was awfully naïve. I'd never read any program evaluation theory or practice. I had to work on all these new theorists and schools of thought and kinds of evaluation.

She approached the fringes of, and soon after entered, the figured world with the lens of an academic.

Helen, who depicted her parents as "Ivy League graduates from the Northeast who came of age in the 1960s and decided to live what they believed by moving to [a rural part of a NE state] to farm and teach," grew up in a utopian-age household. Both her parents worked in their respective social science careers. Helen reflected:

> Pretty much everything we ate my folks had grown and made themselves. We didn't have TV, and we had all kinds of what I later learned to think of as "cultural capital"—a piano, a guitar, book shelves full of books, a huge music collection. Granddad had an enormous old home, art work everywhere, and ran his own company. Mom said he was the classic case of economic determinism, in the sense that before he owned his own business he was literally a card-carrying communist, and afterwards, he started voting Republican. I didn't know what it was until later, but there's also strong tradition in her side of the family of Jewish intellectualism which colored a lot of family discussions, interests.

She added:

> It wasn't until I was older that I understood that they were making principled choices, but that we wouldn't ever have been in real danger of starving or freezing because my grandparents could and would have helped if things got desperate.

Helen told me that her family's utopian lifestyle made her feel different from others. She recalled, "I couldn't relate to their conversations about TV shows; I didn't eat the same kind of food; and I came to understand that my family's disbelief in god was an enormous no-no." This social distancing eventually, according to her identity narrative, primed her for being a social scientist, as the quote below, which I have recycled from a previous section, reflects:

> It took a very long time, but eventually those experiences led me to see myself as an observer, an outsider on the margins. This was both enabling and constraining (nod to Anthony Giddens), enabling because it ultimately led me to social science and constraining because I lived a lot of my early adulthood not participating in my life so much as observing it.

Thus, as was the case with Kathy and Sarah, upbringing and temperament prepared her in an organic way for work as a social researcher. Helen wrote in an email how the idea of social justice began to form early on for her. She recalled:

> By 2nd grade, I was reading voraciously. Reading Roots was one of those pivotal points, because not only was the book totally absorbing (I remember going outside and walking around and around the house reading that book.), but it's when I started to tune into the idea that there were really big and bad things that

> happened to people. I wouldn't have called it this, but it was also when I started to think about inequity and social justice.

This naturalization of liberal social science was neatly interwoven with her notion of social justice.

Helen entered higher education at an early age, after bypassing high school altogether. She said:

> I'd been taking college courses in the summers starting after my 7th grade year, so I knew what to expect. After 9th grade, I decided, with my parents' encouragement, that I wanted to go to college full time. Mom helped [college] establish an early entry program. It was a protected experience—mom was always on campus, we lived three blocks away.

After a year in college, she transferred to a state university that was in state, where she discovered and "fell in love with sociology." She added, "It was the first time that I got into thinking about politics, context, and identity." This helped her forge a sense of activism as she became involved in two "liberal campus activist groups."

> One was the campus gay and lesbian rights group, which I helped start up with a friend. It was uncomfortable because I wasn't a lesbian and wanted everyone to know it, yet there I was proclaiming how important it was to be open minded toward gay people. I learned something about identity, hypocrisy, and what the world looks and feels like to people who are marginalized.

After college, Helen got married, had children, and then decided to return to graduate school. She recalled:

> When I was thinking about graduate school, for a while there was struggling between social work and sociology. I was leaning towards social work because it wasn't going to have a statistics requirement, which in retrospect is really stupid. I can remember talking to my folks about it and them just rolling their eyes and saying "give me a break. It might be hard but you can do this. You got through algebra."

Helen said, she went back to school to "get a doctorate and become a professor. I wanted a group of like-minded peers. I also wanted to learn new stuff, and think about the world." Her graduate school selection process was similar, in some ways, to that of Kathy. She said, "I wanted a program that was supportive of its grad students, that wasn't hung up on one methodology or perspective, and that focused on things I was interested in at the time (social movements, race, gender)."

A northeastern liberal arts college "started courting me hard. They offered me a good assistantship and a generous book allowance." Returning to the "natural" fit with sociology, Helen said:

> Studying sociology gave me all these words and ideas that made me think about all the weird things in my life around me. When I found sociology and ethnography, it was like "ooh." Temperamentally, it was a good fit and it made me think about what it meant to be on the inside versus the outside of groups, to

> have certain kinds of advantages, to not have certain advantages. We learned quickly that in our county while we might have all kinds of cultural capital, it didn't mean shit because we didn't have the social capital, the networks. We didn't have any of that. It helped me put things in broader perspective.

Thus, her advanced degree helped her explore more of her own identity while granting her license to do the kind of liberatory work she idealized.

Helen entered the world of external program evaluation after receiving her MA in sociology. She began her PhD program and then took a leave of absence because she believed:

> I wasn't ready. I was too young and naïve. I felt that if you want to do sociology responsibly you need to know the world better. There was this job announcement for a research assistant position. I was completely overwhelmed because I had the research skills, but I didn't know evaluation and I sure didn't know education.

As she said earlier, Helen scrambled to self-teach about evaluation and education. She recalled:

> It was a great job and I would be able to use what I learned. I would say there was two years where I had to absorb a whole lot about evaluation and about education. Then, I rose through the ranks. I think I was a specialist when I started getting ancy and started trying to figure out what to do next.

After working in the contract world for a few years, she reconsidered her PhD program. She had reached the ceiling in her field. Her dissertation chair invited her to return to the program and complete it long distance.

> It was interesting being a graduate student then because I didn't feel like an idiot quite so much. I had work experience and that gives you a different perspective. I had a good time reading stuff and playing with data. The doctorate furthered my career to the extent that it made me more promotable at [my company] and it put me back into sociology.

Thus, her degree elevated her understanding of the field, which both allowed her to move up at work and improved her chances of publishing her post-academic sociological solo work. But, it also drove the wedge between theory and practice a bit deeper.

Enduring struggles

Evaluation provided Helen a connection to "something I cared about, so I was willing to put up with the discomfort of surprises." It took her three years to come to the realization that she could and might make a career out of contract program evaluation. Encouragement from coworkers and a rapid rising up through the ranks helped secure her position and sense of proficiency in the figured world. Nonetheless, she also described a few enduring struggles.

Helen described a perennial issue in her professional life with trying to reconcile the two facets of her work. She described to me one of the moments that helped define

the need to keep her sociological work separate from her work with clients. The project included interviews with rural principals who provided "heavy duty critiques." When the report was sent to the state department of education, the project officer said, "You can't say these things." Helen recalls her feeling:

> That's the moment when I thought, "They do have the power here and I don't like that. I felt like we were compromising the perspective that the interviewees had shared with us, that we were not holding fast to some idealized, pure reporting. It signaled for me that this kind of work really has constraints around it. Particular constraints having to do with who pays you and what authority they have over you.

She said that as a result of that experience:

> I know my place. I look for intellectual engagements around things that might be controversial elsewhere, outside of work. It made me attend more carefully to how I said things. I liked being able to say things that maybe a client doesn't want to hear in such a way that they have to hear it.

Tied to her bifurcated professional roles, of great concern to Helen is that "evaluation is under-theorized." Similar to the perspective shared by Kathy, Helen said, "People come marching in with their methods and tools, but don't have a good understanding of what evaluation is, what role it plays, how political it is, how it works." She referred to these evaluators as "naïve empiricists" who "listen to the data and they think they are bringing truths that need to be listened to. They don't problematize what

they're doing." This may be an attempt to reconnect a seemingly-unnatural split. But, unlike Kathy, whose academic position permits her to codify in an official capacity the connection between theory and practice (although she does not do it in her contract work), Helen's corporate position provides only tiny space venues for advocating the use of theory in her everyday work.

Helen also described to me a fundamental issue with "bad evaluators" and a controversy in program evaluation about whether or not evaluators ought to receive some form of credentialing. Specifically, she said:

> I'm not convinced that we need more certification, but we do a lot of salvage work where a client has been burned by an evaluator who didn't follow through or didn't do the right kind of analysis. It bothers me to see irresponsible evaluators. There is debate in the field about credentialing. That we need to have more programs specifically for evaluation. Maybe AEA needs to play more of a role like AMA where they police the discipline.

Again, her ambivalence came out.

Helen represents one adaptive response to privatization in higher education and some of the lingering, if not artifactual, values, such as publishing, that correspond to the academic world to which she felt most naturally attached. The role models of knowledge production, beginning with her parents, provided a blueprint of expectations that stated in no uncertain terms, "To be a true sociologist, one must publish and remain connected to the social science." Helen worked as a teacher of sorts, for clients and by conducting

Socratic professional development on the conduct of program evaluation and the interpretation of findings. She was connected to community work as a student mentor, did oversight for interns, has participated on doctoral committees, and did volunteer fundraising for local community groups. Thus, one the one hand, what she described as her normal way of operating emulates her understanding of the academic world, at least more than it does the professional evaluator. At the same time, though, she symbolizes an adaptation of changing university conditions that takes knowledge production into a more Socratic realm.

Concluding Thoughts

The vignettes I offer show some common themes related to social justice, planned and incremental social change, and use of ERD and social science research to bring about or contribute to this change. And, all of the cases reveal an awareness of social position privilege and the ways in which that privilege was translated into contributing tidbits that gave rise to being reflective practitioners. What differed from vignette to vignette, however, were the ways each adapted herself to shifting political economic pressures, how each adjusted her identity narrative to fit with those specific contexts, and the way each made sense of enduring struggles in the figured world of program evaluation.

Concluding Remarks, Discussion, and Further Study

In this final section, I offer a few brief reflections on the entire study, discuss broad implications, and point out areas for possible further research. I treat this not so much as a summary of conclusions or further refinement of the suggestive and emergent model pertaining to the cultural production of the modern program evaluator in education. Instead, it is intended to fray a few threads and allow some of my personal reflections to seep onto the page more translucently.

I have attempted to cover a lot ground in this study while maintaining the emphasis on cultural production as a process. Broadly, I have attempted to describe how people choose to enter the Education Reform Discourse and contribute to both its embodiment and refinement. From a different angle, I have begun to demonstrate how the ERD becomes a tool for activist-minded social scientists that enables them to revise their identity narratives until they are consistent with the figured world of program evaluation. As it turns out, evaluators do not represent the same class conditions upbringing. They are a little more diverse than are their academic counterparts. And, they reside in a betwixt space that balances naturalized quests for understanding human nature with sequentially-oriented planned change.

Throughout this study, I have attempted to avoid the temptation of sharing too much of my own thinking. One critique of a draft of this manuscript pointed out that I neatly side-stepped any reference to the less-than-wonderful program evaluators including the "hacks, money-grubbers, fly-by-nighters who have less scruples and quality-orientation than do these folks." My immediate response to this critique was that

it was not at all my purpose to evaluate the quality of work or the work ethics of program evaluators. In fact, doing so would complicate the way I describe the processes by which they self-author themselves into program evaluation. Having said that, the evaluators with whom I spoke provided numerous examples of ethical challenges they confronted first-hand.

I did not seek out exemplary evaluators. On the contrary, I tapped into a social network of professionals I knew, first. Then, from them, I branched out. Not everyone I spoke with has a universally-wonderful reputation. Others made mistakes, as they pointed out in their identity narratives, from which they reflected and altered their career paths. It is an imperfect world. Participants of this study also described having worked with or having cleaned up after misleading or "hack" evaluators. It is, from the perspective of clients, an open market. Anyone with a Ph.D. is welcome to submit a proposal for an external evaluation. It has been my experience that notoriously dreadful evaluators do not last in a particular market for long.

To put this into perspective, let us consider Austin's pool of education-focused external contract evaluators. At the time of this writing, a former regional education laboratory employs approximately 12 full-time evaluators. Furthermore, the National Center for Educational Achievement has perhaps a dozen evaluators on staff. Add to these the 12 or so total program evaluators that staff Austin's two non-profit organizations focused on external educational evaluation, the 15 or so hired guns in Austin, and, of course, The University of Texas at Austin and other university centers and independent faculty who conduct evaluation and the total rises to a reasonable estimate of

50-100 external evaluators focused on education in Texas' state capital. Such a small pool of evaluators that compete for the same contracts has implications for networking. First, evaluators who are unhappy in one organization often move to others in the same area. For obvious reasons, this can only happen so many times. Second, it means that, especially for larger, multi-part contracts, organizations rely on subcontractors—either individuals or companies. And, third, the major client, the TEA, has staff that are savvy about program evaluation quality. None of these guarantee high quality and, without any credentialing process for evaluators—other than doctorate degrees—anyone with a degree is entitled to apply to do the work. Of the external evaluators in Austin who work on education-related projects, a small handful have made names for themselves as being a "little late on delivery" or "sloppy with data."

Thus, the microcosm of program evaluators seems to regulate and maintain its relatively high quality of research services. It operates as a semi-closed professional community. Furthermore, my exploration of the figured world of program evaluation reveals a diversity of approaches, passionate creativity, and more than a modicum of agency. Or, does it? While the identity narratives and accounts of everyday work illustrate an adaptation of applied and "practical" social scientific research to address real world problems with lots of examples of authority and autonomy, the ever-present client and the enduring bond to that client, the contract, are linkages to a set of conundrums at the crossroads of agency and obligation.

In the opening chapter, when I described some of the pitfalls of an evaluation of a newly-installed advanced placement program, I reserved a few of the deepest cracks for

this postscript. At the conclusion of that evaluation study, the program evaluators submitted a final report with recommendations to the state department of education. A change in the political balance of the state put the GEAR UP program out of favor. The findings the state department of education submitted to the U.S. Department of Education, in greatly reduced form via the ED 524B, omitted reference to deficits in the program while highlighting the intractability of the target population. Soon afterwards, the program was ended.

What I would like to point out here is that the macropolitical realm is not simply a contextual feature. It is the muscle that keeps the neoliberalized ERD machine moving. The above example shows one way in which this muscle is capable of limiting the dissemination of particular evaluation knowledge. Another, more insidious, macropolitical tactic is to not only own the proprietary knowledge of the evaluation, but to also colonize the knowledge producers. Throughout this study, I have depicted, in a way that might appear somewhat as a vindication of evaluators, a self-monitoring system that keeps itself morally sound by eschewing any attachments to particular political agendas. The methodological approach of talking with people about their reflections and memories of how they came to do what they do may, without some closing notes, appear as a naïve accounting with reference to the macropolitical realm, however. In fact, knowingly or not, just as in the case of pharmaceutical companies, some program evaluators do succumb to pressures to produce findings that will support particular programs.

Furthermore, while the approach provided a powerful entrée into the process of becoming an evaluator, it was difficult to reconcile the contradiction between the appearance of agentic control and the need to adapt, without complaint, to privatized production of knowledge. In this project, I characterized the emergence of two responses to the expanding neoliberal realm. The types also represent a bridging of theory and action by mixing up traditionally-discrete realms. Some evaluators working in academic settings are bringing in money from a variety of sources, as the market itself, of no-strings-attached wells dry up—a trend that began with the move of social scientific research being affiliated primarily with museums, hospitals, university centers (such as those at the University of Chicago during the School Survey Movement), and other institutions to the emergence of funding for individual faculty funding (e.g., field-initiated research). Meanwhile, post-academics respond to a historically-indexed moment in which intellectual freedom became an expectation. They may represent a forefront of an emergent shift in funding, focus, and, ultimately, the purpose of social research in education, most especially in terms of program development.

This study also underscores how some of the thinking linked to the ERD, especially newly-packaged claims regarding equity and excellence, are neither new nor addressed in drastically innovative ways. They were not new when the 1965 mandate was created. They were not new during the 1940s or 1950s when data were being amassed to compare spending and staff, resources, and infrastructures to curricular quality. As I noted in Chapter 2, the call to use data to inform steps toward equity and excellence preceded the School Survey Movement and, perhaps, Reconstruction. I do not

claim that no change has occurred. Indeed, along the way are who decides how this will be done and by what means. And, I attempted to interweave compelling evidence that suggests social capital ala social networks and socially-positioned gatekeepers helps determine, just as it would in any other figured world, the requirements for entry into the figured world and, therefore, help monitor and maintain its frontiers.

Implications of this study

Maybe "implications" is not the best word to describe what I want to do with this section, since I am not offering a panacea solution that will make the process of becoming an effective or "good" evaluation more efficient. This study has humbled me to the usefulness of NCLB, a law that, when first introduced and enacted, seemed to promise an unavoidable reduction in my role as a mostly qualitative formative evaluator. In fact, however, my work expanded tremendously.

Over the decade since the initial rollout of that dreadful piece of legislation, I saw a lot of works that lamented the loss of research designs that did not fall under the auspices of NRC's legion of gold and silver standards. And, aside from conservative pieces that applauded the creation of standards, that no one seemed able to achieve (e.g., AYP's rigid and seemingly-commonsensical connection to teacher quality and funding), I saw little that considered benefits of the legislature. The evaluators in this study helped me acknowledge and appreciate an unintended effect of NCLB, namely the expansion of program evaluation into schools, district offices, and other sites of the educational apparatus. There was a certain movement from evaluation done <u>to</u> targets of reform to being done <u>with</u> those targets. Evaluators needed to defend their findings, discuss how

those findings might influence practice, and helped clients collect and analyze their own data.

For further study

Some of the false starts for sections of drafts of this study will be touched upon here. As I pored over interviews and extant documents, especially Abert's trilogy memoir that described, in great detail, the emergence of program evaluation from an insider perspective at the HEW (1979), it was difficult to remain entirely focused on the matter at hand. This realization may reflect the methodological approach I selected for the study. Oral history combined with personal reflections certainly helped me lay out the territory and understand how people see their professional lives unfold. But, it did not allow for much assembly of action with narrative. Thus, one particular follow up might include a participant observation approach in a localized setting of program evaluation work. I believe such a study holds some promise for identifying and exploring more deeply the spaces between what people say they do and what they appear to actually be doing. Such a critical ethnography has the potential to contribute to the critical literature on how program evaluators position themselves and are positioned in everyday work and may, therefore, help define the relationships between identity production, the macropolitical realm, and the ERD.

Another area in need of further exploration, and from a critical lens, is the social position of program evaluators. The lack of diversity among program evaluators seemed to be an issue from my limited first-hand experience. The participants, however, especially those who have worked in academic settings, helped me understand that, in

comparison to academia, while problematic, especially in the practice of interpreting data that affects programming intended to enhance the educational opportunities for minority and economically disadvantaged students, the cleavage is not as big. In fact, the shortage of minority evaluators might speak to a shortage of practitioner oriented social scientists completing graduate programs.

Of course, another key facet to the evaluation story is gender. Evaluation in gendered in favor of white women, which is a shift from what it was 20 years ago. Most participants of this study explained this by pointing to evaluation's "helping" nature. This explanation is not sufficient, in my mind, at least. It seems to me that part of the issue is tied to a saturation of positions and an over-representation of men in university settings along with a huge increase in women earning advanced degrees in social sciences. This, as one might imagine, could be its own separate and additional exploration.

Additional work could also be done on the emerging typology. Specific questions might include what are the implications of applied programs and practitioner-oriented departments in academia? Is this trend indicative of a further withering of social science academic work? Or, have social sciences grown sufficiently in numbers, scale, and voice that there is room for multiple specializations? Clearly, there are different ways of conceptualizing expansion into new areas (e.g., creation of the Cultural Studies in Education program at The University of Texas at Austin) that infuse critical social science thinking into traditionally less critical realms such as education.

Appendices

Interview Protocols

Personal and Professional History

Tell me about your childhood. Give me a broad description of yourself from your earliest recollections through grade school. Also tell me about things that really captured your interest, especially bothered you, and any skills or talents you recall.

Tell me about your family. What activities were especially important for your family? What values were most important to your family? Please provide a few examples. What was most important to you in terms of values and activities? How about high school days?

Tell me about your postsecondary education. Did you go straight into college? Tell me about your college selection process and how you chose the academic path you took. What motivated you to study what you did? At the time, what did you hope to get out of your college experiences? What values or ideals were most important to you during your years in college?

Tell me about graduate school. Why did you choose to study what you did? What did you hope to get out of your university experience? Compare yourself to others in your cohort or program. What values or ideals were most important to you during those years?

Tell me about your professional life. How would you describe your current role? What is your current title? Tell me about the path to your current position in relation to graduate school. If possible, provide a copy of your CV or resume. What inspired you to consider program evaluation? How does it relate to your social science education?

Was there any relationship between your academic choices and opportunities and your social position (gender, ethnicity, class)?

Was there any relationship between your career choices and opportunities affected and your social position (gender, ethnicity, class)?

Tell me some shortcomings or gaps in program evaluation. How might they be addressed?

What would you tell a budding social scientist who was considering program evaluation as a career path?

Questions for Everyday Life Interviews

Identity Narrative

- What is "evaluation?" How does the work of evaluators differ from that of academics? Tell me what being an evaluator means to you personally (want them to talk about ways evaluation defines who they are).
- Tell me how you were incorporated into the field. What do you need to learn to be an effective evaluator?
- Did you encounter any struggles? Was there ever a time you felt or were expected to feel that you made a mistake in becoming an evaluator? How were these struggles resolved?
- What about the industry surprised you the most when you first started doing evaluations? Tell me about those. How did you react?
- Have you had any major shifts in thinking about your role as an evaluator over time? Have you had any major revisions to the way you think about yourself as a social scientist? What prompted those changes?

Projects

- Give me an example of a specific contract evaluation you have worked on. Please describe its lifecycle from its conceptualization to its finish.
- How was the contract identified?
- How were decisions about method, theory, staffing, budget, and reporting made?
- Were there any tense moments or conflicts? How were they managed?
- Talk to me about sample selection. Tell me about instrumentation.
- Tell me about data collection, handling, and analysis. Were there any conflicting findings? How was that situation handled?
- How about communication with stakeholders. Who could communicate with them?
- What tasks are involved in program evaluation?
- Who did the work? Tell me about the work team. Were there any subcontractors involved? Tell me about that.
- How were tasks managed?
- Tell me about the report writing process. Who determined how that would be done? What were some of the rules for doing that? What is the process? Tell me about sanitization.
- Who interacted with clients?
- What other kinds of communications were there?

Questions about the Evaluation "Field"

- What major changes in evaluation have occurred in the course of your career? What brought them about?
- What kinds of organizations are involved in evaluations? What makes them different?
- How are new evaluators incorporated into the field? What do they need to learn?

Human Resources and Management

Details about the firm or organization

- How does your organization market its services?
- What are the major considerations for prospective projects?
- What are non-negotiable rules for the conduct of evaluations?
- Tell me about how time gets allocated.
- In evaluation work in your organization, what gets rewarded?
- What gets sanctioned?

Staffing

- What are the most important attributes of an evaluator?
- What is the process for new staff hires? What do you look for? (get job descriptions for this)
- How are new evaluators incorporated into the field? What do they need to learn?
- How does an evaluator move? What is the professional trajectory of an evaluator?
- How are new evaluators incorporated into the field? What do they need to learn?

Questions about Evaluation "Field"

- What major changes in evaluation have occurred in the course of your career?
- What brought them about?
- What makes for a good year in an evaluation organization?
- What other kinds of organizations, besides yours, are involved in evaluations? Tell me about the types of organizations that do external contract evaluations. What makes them different?
- How are new evaluators incorporated into the field? What do they need to learn

Case Study Interview Protocol

The purpose of this interview was to provide in-depth illustrations of the themes that were generated from the first two interviews. Therefore, I developed each follow up case study interview uniquely.

Part I. Follow Up Questions

[The purpose of this portion is to go a little deeper on general themes that might have come up during the first two interviews]

Sample question: In the second interview, you mentioned that it was difficult for you to associate research services with a contract orientation (focused on clients needs and given a set dollar amount and for a pre-specified period). You told me that you eventually figured it out and came to accept it, but I was wondering if you could help me understanding how your thinking around that change happened.

Part II. Turning Points

[In this portion of the interview, I pick up on particular turning points identified in the first two interviews. The focus will be on recurring themes (major theoretical facets of the identity production process) such as reference to internalized voices, adherence to the reform discourse, perceptions of the relationships between career focus and social position]

Sample question: In the first interview, you told me that your parents were disappointed when you told them you wanted to study sociology in graduate school. In what ways did that affect you? Did that perception change over time? If the voice of your parents was to steer you away from that field, whose voice(s) steered you toward it?

Part III. Professional Future

Tell me what you would like to do with your career long-term.
What will you do to get there?
Is there anything you might have to unlearn or relearn? How will you do that?

Bibliography

Abert, J. G. (1979). Program Evaluation at HEW: Research Versus Reality. Marcel Dekker Inc: New York.

Abert, J. G. & Kamrass, M. (1974). Social Experiments and Social Program Evaluation: Proceedings of the Washington Operations Research Council Symposium. Ballinger: Cambridge.

Adelman, C. (Ed.) (1984). The Politics and Ethics of Evaluation. London: Croom Helm.

Althusser, L. (1970). *Ideology and Ideological State Apparatuses* in Lenin and Philosophy and other Essays (1971).

Agar, M. (1986). The Professional Stranger: An Informal Introduction to Ethnography. Academic Press: Berkeley.

Apple, M. (1978). Ideology, Reproduction, and Educational Reform. Comparative Education Review, 22(3), 367-387.

Apple, M. (1999). *The Absent Presence of Race in Educational Reform.* Race, Ethnicity and Education, 2(1), 9-16.

Apple, M. (2001). *Comparing Neo-Liberal Projects and Inequality in Education*. Comparative Education, 37(4), 409-423.

Apple, M. (2004). Ideology and Curriculum. New York: Routledge.

Ayres, L. P. (1917). The Cleveland School Survey, Summary Volume. Survey Committee of the Cleveland Foundation: Cleveland.

Hirst, P. & Zeitlin, J. (1991). *Flexible Specialization Versus Post-Fordism: Theory, Evidence and Policy Implications*. Economy & Society, v. 20, no. 1.

Bakhtin, M. (1981). The Dialogic Imagination . University of Texas Press.

Bell, S. (1997). *Crafting a Non-Partisan Evaluation in a Partisan World: The Urban Institute New Federalism Evaluation*. Paper presented at the American Evaluation Association Conference, San Diego, November 1997.

Boas, F. (1911). The Mind of Primitive Man. New York: Macmillan.

Boeing financial report (2006). http://www.boeing.com/companyoffices/financial/finreports/annual/06annualreport/index.html. Accessed April 2, 2007.

Bogdan, R. & Biklen, S. (1992). Qualitative Research for Education: An Introduction to Theory and Methods. Boston: Allyn-Bacon.

Bourdieu, P. (1977). Outline of a Theory of Practice. Cambridge.

Bourdieu, P. & Wacquant, L.. 1992. An Invitation to Reflexive Sociology. Chicago: Chicago University Press.

Brewer, J. D. (2000). Ethnography. Buckingham, England: Open University Press.

Buchanan, G. & Wholey, J. (1974). *Overview of Evaluation at the Federal* Level. In Abert, J. G. & Kamrass, M. (Eds). Social Experiments and Social Program Evaluation: Proceedings of the Washington Operations Research Council Symposium. Ballinger: Cambridge.

Calhoun, C. (1994). Social Theory and the Politics of Identity. Blackwell: Malden: MA.

Callahan, R. (1962). Education and the Cult of Efficiency. Chicago: University of Chicago Press.

Campbell, D. & Stanley, J. (1963). Experimental and Quasi-Experimental Designs for Research. Chicago: Rand-McNally.

Carroll, T. G. (1992). *The Role of Anthropologists in Restructuring Schools*. In Hess, G.A., Jr. Empowering Teachers and Parents: School Restructuring Through the Eyes of Anthropologists. Bergin & Garvey: Westport.

Castells, M. (1997) The Power of Identity: The Information Age: Economy, Society, and Culture. Blackwell: Malden: MA.

Caswell, H. L. (1929). *Is the School Survey Movement Dead*? Peabody Journal of Education. 7(2). Sept., 1929, 108-114.

Caswell, H. L. (1932). *Definition and Classification of Surveys*. Peabody Journal of Education. 10(3). Nov., 1932, 179-181.

de Certeau, M. (1984). The Practice of Everyday Life. IC Press: Berkeley.

Chelimsky, E. (1991). *On the Social Science Contribution to Government Decision-Making*. Science. Oct. Vol 254.

Chelimsky, E. (1987). *The Politics of Program Evaluation*. Society. Nov/Dec.

Clifford, J. (1988). The Predicament of Culture: Twentieth-Century Ethnography, Literature, and Art. Cambridge: Harvard University Press, 1988.

Clifford, J, & Marcus, G., eds. (1986). Writing Culture: The Poetics and Politics of Ethnography. Berkeley: University of California Press.

Coleman, J. (1966). Equality of Educational Opportunity. Arno Press.

Davies, C. A. (1999). Reflexive Ethnography: A Guide to Researching Selves and Others. New York: Routledge.

Datta, L-E. (2000). *Seriously seeking fairness: strategies for crafting non-partisan evaluations in a partisan world*. The American Journal of Evaluation. Vol 21, No. 1, Winter 2000.

Della Porta, D. & Diani, M. (2006). Social Movements, An Introduction.

Dey, I (1993). Qualitative Data Analysis: A User-Friendly Guide for Social Scientists. New York: Routledge.

Dole, C. & Csordas, T. J. (2003). *Trials of Navajo Youth: Identity, Healing, and the Struggle for Maturity*, Ethos, Vol. 31, No. 3, pp. 357-384

Donaldson, S.I. (2001). *Overcoming Our Negative Reputation: Evaluation Becomes Known as a Helping Profession*. American Journal of Evaluation, 22, 355-361.

Eddy, E. (1976). *Future Directions for Educational Anthropology*. Council on Anthropology and Education Quarterly (7)1. 29-32.

Eisenhart, M. (2000). The Fax, the Jazz Player and the Self-Storyteller. In B. Levinson (Ed). Schooling the Symbolic Animal : Social and Cultural Dimensions of Education. Rowman & Littlefield Publishers: Lanham, MD.

Fabian, J. (1983). Time and the Other: How Anthropology Makes its Object. New York: Columbia University Press

Fairclough, N. (1995). Critical Discourse Analysis. Boston: Addison Wesley.

Foley, D. E. (1995). The Heartland Chronicles. University of Pennsylvania: Philadelphia

Foucault, M. (1972). Archaeology of Knowledge. New York: Pantheon.

Foucault, M. (1977). Discipline and Punish: The Birth of the Prison. Vintage: New York.

Foucault, M. (1979). The History of Sexuality. Volume 1: An Introduction. London: Penguin Books

Freeman, H. (1974). *Evaluation Research: A Growth Industry*. In Abert, J. G. & Kamrass, M. (Eds). Social Experiments and Social Program Evaluation: Proceedings of the Washington Operations Research Council Symposium. Ballinger: Cambridge.

Glaser, B. & Strauss, A. (1967). The Discovery of Grounded Theory. Chicago: Aldine.

Gray, W. S. (1917). The Teaching of Reading and Writing: An International Survey. Bloomington, IL: Public School Publishing.

Guba, E. G., & Lincoln, Y. S. (1981). Effective Evaluation: Improving the Usefulness of Evaluation Results through Responsive and Naturalistic Approaches. San Francisco: Jossey-Bass.

Guba, E. G., & Lincoln, Y. S. (1989). Fourth Generation Evaluation. Newbury Park, CA: Sage.

Gramsci, A. (1971). Selections from the Prison Notebooks. New York: International Publishers.

Hammersley, M. & Atkinson, P (1995). Ethnography: Principles in Practice. New York: Routledge.

Hanus, P. H. (1920). School Administration and School Reports. Houghton Mifflin: Boston.

Harvey, David. (2005). A Brief History of Neoliberalism. Oxford: Oxford Press.

Harvey, D. (1990). The Condition of Postmodernity: An Enquiry into the Origins of Culture Change. Cambridge, MA: Blackwell, 1990.

Henry, G. (2001). *How Modern Democracies Are Shaping Evaluation and the Emerging Challenges for Evaluation*. American Journal of Evaluation Sept. 2001 22: 419-429,

Hodgson, G. (1978). America in Our Time. New York: Vintage.

Holland, D.C. & Lave, J. (2001) *History-in-Person: An Introduction*. In Holland, D.C. & Lave, J. (Eds.) History-in-Person, Enduring Struggles, Contentious Practice, Intimate Identities, Santa Fe: School of American Research

Holland, D., Lachicotte, W., Skinner, D., and Cain, C. (1998). Identity and Agency in Cultural Worlds. Harvard: Cambridge.

House, E. R. (1993). Professional Evaluation: Social Impact and Political Consequences. Newbury Park: Sage.

House, E. R. (1979). *The Objectivity, Fairness, and Justice of Federal Evaluation Policy as Reflected in the Follow Through Evaluation*. Educational Evaluation and Policy Analysis. (1)1.

House, E. R. (1978). *Evaluation as Scientific Management in U.S. School Reform*. Comparative Education Review. Oct.

Jameson, F. (1984). Postmodernism, or the Cultural Logic of Capitalism. New Left Review 146 July/August.

Judd, C. H. (1918). Survey of the St. Louis Public Schools. Yonkers-on-Hudson: New York.

Kliebard, H. M. (1999). Schooled to Work: Vocationalism and the American Curriculum, 1876-1946. Teachers College Press: New York.

Klingner, J., Sturges, K. & Harry, B. (2003). *Conducting Ethnographic Observations of Reading in Elementary Schools*. In Vaughn, S. and Briggs, K. (Eds.). Reading in the Classroom: Systems for Observation of Teaching and Learning. Brookes: Baltimore.

Lachicotte, W. (2002). *Intimate Powers, Public Selves: Bakhtin's Space of Authoring*. In Mageo, J. (Ed.) Power and the Self.

Lagemann, E. C. (2000). An Elusive Science: The Troubling History of Education Research. U Chicago Press: Chicago.

Lagemann, E. C. (1997). Contested Terrain: A History of Education Research in the United States, 1890-1990. Educational Researcher. 26(9), 5-17.

LeCompte, M. & Schensul, J. (1999). Designing and Conducting Ethnographic Research. Lanham, MD: Alta Mira Press

Lee, G. C. (1949). The Struggle for Federal Aid: First Phase. New York: Bureau of Publications, Teachers College, Columbia University.

Leont'ev, L. (1974/5) *The Problem of Activity in Psychology*. Soviet Psychology (Winter 1974-75).

Lichten, W. (2000). *Whither Advanced Placement?* Education Policy Analysis Archives. Vol 8, No. 29.

Madaus, G., Scriven, M. & Stufflebeam, D. (1983). *Program Evaluation: An Historical Overview*. In Madaus, G., Scriven, M. & Stufflebeam, D. (Eds.). Evaluation Models. Boston: Kluwer-Nijhoff.

Manicas, P. (1987). A History and Philosophy of the Social Sciences. Oxford: Blackwell.

Marcus, G., & Fisher, M. (1986). Anthropology as Cultural Critique: An Experimental Moment in the Human Sciences. Chicago: Univ. of Chicago Press.

Marx, S. (2006). Revealing the Invisible. New York: Taylor and Francis.

Miles, M. & Huberman, M. (1994). Qualitative Data Analysis: A Sourcebook of New Methods. Beverly Hills, CA: Sage.

Moley, R. (1923). The Cleveland Surveys. The Survey, 50. May 1923, 229-31.

Morell, J. (1990). *Evaluation: Status of a Loose Coalition*. Evaluation Practice.

Morell, J. & Flaherty, E. (1978). *The Development of Evaluation as a Profession: Current Status and Some Predictions*. Evaluation and Program Planning, 11(3), 213-219.

National Commission on Excellence in Education (1983). A Nation at Risk.

National Education. (1871). Legislative Testimony of the Hoar Bill.

Oakes, J. (1985). Keeping track: How schools structure inequality**.** New Haven, CT: Yale University Press.

OECD. (2003) Development Co-Operation 2003 Report. Accessed online at http://www.oecd.org/dataoecd/62/12/45342482.pdf. April 1, 2007.

Patrizi, P. & McMullan, B. (1998). *Evaluation in foundations: The unrealized potential*. Paper presented at the W.K. Kellogg Foundation Evaluation Unit

Popkewitz, T. (2001). *The Production of Reason and Power: Curriculum History and Intellectual Traditions*. In Popkewitz, T.S, Franklin, B, & Pereya, M. (Eds.). Cultural History and Education: Critical Essays on Knowledge and Schooling. New York: Routledge.

Popkewitz, T. (2000). *The Denial of Change in Educational Change: Systems of Ideas in the Construction of National Policy and Evaluation*. Educational Researcher, 17-29.

Popkewitz, T. S. & Brennan, M. (1998). *Restructuring of Social and Political Theory in Education: Foucault and a Social Epistemology of School Practices*. In Popkewitz, T. S. & Brennan, M. (Eds.). Foucault's Challenge: Discourse, Knowledge, and Power in Education. Teachers College Press: New York.

Popkewitz, T. S. (1991). A Political Sociology of Educational Reform: Power/Knowledge and Power in Teaching, Teacher Education, and Research. New York: Teachers College Press.

Rist, R. (1990). Program Evaluation and the Management of Government. New Brunswick, NJ: Transaction Books.

SAIC company website. (http://www.saic.com/about). Accessed March 10, 2011.

Sanjek, R. (1990). Fieldnotes: The Makings of Anthropology. Cornell University Press: Ithaca, NY.

Scriven (1996). *The Theory behind Practical Evaluation*. Vol 2(4): 393-404. Evaluation. New York: Sage.

Sfard, A. & Prusak, A. (2005) *Telling Identities: In Search of an Analytic Tool for Investigating Learning as a Culturally Shaped Activity*, Educational Researcher, Vol. 34, No. 4, 14-22

Shumar, W. (1997). College for Sale: A Critique of the Commodification of Higher Education. Falmer Press: Washington, DC.

Somers, M. & Gibson, G. (1994) *Reclaiming the Epistemological "Other": Narrative and the Social Constitution of Identity* in Craig Calhoun. (ed) Social Theory and the Politics of identity. Oxford & Cambridge: Blackwell Publishers.

Suchman, E. (1967). Evaluation Research. New York: Russell Sage

Sullivan, S. (2006). Revealing Whiteness: the Unconscious Habits of Racial Privilege. Indiana University Press: Bloomington.

Taussig, M. (1999), Defacement: Public Secrecy and the Labor of the Negative. Stanford: Stanford Univ. Press.

Taylor, F. W. (1911). The Principles of Scientific Management. New York: Harper & Brothers.

Texas Center for Educational Research. (2006). Summative Evaluation Report for Year 6 Texas GEAR UP. Submitted to the Texas Education Agency.

Travers, R. (1983). How Research has Changed American Schools: A History from 1840 to the Present. Kalamazoo: Mythos Press.

Torres, C. A. (1999). *Critical Theory and Political Sociology of Education: Arguments*. In Popkewitz, T. & Fendler, L (Eds.). Critical Theories in Education: Changing Terrains of Knowledge and Politics. Routledge: New York.

Trouillot, M. R. (1995). Silencing the Past: Power and the Production of History. Beacon: Boston.

Tyack, D. B. (1974). The One Best System: A History of American Urban Education. Harvard: Cambridge.

Tyler, S. (1986). *Post-Modern Ethnography: From Document of the Occult to Occult Document*. In Clifford, J., and G. Marcus, (Eds.). Writing Culture: The Poetics and Politics of Ethnography. Berkeley: University of California Press.

U.S. Department of Education (2001). No Child Left Behind Act of 2001. http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Urrieta, L. Jr. (2009). Working from Within Chicana and Chicano Activist Educators in Whitestream Schools. University of Arizona Press: Tucson.

Urrieta, L. Jr. (2006). *Community Identity Discourse and the Heritage Academy: Colorblind Educational Policy and White Supremacy*, International Journal of Qualitative Studies in Education, Vol. 19, No. 4, pp. 455-476

Vanderwalker, N. (1898). *Some Demands of Education upon Anthropology*. American Journal of Sociology 4(1).

Voloshinov, V. (1929/73). Marxism and the Philosophy of Language. Part II.

Vygotsky, L. S. (1978). Mind in Society: The Development of Higher Psychological Processes. Cambridge, MA: Harvard University Press.

Water, M. (1990. Ethnic Options: Choosing Identities in America. Berkeley: University of California Press

Waktins, J. (2001). *Re-Searching Researchers and Teachers: Comment on "Not Talking Past Each Other."* In Anthropology & Education Quarterly. V 342, No. 3.

Weiler, H. (1990). *Decentralised Evaluation in Education Governance: An Exercise in Contradiction*? In Granheim, M., Kogan, M. & Lundgren, U. (Eds.). Evaluation as Policymaking: Introducing Evaluation into a National Decentralised Educational System. London: Jessica Kingsley.

Weiss, C. W. (1987). *Evaluating Social Programs: What Have we Learned?* Society. Nov/Dec.

Wolcott, H. (1994). Transforming qualitative data: Description, analysis, and interpretation. Sage: Thousand Oaks.

Wortham, S. (2003). *Accomplishing Identity in Participant-Denoting Discourse*, Journal of Linguistic Anthropology, Vol. 13, No. 2, 189-210

Vita

Keith M. Sturges was born in Bethesda, Maryland on January 25, 1968. After completing his undergraduate work in Anthropology at the University of Maryland at College Park in 1993, where he graduated magna cum laude, Keith worked as a junior faculty member for the University of Maryland system where he drafted a curriculum for a PhD program in Applied Anthropology and worked as a Research Associate for the Cultural Systems Analysis Group where he got his first taste of contract research. He earned seven academic awards as an undergraduate and wrote three winning contract proposals. Keith then enrolled in the University of Maryland at College Park's Master of Applied Anthropology program where he was awarded a Foreign Language and Area Studies fellowship to enable his work on the use of remittances in development in rural Panama. He worked as a researcher at for- and non-profit contract research organizations, regional education laboratories, and at the University of Miami where he published on topics related to community development, parental involvement in education, the overrepresentation of minority children in special education, and education reform. In 2004, he began studies at the University of Texas at Austin for, what seemed at the time, pragmatic reasons. He became a full-time student in 2006 when he received the William S. Livingston Graduate Fellowship.

Permanent Address: 3208 Hillmeade Court
Nashville, TN 37221

Email: keithmsturges@yahoo.com

This dissertation was typed by the author.