Culture, Cognition, and Evaluation of STEM Higher Education Reform: A Mixed-Methods Longitudinal Study

Proposal by Susan B. Millar, Matthew T. Hora, and Charles W, Kalish

submitted through the Wisconsin Center for Education Research
on Jan. 7, 2008
to the NSF's
Research and Evaluation on Education in Science and Engineering (REESE) Prsogram

NSF proposal number: 0814724

Culture, Cognition, and Evaluation of STEM Higher Education Reform:
A Mixed-Methods Longitudinal Study

Project Summary

Proposal Type – Empirical Research; Topical Strand – Frontier Research

This proposal addresses the REESE Frontier Research topics *Measurement, Modeling, and Methods* and *Cognitive Processes Underlying STEM Learning and Teaching*. The domain of the proposed work is STEM education improvement efforts in higher education. In seeking to understand the slow and spotty progress in implementing proven reform-based practices in higher education, researchers increasingly look to cognitive frameworks to explain institutional inertia and individual reactions to change initiatives. In so doing, they encounter the methodological and conceptual challenges of determining the appropriate unit of analysis for measuring and modeling culture and cognitive variables. The purpose of the proposed study is to address both the policy issue of slow adoption of STEM education reform and the methodological unit-of-analysis problem by deepening understanding of the *relationships between* individual cognition and institutional contexts. The goals of the study are to integrate cognitive and cultural factors into a mixed-methods evaluation design; to field-test this evaluation design in the area of higher education reform in order to assess its utility and generate new knowledge; and to disseminate the findings to researcher and practitioner communities across multiple disciplines.

The proposed study is guided by the following core questions: (a) What mental models do individual faculty have of effective pedagogy? (b) Which groups (e.g., department, professional rank) exhibit cultural consensus on these cognitive structures? (c) How do individual and institutional characteristics covary with individual cognitive structure and group cultural consensus? (d) How do individual- and group-level variables covary with teaching practices? (e) Which individual- and group-level characteristics, if any, contribute to changes in cognitive structure or cultural consensus over time? and (f) What effect does participation in a reform effort have on individual cognitive structure, group cultural consensus patterns, and individual teaching practices? To answer these questions, the study will use a longitudinal, mixed-methods design that will follow STEM faculty from mathematics, life sciences, and physical science departments over 3 years. Data on faculty views of effective pedagogical practices for undergraduate STEM courses, classroom practices, and attitudes toward reform activities will be gathered from STEM educators at three research-intensive universities through Web-based surveys and ethnographic interviews. The data-gathering instruments and analysis processes (Pathfinder analysis, cultural consensus analysis, and hierarchical linear modeling) have been extensively tested and validated in various disciplines, and will directly address the unit-of-analysis problem.

Intellectual merit. The proposed research combines longitudinal and mixed research methods in a conceptual framework that addresses concerns about the need for inter-disciplinary methodologies, and the validity of aggregating individual-level data to produce conclusions about higher organizational levels. It does so in ways that promise to (a) contribute new knowledge about the dynamics that underlie efforts to scale up proven innovations and (b) provide better methods and measures for use in STEM evaluation studies. It presents a carefully designed research plan that will be critiqued and strengthened by a group of highly qualified external advisors and implemented by a highly qualified research team.

Broader impact. Because the research area addressed by the proposed project not only is important for STEM educators, evaluators, and policymakers, but also addresses foundational issues facing the disciplines of cognitive science, education, organizational behavior, and cultural anthropology, project findings will be of significant value to researchers, evaluators, and policymakers in the U.S. and abroad. In particular, our research seeks to incorporate recent developments in measuring and modeling culture, cognition, and context into STEM evaluation designs, and thereby to advance the theoretical and methodological state-of-the-art for STEM education research and evaluation and help transform STEM education research and policy. Finally, results from this research will be disseminated through case study narratives provided to each research site, presentations at national meetings of researchers and evaluators of STEM education reform efforts, a Web site, and publications designed for both academic and practitioner audiences.

Culture, Cognition, and Evaluation of STEM Higher Education Reform: A Mixed-Methods Longitudinal Study

Project Description

I. BACKGROUND AND SIGNIFICANCE

This Empirical Research proposal addresses the REESE Frontier Research topics *Measurement, Modeling, and Methods* and *Cognitive Processes Underlying STEM Learning and Teaching*. The domain of the proposed work is STEM education improvement efforts in higher education. The work directly relates to the problem that, despite many demonstrably successful efforts over the last two decades to improve STEM education at the individual instructor or course level, adoption of these improved teaching resources by entire departments, colleges, or institutions has rarely occurred. The cognitive frameworks of individual faculty[1] and the social, cultural, and structural contexts in which faculty operate are consistently cited for the slow rate of adoption of inquiry-based STEM pedagogy. We contend that the *relationships between* individual cognition and institutional contexts are not well understood and that research findings that provide a good understanding of these relationships may enable policymakers and implementers to more successfully foster adoption of improved teaching methods and resources at department, college, and institutional levels.

We have three interrelated goals for this project. Our *first goal* is to integrate cognitive and cultural factors into a mixed-methods evaluation design so that it will be possible for evaluators and policymakers to systematically (a) identify the specific mechanisms by which culture, institutional context, and individual faculty cognition and pedagogical practices influence one another and (b) determine whether reform initiatives result in changes in any of these areas. Our *second goal* is to field-test this evaluation design in the area of higher education reform in order to assess its utility and generate new knowledge. Our *third goal* is to disseminate our findings by compiling narrative case studies for each of our research sites, preparing working papers and publishing research articles, and presenting findings at professional meetings to a wide audience of evaluators, researchers, practitioners, and policymakers.

The outcomes of this project will be of considerable interest to national audiences of researchers, evaluators, and policymakers. First, we believe, for the reason stated above, that research findings about the interaction between cognition and cultural contexts and about the ways in which these interactions affect the adoption of successful STEM education methods and resources will be of value to education researchers and STEM education policymakers. Second, there is a need in STEM education for evaluation designs that effectively measure and model constructs at multiple organizational levels and that can estimate program effects at various analytic levels. Findings about an evaluation design that seeks to meet these needs will be of considerable interest to the STEM evaluation community. Third, institutional leaders and policymakers will be able to use the research findings regarding the complex relationships between culture, context, and individual cognition to design more effective initiatives and policies.

A. Overview

At the outset, we identify conceptual and methodological challenges that are central to integrating cognitive and cultural variables into evaluation designs—namely, determining the appropriate methods and constructs for measuring and modeling culture and cognition at multiple organizational levels. Research in this area must attend to recent developments in cognitive science, educational research, organizational behavior, and cognitive anthropology, which collectively suggest viable methods for eliciting and analyzing culture and cognition as they function at different organizational levels. For example, the traditional method of describing the cultural features of organizational units by aggregating data about the individuals within these units is now called into question. Research in anthropology and organizational behavior has advanced the argument that organizational cultures are not static and that homogenous sets of values and beliefs

[1] By *faculty*, we mean all people, including graduate students, who hold undergraduate teaching positions—whether full- or part-time, tenured or untenured—in postsecondary institutions, except for emeritus faculty and postdocs.

cannot be uniformly ascribed to all members of an organizational unit such as a mathematics department. Instead, while pervasive cultural models may operate within such units, subcultures nested within these groups (e.g., logicians, adjunct faculty) may share different or even opposing values and beliefs. While individuals within these nested organizational units may internalize some of these organizational cultural models, individuals have their own cognitive schemas, or personal mental models, that shape how they receive, interpret, and act on information. As a result, leading researchers now question the adequacy of using constructs valid at the individual level to ascribe values and assumptions to larger organizational units. In other words, we have identified in the processes commonly used for studying cultural phenomena a methodological problem pertaining to the unit of analysis: it is necessary to measure and model cognitive processes at the individual level, cultural consensus at the group level, and the influence of contextual factors at both of these levels. This is important because the specific mechanisms by which these multilevel variables interact, and if and how they influence teaching practices, are not well known.

To address these methodological challenges, our interdisciplinary team of STEM evaluators, cultural anthropologists, education researchers, and cognitive scientists will draw on methods and theory from our various disciplines. We propose a 3-year mixed-methods longitudinal study that will explore the complex relationships among culture, context, and the pedagogical beliefs and practices of individual faculty, and that will use the theory of cultural models from cognitive anthropology to explore these relationships and their implications for reform. We will address the fundamental unit-of-analysis problem identified above by using two distinct analytic procedures to analyze individual-level data: (a) *Pathfinder analysis*, to identify the structure of individual faculty's mental models, and (b) *cultural consensus analysis*—a type of factor analysis that uses respondents as variables rather than aggregating their scores—to identify cultural groupings at higher organizational levels. A strength of cultural consensus analysis is that it explores within- and between-group variance with respect to cultural consensus. Further, we will use hierarchical linear modeling (HLM) to explore relationships among individual cognition, group cultural consensus, and contextual variables and to tease out the effects of participation in STEM education reforms on teaching practices. Finally, we will capitalize on the relative strengths of quantitative and qualitative methodologies by complementing our statistical analyses with findings from our fine-grained ethnographic exploration of faculty experiences with and perceptions of their institutional contexts. Our research team has the expertise needed to conduct this work, and an outside team of highly qualified advisors will critique and assess our work on a regular basis (see Evaluation Plan section).

Our study is guided by the following six research questions:

1. What mental models do individual faculty have of effective pedagogical approaches for undergraduate STEM courses?
2. Which groups (e.g., department, professional rank) exhibit cultural consensus on these cognitive structures?
3. How do individual and institutional characteristics covary with individual cognitive structure and group cultural consensus?
4. How do individual- and group-level variables covary with teaching practices?
5. Which individual- and group-level characteristics, if any, contribute to changes in cognitive structure or cultural consensus over time?
6. What effect does participation in a reform effort have on individual cognitive structure, group cultural consensus patterns, and individual teaching practices?

Answers to these questions will generate knowledge of national significance about the cognitive processes underlying STEM teaching and effective and appropriate methods of integrating measures of cognition, culture, and context into STEM education evaluation designs. Research data will be gathered at three public research universities that have a strong history of involvement in STEM education reform: the University of Wisconsin–Madison (UW-Madison), the University of Washington–Seattle, and the University of Colorado at Boulder. In developing our research design and constructing our data-gathering instruments, we will build on a pilot study that evaluated the attempts of a National Science Foundation

(NSF) Math and Science Partnership project to foster cultural change in institutions of higher education (Hora & Millar, 2006). A Web-based survey (*N* ~ 825) will be administered to the entire population of faculty (including tenure-track, adjunct, and instructional academic staff) in mathematics, life sciences, and physical science departments at these institutions. A sample of the survey respondents also will be selected for ethnographic interviews (*N* ~ 54). These procedures will elicit data on respondents' mental models of (a) effective pedagogical practices for undergraduate STEM courses, (b) classroom practices, and (c) experiences with reform activities.

B. Relation to REESE Program Goals

This research directly addresses a number of goals of the Measurement, Modeling, and Methods topic of the REESE program. One goal of the program is to develop appropriate and robust ways to measure and model constructs at multiple organizational levels. The proposed work addresses this goal by using the constructs of individual cognitive structure and group cultural consensus, which will be measured through a paired comparison of survey and ethnographic interviews. Our procedures for analyzing the survey data—Pathfinder analysis and cultural consensus analysis—have been extensively used and validated in their respective disciplines of cognitive psychology and cultural anthropology, and thus we are confident in their utility for addressing our research questions. Two more goals of the program are (a) to avoid the methodological pitfall of aggregating individual-level data to produce conclusions about higher organizational levels and (b) to estimate effects at various analytic levels. The research proposed here addresses both of these goals by using multilevel theory to estimate the effects of reform participation on individual- and group-level variables, including classroom practices, individual cognitive structure, and group cultural consensus. We address the goal of increasing the validity and reliability of conclusions by using robust quantitative procedures (such as multilevel modeling), making appropriate use of statistical methods, and using qualitative procedures (such as data triangulation) that ensure reliability. We address the goal of developing methods that combine and aggregate different forms of evidence within a single design through our mixed-methods design, which incorporates methodologies and interpretive frameworks from both quantitative and qualitative research paradigms. We also address the goal of integrating current research from cognitive science with education research in order to improve STEM evaluation designs. Results from this research will demonstrate to evaluators how to conduct status designs (i.e., point-in-time evaluations) of institutional culture at different organizational levels, how to measure and model cognitive and cultural variables in experimental or quasi-experimental designs, and how to conceptualize institutional culture for interpretive or case study designs. Finally, this research addresses the goal of making findings accessible to policymakers and the lay public by developing case study narratives that are easy to read and relevant to these audiences.

C. Relation to Long-Term Goals of the Principal Investigator

This proposal extends the STEM education evaluation work of the principal investigator (PI) and other members of the research team by developing multidisciplinary methods that draw on the cognitive sciences, education research, and evaluation. The long-term goals of the PI are to contribute to the understanding of change processes associated with making improvements in STEM teaching and learning, primarily in higher education and secondarily in the K–12 system. In particular, the PI is interested in learning what is needed to move from the *proven pilot stage*, commonly achieved at the level of individual faculty and courses, to the *full-scale implementation stage*, where evidence-based practices become the new practices in the everyday life of faculty and their institutions. In line with this goal, the research proposed here will improve our understanding of the relationship between change processes at the individual level and those at the level of the larger social units in which individuals work.

D. Relation to Existing Knowledge

The performance of U.S. students in mathematics and science has become an increasingly pressing problem, particularly in light of its implications for the future competitiveness and employability of U.S. citizens (Committee on Science, Engineering, and Public Policy, 2006; U.S. Office of Science and Technology Policy, 2006). As a result, the quality of teaching in higher education has come under fire, and since the publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983), we have seen a cascade of criticisms of higher education. Critics have noted that many undergraduates with STEM majors graduate with substantial deficiencies in their content knowledge (e.g., Handelsman et al., 2004) and that poor instruction and academic assistance lead to high rates of attrition among undergraduate science majors (Seymour & Hewitt, 1997). In response, there has been a significant investment in STEM education reforms in higher education, particularly by the NSF Division of Undergraduate Education (DUE). However, it is apparent that the improved teaching and learning methods and resources that have resulted from this investment are, for the most part, used only by the innovators, either on their own or in small groups. Researchers have found that institution-level adoption has been slow and spotty (Stensaker & Norgard, 2001) and that major structural and cultural barriers impede the widespread implementation of proven practices in STEM disciplines (Ertmer, 1999; Johnson, 2000). Among the reasons cited for the slow rate of change in higher education are (a) the structural organization of academic institutions (Birnbaum, 1988); (b) the insufficient attention paid by reform efforts to the structural and cultural realities of the university system (Thorn, 2003; McAlpine & Weston, 2002), particularly those at the departmental level (Walczyk, Ramsey, & Zha, 2007); (c) the barriers faculty conceptions of change processes and effective pedagogy pose to successful adoption of reform-based practices (Sunal et al., 2001); and (d) the tendency for change efforts to focus more on innovation than on managing change and to depend on intermittent external funding (Tobias, 1992).

Researchers in education reform and organizational behavior are increasingly drawing on cognitive frameworks to explain institutional inertia and individual reactions to change initiatives. Of particular interest is schema theory, which states that new information and experiences are mediated through an individual's unconscious mental structures, which selectively frame and connect new ideas in terms of what is already encoded in memory (Schank & Abelson, 1977; Rumelhart, 1980). Johnson-Laird (1980) introduced the idea of mental models, which are specific knowledge structures that are constructed to represent a new situation using generic knowledge that may also underlie an individual's ability to causally account for the physical environment (Tienson, 1990; Brewer, 1987). These insights from cognitive psychology are widely used in research on faculty pedagogical practices, which is finding that exemplary instructional practices entail development of complex and sophisticated cognitive schema and reasoning skills (Borko & Livingston, 1989; Dunkin, 1995; Hativa, Barak, & Simhi, 2001). In concert with advances in cognitive science, some researchers are focusing instead on the structure and content of specific types of knowledge that actors use in teaching, including individuals' underlying procedural knowledge for teaching (Villachia et al., 2001; Housner, Gomez, & Griffey, 1993), and the relationship between procedural knowledge and student performance (Acton, Johnson, & Goldsmith, 1994). Yet other researchers focus on how individual cognitive processes are influenced by, and even embedded in, specific situations and cultural artifacts (Resnick, 1991; Lave & Wenger, 1991; Hutchins, 1995). The idea that structural and social features of an organization can shape the character of cognitive activities is being explored as a way to explain collective patterns of beliefs and practices in institutions (Narayanan & Armstrong, 2005; Coburn, 2001). In higher education, there has been some focus on the academic department as a locus for meaning formation (Van Maanen & Barley, 1985), as disciplinary groups may have their own epistemologies, beliefs, and expectations about teaching and academic life (Becher & Trowler, 2001; Austin, 1996).

These lines of inquiry have implications for STEM education reform, as researchers are finding that individuals' mental models deeply influence the types of environmental information to which they attend and the way in which they interpret reform initiatives (Nisbett & Ross, 1980; Spillane, Reiser, & Reimer, 2002). As a result, some researchers argue that organizational change depends on changes in individuals' mental models (Argyris & Schon, 1978; Huber,1991) and that reforms to improve teaching will be ineffective without efforts to make the social and cultural contexts of academic departments more conducive to better teaching (Knight & Trowler, 2000; Schroeder, 2001; Millar, 2003). However, there has been little empirical research on the relationship between institutional contexts and individual faculty mental models for teaching practice (Gomez, Hadfield, & Housner, 1996; Kane, Sandretto, & Heath, 2002). Indeed, the research on mental models conducted by scholars in several disciplines suffers from a lack of consensus regarding definitions, and procedures for eliciting and analyzing data (Lakomski, 2001). This is important because consensus on definitions and procedures is particularly important for research in non-laboratory settings where individual agents are engaged in non-standardized practices such as teaching (Narayanan & Armstrong, 2005).

There are additional conceptual and methodological developments on the effect of the institutional context on individual faculty that must be, and to some degree have been, addressed by researchers. First, the theory of organizational culture that has been dominant in business and education tends to use the culture concept to refer to a unitary and stable set of values and beliefs uniformly shared by organizational members (Smircich, 1983; Martin, 2002; Bate, 1997). This tendency has led scholars to focus almost exclusively on individual faculty or the institution as a whole, largely ignoring the intermediate levels of academic departments (Trowler & Cooper, 2002). Increasingly, anthropologists recognize that studying culture in a postindustrial world requires abandoning the notion of isolated groups of people who uniformly share a stable and unchanging set of values and practices (Clifford & Marcus, 1986; Mintz, 1986; Wolf, 1982). This development has been mirrored in organizational studies by a focus on organizational subcultures and the ways in which the policies and practices of particular units (and the units in which they are nested) affect the diverse ways in which individuals act (Kuh & Whitt, 1988).

Second, research in cognitive anthropology has contributed to these dialogues a theory of culture that is based on explanatory structures for specific domains (cultural models) that are acquired from an individual’s social and cultural environment and are deeply embedded in the cognitive processes of the brain (Hutchins, 1995; Shore, 1996; D’Andrade, 1995). These cultural models are instantiated in the form of public artifacts (e.g., course syllabi, rituals) that contain symbolic information about a group’s values and behavioral norms, and are internalized by individuals as personal mental models (Shore, 1996). Furthermore, when cultural schemas are constantly present and activated in the environment, this serves to reinforce certain neural connections that can be very durable and strong (Strauss & Quinn, 1998). This conception of culture is entirely consistent with the above-described literature on the role of individual cognition in resisting reform initiatives.

A third development centers on questions regarding the validity of measuring culture by aggregating individual-level data, generally obtained using surveys of attitudes or beliefs, to produce conclusions about higher organizational levels using mean scores (Ashkanasy, Wilderom, & Peterson, 2000). The validity of this procedure has long been questioned on the grounds that individual-level psychological measures cannot be validly aggregated to account for group-level phenomena, which may operate according to an entirely different logic (Glick, 1985; Ashkanasy et al., 2000). Procedures that rely on aggregation also run the risk of the ecological fallacy problem (Robinson, 1950; King, 1997), in which aggregate results may substantially differ from individual-level results. In response, researchers are increasingly arguing for more sophisticated approaches to modeling multilevel phenomena that account for variability among individuals and the impact of group membership (Porter & Umbach, 2001). Furthermore, some researchers advocate the continued use of traditional methods of field ethnography to study complex behavioral phenomena (Agar, 1996; Bernard, 2002).

Despite conceptual and methodological advances in this area of research, the constructs of culture and cognition have not been widely integrated into STEM education evaluation in a systematic and robust fashion. While a chief aim of the evaluation of the NSF DUE Course and Curriculum Development Program has been to assess whether a project effected changes in institutional culture, the construct *culture* has not been systematically operationalized or measured (Eiseman, Fairweather, Rosenblum, & Britton, 1998). For example, some researchers view culture less as an organizational attribute and more as a conceptual lens through which to analyze issues such as the achievement gap and minority retention in STEM fields (Mertens & Hopson, 2006). Researchers increasingly stress methodological issues in STEM education evaluations, motivated by the expectation that the resulting findings will more effectively support decisions about program replicability and contribute to knowledge-based claims that support education research (Lawrenz & Huffman, 2006; Kelly & Yin, 2007). Some researchers in the STEM education reform community also are surfacing difficulties associated with studying change processes in complex institutional environments, and with how evaluation findings often need to be more carefully interpreted in light of local contextual factors (Patton, 2006; Anderson & Helms, 2001). Taken together, these examples point to the need to develop robust and replicable procedures for measuring and modeling culture and cognition to evaluate STEM education reforms. As is evident from this short review of the literature, the interests of diverse disciplines are converging on the task of understanding the role that culture and context play in influencing individual cognitive behavior in complex institutions.

For this research, we will draw on the methodological and conceptual advances in mixed-methods research, modeling cognitive structure, cultural consensus theory, and multilevel modeling. The following are some of these advances.

- *First*, researchers are increasingly using mixed-methods research designs to address research problems, drawing from qualitative and quantitative methodologies (Tashakkori & Teddlie, 1998; Johnson & Onwuegbuzie, 2004). This development is particularly salient in STEM education evaluations, where some researchers and evaluators argue for methodological pluralism (Lawrenz & Hoffman, 2006; Patton, 2006) and for the unique value of a mixed-methods approach to conducting evaluation in the complex and multidimensional environments of educational settings (National Research Council, 2002).
- *Second*, researchers in cognitive psychology and organizational research have been testing and validating methods to elicit, represent, and systematically analyze the structure of individuals' cognitive processes (Mohammed, Klimoski, & Rentsch, 2000; Shavelson et al., 2002; Hodgkinson & Clarkson, 2005). Techniques in causal mapping represent cognition as a system of cause-effect relations in a specific domain, which are elicited through surveys or interviews and represented pictorially or through data matrices (Narayanan & Armstrong, 2005). A particularly robust analytic method is the previously mentioned Pathfinder analysis, which is based on graph theory in mathematics and produces psychological scaling depicting the underlying structure of concepts in a particular domain (Schvaneveldt, Durso, & Dearholt, 1990). The Pathfinder algorithm transforms paired comparison ratings into a network structure and has been used to assess the structure of knowledge growth for academic subject matter (Acton et al., 1994; Curtis & Davis, 2003).
- *Third*, methods of measuring group consensus on specific cultural knowledge are being widely used in cultural anthropology and cognitive psychology (Medin & Atran, 2004; Weller, 2007). Interview or survey data are submitted to the cultural consensus model (CCM), which is a type of principal components analysis used to determine consensus within and across groups with respect to a specific cultural domain (Romney, Batchelder, & Weller, 1987; Medin & Atran, 2004). CCM addresses the unit-of-analysis problem by using respondents as variables, rather than aggregating their scores, in order to identify emergent cultural groupings at higher organizational levels.
- *Fourth*, methods of analyzing relationships across multiple levels have been used with excellent results in education research. These methods include multilevel models, such as HLM, that allow simultaneous examination of the effects of group- and individual-level variables on individual-level outcomes, while also accounting for the non-independence of observations within groups (Ethington, 1997; Umbach & Porter, 2002). One application relevant to this study is an examination of the influence of departmental contexts on faculty scholarly output, which allows for exploration of data within nested sources of variability, such as faculty within departments within universities (Umbach & Porter, 2002).

II. RESEARCH DESIGN AND METHODOLOGY

A. Research Design

This study will use a longitudinal design, which involves observations of the same item or participants over a period of time and is used in many disciplines. While longitudinal designs have less power than experimental designs in detecting causal relationships, they have more power than point-in time or cross-sectional studies (Bauer, 2004). The type of longitudinal design we propose to use is a panel study, which means data will be collected from the same participants at different points in time. Data will be collected in two phases: Phase I, in the spring of 2009, and Phase II, in the spring of 2011 In order to maintain contact with interviewees from Phase I, we will create and update as needed a database of all respondent contact information. We have successfully managed this tracking work in previous longitudinal studies conducted with small samples of STEM faculty (Hora & Millar, 2006; Hora, 2007). We will provide potential participants with information about the study, including their rights as human subjects and any potential risks, and will include only those who sign a detailed consent form.

We will use a mixed-methods design that integrates surveys, interviews, and document analysis so that we can triangulate across numeric trends and vivid personal narratives. This triangulation will enable us to (a) better understand the mechanisms by which individual faculty cognition, culture, institutional context, and pedagogy influence one another and (b) determine if a reform initiative resulted in changes in any of these areas. We propose to use a survey consisting of a paired comparison exercise (see below) and instruments that elicit information on teaching practices and attitudes toward STEM reform. The paired comparison data will be analyzed using two procedures: (a) Pathfinder analysis to measure the structure of faculty knowledge (mental models) for effective pedagogies for undergraduate STEM instruction; and (b) cultural consensus analysis to explore group-level cultural models. These data will be analyzed using a multilevel model, which will explore relationships among variables at the individual and group levels in relation to our dependent variable of teaching practice. **This sequential analytic process takes into account the need to analyze the influence of the institutional context on individual cognition and group cultural consensus *before* exploring the impact of cognition and culture on teaching practices.** Findings from these quantitative analyses will be complemented by findings from our qualitative analysis of the ethnographic interviews and institutional artifacts (e.g., course syllabi). Table 1 presents the relationships among our research questions, types of data, and analytic methods.

Table 1
Relationship Among Research Questions, Types of Data, and Analytic Methods

Research questions (RQs)	Measures & Data	Analytic methods	Results
1. What are the cognitive structures of individual faculty knowledge of effective pedagogy?	Paired comparison survey using constructs from pilot study; interview	Pathfinder analysis of paired comparison data; thematic analysis of interviews and artifacts	*Individual level:* Map of coherence, complexity, and construct centrality; visual depiction of cognitive structure
2. Which groups (e.g., department, professional rank) exhibit cultural consensus on these cognitive structures?	Paired comparison survey; interview	Principal components analysis of paired comparison data by groups (i.e., department); thematic analysis of interviews and artifacts	*Individual level:* Cultural competency score; membership in cultural group *Group level:* Groups with cultural consensus
3. How do individual characteristics (Level 1) or institutional characteristics (Level 2) covary with individual cognitive structure and group cultural consensus?	RQ1 and RQ2 results; demographic and institutional characteristics; STEM reform survey; interview	HLM; thematic analysis of interviews and artifacts	Significant relationships between Level 1 and 2 independent variables and cognitive structure and cultural consensus (dependent variables)
4. How does cognitive structure (Level 1) or group cultural consensus (Level 2) covary with teaching practices?	RQ1 and RQ2 results; teaching practices survey; interview	HLM; simultaneous equations estimation; thematic analysis of interviews and artifacts	Regression estimates of the direction, magnitude, and statistical significance of cultural consensus measures; estimates of bi-directional effects between teaching practice and attitudes

Research questions (RQs)	Measures & Data	Analytic methods	Results
5. Which individual characteristics (Level 1) or institutional characteristics (Level 2), if any, contribute to changes in cognitive structure or cultural consensus over time?	RQ1 and RQ2 results; demographic and institutional characteristics from survey; interview	HLM; binary and ordered logistic regression; thematic analysis of interviews and artifacts	Regression estimates of the direction, magnitude, and statistical significance of contextual influences on cognitive structure and cultural consensus
6. What effect does participation in a reform effort have on cognitive structure, cultural consensus, and teaching practices?	RQ1 and RQ2 results; STEM reform survey; teaching practices survey; interview	HLM; binary and ordered logistic regression; Heckman selection modeling; thematic analysis of interviews and artifacts	Regression estimates of the direction, magnitude, and statistical significance of contextual influences on cognitive structure, cultural consensus, and teaching practices—in particular, the separate identification of causal and non-causal (selection) effects

Research Sites

In designing this study, we elected to draw study participants from universities with a history of STEM education reform activities, rather than from particular reform programs. Selecting institutions rather than specific reform efforts will allow us to capture the role of culture, context, and cognition in a random sample of faculty at individual institutions and to compare groups of participants and nonparticipants at the same institutions. We selected three institutions using the following criteria: (a) public research-intensive institutions as defined by the Carnegie Foundation for the Advancement of Teaching (2007); (b) institutions with undergraduate enrollments of similar size based on figures from fall 2006; and (c) institutions with similar 4-year averages of NSF DUE funding. Based on these criteria, we selected UW-Madison, the University of Washington–Seattle, and the University of Colorado at Boulder. Collectively, these institutions provide participant and regional diversity. It is important to note that this study is not designed to evaluate the quality or efficacy of individual STEM education reform efforts at these institutions, but rather to enable us to study the effects of reform programs within specific institutional contexts.

Our first step will be to identify all or nearly all of each participating institution's current and recent STEM education reform efforts. To do this, we will interview relevant key leaders (e.g., principal investigators, administrators, especially active faculty). Using findings from these interviews and any program documents or reports we can obtain, we will identify and classify all three institutions' programs within the first 2 months of the study. Data on other institutional characteristics will be gathered through these preliminary interviews and through analysis of documents from the offices of institutional research at each research site. These data will include levels of research funding, level of NSF DUE funding, number of undergraduates in departments, number of faculty in departments, and number of active STEM education reform efforts.

Methods

Research Component I: Survey

The purpose of the survey component of this study is to obtain data from a large probability sample so that we can generalize from a subset of our sampling population to the population as a whole. Furthermore, survey research is a cost-effective and efficient data collection procedure that is particularly useful for resource-limited evaluation applications. The sample for the survey will be drawn from the population of faculty in the **mathematics, life sciences, and earth sciences** departments. These disciplines were selected for the prominent and ubiquitous role they play in undergraduate STEM education. The sample frame includes faculty defined broadly[2] because graduate students and, increasingly, untenured teaching staff (Baldwin & Chronister, 2001) often teach undergraduate courses, and thus may constitute a significant factor in the success of reform initiatives (Seymour, 2005). Our current estimate of faculty numbers (approximately 275 in these departments at each university) is low because it does not include graduate teaching assistants and non-tenured faculty. We anticipate sending surveys to the entire relevant faculty population at each research site. As a result, we will administer an estimated 1200 surveys, and assuming a 50% response rate, we will receive approximately 600 completed surveys.

We will develop a Web-based survey instrument in close collaboration with the UW-Madison Survey Center, which will conduct all necessary programming, sampling, survey administration, and data reporting. The survey center will use the most recent and proven methods for improving survey response rates, including sending a preliminary letter with a small incentive and making reminder phone calls. To ensure that we are including valid and reliable measures in the survey, we will draw on existing instruments that have been tested and validated. The survey instrument itself will also be pretested and validated as a whole (Dillman, 2000). We will administer the first survey in early 2009. The follow-up survey will be administered in early 2011 to the entire population of faculty, including new hires at each institution. The 2-year interval between data gathering points should allow sufficient time for changes to take place in individual faculty cognitive structures, cultural groups, and teaching practices. The survey will include the following items:

1. *Individual characteristics.* Items will ask for age, gender, rank and/or type of appointment, time at institution, department, discipline, research specialty, professional affiliations, and experiences with STEM reforms (including name of reform, duration of participation, and level of involvement).
2. *Paired comparison.* Building on work by Acton et al. (1994) and Curtis and Davis (2003), we will devise a paired comparison exercise that asks participants to select 10 constructs from a set of 35 pertaining to effective pedagogy in STEM courses. These constructs will be derived from pilot studies of the SCALE project (Hora & Millar, 2006). Participants will also be able to create up to two new paired constructs. Participants will then be asked to judge the relatedness of each construct combination, using a 9-point rating scale ranging from *highly similar* (9) to *highly dissimilar* (1). To enhance psychological validity and reduce the time required to conduct these ratings, we will ask respondents to make their judgments quickly (Clariana & Wallace, in press).
3. *Attitudes toward STEM reform.* Items will elicit respondent attitudes toward STEM education reform, including perceived barriers and supports in their home institutions. We will adapt items from survey questions designed for and validated by the Center for the Integration of Research, Teaching, and Learning (CIRTL), among other projects.
4. *Reported teaching practices.* Items will be designed to gather data on philosophical beliefs and teaching practices. These items will be adapted from the Statistics Teaching Inventory (STI), a new instrument developed by the NSF-funded ARTIST project at the University of Minnesota that will be tested and further refined in spring 2008 (Garfield, delMas, & Zieffler, 2007).

We will analyze the survey data in three steps, each of which will employ appropriate and robust analytic methods to answer the six research questions. We will undertake Steps 1 and 2 (see below) after receiving all responses from Phase I and Phase II surveys. Step 3 will occur after our analysis of the Phase II data is complete. This 3-step procedure is informed by a theory of cultural models as detailed above.

Step 1: Analyzing paired comparison data to produce approximations of the cognitive structures of individual faculty knowledge and identify levels of cultural consensus within or between groups. The paired comparison data will initially be analyzed using the Pathfinder (PF) algorithm to identify the cognitive structures of individual faculty knowledge pertaining to effective pedagogy in STEM courses (Research Question 1). PF is a data reduction technique for large amounts of proximity data (i.e., paired

[2] As indicated in footnote 1, we include as "faculty" all people who teach undergraduates, including graduate students—whether full- or part-time, tenured or untenured, except for emeritus faculty and postdocs. The latter are not included as they are unlikely to teach undergraduate courses, and are difficult to identify in campus databases.

comparison results) that transforms data into an adjacency matrix, which consists of rows and columns whose cells are populated by weighted relationships between two constructs. These data will then be analyzed to capture a variety of cognitive characteristics for each participant and to graphically depict an associative network structure (or *mental model*) (Cooke, Durso, & Schvaneveldt, 1986). PF regards the degree of relatedness between individual constructs as an estimate of psychological distance, so highly related concepts are separated by fewer links and less related concepts are separated by more links or may be unconnected. The resulting analyses will provide information about cognitive coherence, complexity, and construct centrality, each of which have been shown to have high construct validity when analyzed using PF (Mohammed et al., 2000). We will use the Knowledge Network and Orientation Tool (KNOT, n.d.) software to run the Pathfinder algorithm. This will initially produce a PFNet representation, which is an estimate of the weight or cost of the link connecting each pair of nodes. From the PFNet, we will produce the following measures for each participant: (a) *cognitive coherence*, which measures the internal consistency of a participant's knowledge representation and can be interpreted to estimate the degree of development of each participant's cognitive framework regarding STEM instruction; (b) *cognitive complexity*, which measures the number of links in each participant's network; and (c) *construct centrality*, which identifies the most salient linked propositions for each participant. The results from these procedures will be used by the multilevel model as individual-level variables for further analysis.

Next, we will analyze the paired comparison data using cultural consensus analysis in order to identify the level of consensus within or between groups on effective pedagogy in undergraduate STEM courses (Research Question 2). We will analyze the adjacency matrices developed through the PF analysis by stacking the columns to create a single column for each respondent and then conducting a factor analysis using SPSS analysis software (Lynch & Medin, 2006). The purpose of the first procedure will be to assess the level of consensus *within academic departments and disciplinary groups*. The model will be run to test within-group consensus on effective pedagogy, which will determine if the data, overall, conforms to a single-factor solution (Romney et al., 1987). In the event that the analysis indicates that there is more than one factor in the solution—or no single response pattern—we will use analysis of variance (ANOVA) to explore other variables (e.g., gender, rank) that may explain the variation. The purpose of the second procedure will be to assess the level of consensus *between academic departments and disciplinary groups* and will follow the steps outlined above. It will also be possible to estimate individual competency using the pattern of inter-informant agreement indices on the first factor, assuming there is consensus (Weller, 2007). While it is important not to interpret these competency scores as expertise scores, they do indicate how well the responses of an individual correspond to those of the group (Atran, Medin, & Ross, 2005). These scores will be submitted to the multilevel model as individual-level variables for further analysis, and the groups that exhibit cultural consensus will be submitted as group-level variables.

Step 2: Analyzing Step 1 results with other data to identify relationships among the individual- and group-level variables and our key dependent variable of teaching practices. We will use the results from the PF and cultural consensus analyses, and data from the other survey items, to explore relationships among variables. We will use an ANOVA procedure to explore the relationships among the data from the PF and cultural consensus analyses both within and between academic departments and disciplinary groups, and we will study these relationships with additional demographic and institutional covariates and survey data on attitudes toward STEM reform in an HLM framework (Research Question 3). We will repeat this procedure after Phase II to gain insight into a third level of variation (across time) in individual cognitive structure or group-level cultural consensus. This procedure will attempt to identify particular variables that contribute to the dynamic processes of attaining cultural consensus (Research Question 5).

Next, we will use HLM to ascertain whether cognitive structure and cultural consensus covary with teaching practices (Research Question 4). For this procedure, results from the PF (individual cognitive coherence, complexity, and construct centrality) and cultural consensus analyses (individual cultural competence scores, groups with cultural consensus) will be tested against data on teaching practices obtained through items adapted from the STI instrument. Since we have two levels of data, the multilevel model can be conceptualized as a two-stage system of equations. In the first stage (Level 1), a separate individual-level regression is defined for each group or higher level unit. In the second stage (Level 2), each of the group or context-specific regression coefficients defined in Equation 1 is modeled as a function of group-level variables. The model includes the effects on the individual-level outcome of (a) group-level variables, (b) individual-level variables, and (c) interaction between group- and individual-level variables. We will use simultaneous equation modeling to account for the potential reciprocal effect between teaching practices and group consensus. This method yields a valid estimate of the effect of one factor on the other in each direction (and, most notably, of the causal effect of cultural attitudes on teaching practices). A set of instrumental variables—those exogenous with respect to this feedback between factors—will be used in two-stage least squares regression and will identify this causal estimate along with other relationships between covariates inside the hierarchical structure (instructors within departments).

Step 3: Analyzing data longitudinally to identify the effects of participation in reform. We will identify the effects of participation in STEM education reforms on individual cognitive structure, cultural consensus, and teaching practices (Research Question 6). This analysis will provide rich information about the dynamic nature of the intervention, using covariate regression analysis (HLM, binary logistic, or ordered logistic regressions as appropriate to the dependent variable) to identify not only causal effects of the reform, but also predictions for potentially heterogeneous effects for departments of different initial patterns of cultural agreement. The longitudinal structure of the data will also permit us to study the factors underlying individuals' decisions to participate in the reform. Binary logistic regression techniques will be used to identify the influence of individual-level factors on the propensity to seek reforms. This analysis will be forward-looking in predicting the rates at which different types of individuals are likely to adopt specific teaching innovations. Furthermore, by using selection bias elimination procedures such as the Heckman two-step process, we will group individuals into treatment categories to better approximate a randomized field trial.[3]

Research Component II: Interviews and Artifact Analysis

The interview component of the study will enable us to obtain insight into the specific processes by which culture, context, and individual cognition influence the pedagogical practices of faculty. The interviews will also serve as an opportunity to validate survey findings, particularly results from the paired comparison exercise, increasing the likelihood that the quantitative analyses accurately represent individual cognitive processes (Shavelson, Ruiz-Primo, & Wiley, 2005). Like Component I, Component II is organized into two phases, with interview data collected after survey data (see Work Plan).

Following administration of the Web-based survey, we will interview 18 faculty from each institution ($N = 54$). To identify prospective participants, we will develop a purposive sample of 30 faculty from the faculty lists used in the survey component, including an extra 12 to allow for respondents who decline to participate. At each institution, we will select at least 6 participants who have not participated in STEM education reform programs.[4] Interviews will be face-to-face and will last 30–45 minutes, a duration that we have found feasible for faculty with demanding workloads. The interview protocol will be semistructured and will include the following components: (a) review and discussion of the participant's cognitive map; (b) questions regarding the participant's teaching practices, their origins, and their contextual influences; (c) questions regarding the participant's perspectives on the institutional culture and related contextual influences and agreement or disagreement with these cultural norms; and (d) questions regarding STEM education reform programs and, if applicable, the participant's experiences with them. The interviews will be digitally recorded and then transcribed and entered into an NVivo project. We will also collect artifacts

---

[3] These processes permit inference on how individuals may be simultaneously more interested in and more capable of implementing reforms.

[4] We have found that including nonparticipants in an evaluation study is an extremely useful way to explore the impact of the institutional context on change processes (Hora & Millar, 2006). Recognizing that inclusion of nonparticipants may not be a cost-effective model for evaluators to follow, we will assess the value of including nonparticipants in the interview sample and include this assessment in our final recommendations.

salient to undergraduate STEM courses in each department, including course syllabi, policies governing undergraduate education, and instructor course descriptions.

Analyzing qualitative data is largely an inductive activity, involving reducing complex data sets into smaller units or categories according to a classification scheme derived from the major themes and patterns in the data (Miles & Huberman, 1994; Bernard, 2002). To analyze the interview transcripts, we will use the grounded theory approach of Strauss and Corbin (1990), which will entail developing a coding paradigm and systematically applying it to the transcripts. A coding paradigm is a structured coding scheme used to analyze data and identify discrete themes and patterns (Strauss & Corbin, 1990; Ryan & Bernard, 2003). We will develop the coding paradigm through an iterative process by reviewing field notes and a sample of interview transcripts to capture themes, patterns, and trends in the data (Ryan & Bernard, 2003). To establish inter-coder reliability, members of the research team will individually code a set of three interviews, and reliability will be established using Cohen's Kappa with a cutoff of .80 (Bernard, 2002). Next, we will use the matrix coding function in NVivo to identify high incidences of cross-coded text and will constantly compare these highly referenced themes with field notes and survey findings (Glaser & Strauss, 1967; Bernard, 2002). Then, we will organize the recurring themes in a time-ordered display to identify relationships among individual themes over time (Miles & Huberman, 1994). We have found that causal networks are an effective and engaging tool through which to convey complex qualitative data to a wide audience (Hora & Millar, 2006). Finally, we will assess findings from the interviews in light of the documentary artifacts we have collected in order to ascertain how individuals' cognitive processes were influenced by, or instantiated in, these artifacts (Halverson & Clifford, 2006). We will validate the findings using standard techniques in qualitative research, including exploring countervailing evidence and member-checking research findings with participants (Miles & Huberman, 1994; Bernard, 2002).

Data Interpretation Procedures and Dissemination to Participating Institutions

Research findings from the quantitative procedures will be constantly compared to findings from the qualitative component, and these data will be integrated to form a coherent narrative of institutional culture and practice (Polkinghorne, 1995). We will contextualize our statistical analyses with explanatory narrative from our qualitative analyses to produce conclusions that integrate both types of data. We will note which findings are based on strong evidence and which are speculative in nature, regardless of the research methodology. To make the data accessible and usable, we will ascertain the unique needs of our audiences and vary the level of detail in our reports accordingly (Frechtling, Sharp, & Westat, 1997). We will construct narrative case studies of each research site with the intention that they be of value to administrators and STEM education leaders at the sites. A case study approach is well suited to the in-depth analysis of complex issues and processes like organizational culture (Shulman, 1987; Stake, 1994). The case study for each research site will also make recommendations for future STEM education reform planning and evaluation, including suggestions for overcoming the identified barriers and maximizing opportunities for program implementation.

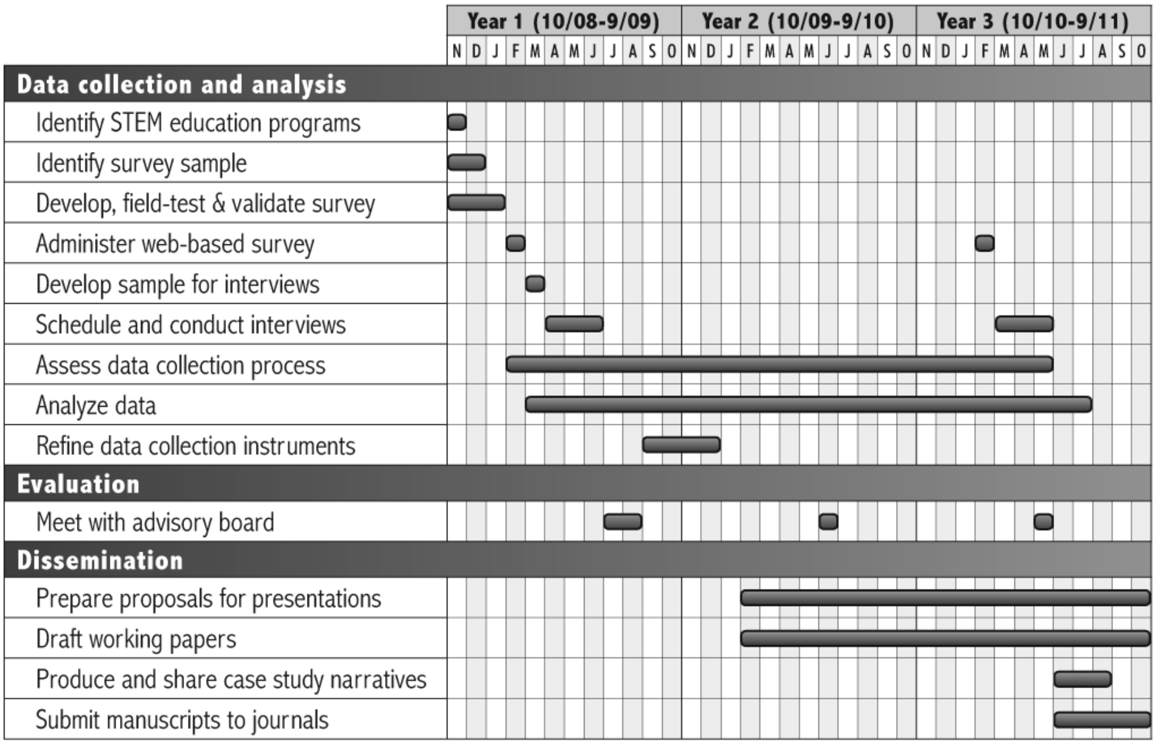
Implications for STEM Education Evaluation

We anticipate that our study will yield findings on (a) the ways that culture, cognition, and context interact to influence teaching practices and reform implementation and (b) appropriate methods for integrating cultural and cognitive factors into evaluation designs. Results from this research will demonstrate to evaluators how to (a) conduct status designs (i.e., point-in-time evaluations) of institutional culture at different organizational levels, (b) measure and model cognitive and cultural variables in experimental or quasi-experimental designs, and (c) conceptualize institutional culture for interpretive or case study designs. Each of these contributions will be valuable to an evaluation community that is striving for improved methodological rigor, responsiveness to multiple disciplines, and insights into the cultural and cognitive aspects of education reform.

B. Work Plan

Table 2 depicts our work plan over the 3 years of the proposed project.

Table 2
Work Plan



III. PROJECT PERSONNEL AND MANAGEMENT

A. Principal Investigator and Senior Personnel

The PI, **Susan Millar**, will take final responsibility for the scientific and technical direction of the project. Working with the research team, the co-PIs, and the advisory board, she will ensure the scientific quality of the project. Co-PI **Charles Kalish** will devote 8% of his time to the project. He will act as the project's methodology and measurement leader. Co-PI **Matthew Hora** will devote 75% of his time to this project. Hora will act as the project director. In this role, he will (a) collaborate with Millar and Kalish to ensure the scientific quality of the project; (b) with direction from Millar, manage the budget and ensure that work proceeds according to the timeline; (c) prepare for the team's annual meetings with the advisory board. He also will act as the project's primary researcher, implementing all research activity related to the ethnographic interviews and taking overall responsibility for data analyses, writing, and dissemination of findings. A half-time graduate student (TBD) will assist Hora with these qualitative research duties. **Jeffrey Watson**, a researcher at WCER, and **Nicholas Mader**, a doctoral dissertator in economics at UW-Madison and a WCER researcher, will devote 10% and 50% of their time, respectively, to the project. Watson and Mader will work closely with Hora to develop data management procedures and collect and analyze all quantitative data. The student hourly, who is budgeted for 780 hours per year at $10/hour, will enter data into software applications, transcribe interviews, and assist the research team as requested.

B. Graduate Student Training Opportunities

This study will provide opportunities for two graduate students at UW-Madison in educational psychology, economics, or other aligned fields. The project will be structured to train graduate students in mixed methods with a particular focus on STEM evaluation. In this way, we will help train the next generation of STEM evaluators and educators in cutting edge methods. Students will also help with instrument design and field testing, administering instruments, analyzing data, and writing. They will take the lead in developing data management and dissemination procedures. Their involvement in writing case studies and presenting will give them the opportunity to develop technical and presentation skills.

C. Project Capacity: Results from Prior NSF Support

Susan Millar is the PI on “Engaging Critical Advisors to Formulate a New Framework for Change: Expansion of ‘Toward a National Endeavor to Marshal Postsecondary STEM Education Resources to Meet Global Challenges,’” NSF DUE-0744106 ($660,109), funded from October 1, 2007, through December 31, 2010, and “Toward a National Endeavor to Marshal Postsecondary STEM Education Resources to Meet Global Challenges: A Planning Proposal,” DUE-0732521 ($99,733), funded from June 15, 2007, through May 30, 2008. These two grants, which are devoted to achieving the same purpose, are in the early implementation stage. Millar was also a PI on NSF award #0127725, “Assessing Student Learning and Evaluating Faculty Capacity Development in the NSF-Funded Regional Workshop Project,” granted to Richard Iuli, University of Rochester. That grant was funded at $640,000 for 4 years (2001–2005); additional research and dissemination continued through August 2007 under a no-cost extension. This award resulted in a publication in *Metropolitan Universities* titled “Using Workshops to Improve Instruction in STEM Courses” (Connolly & Millar, 2007). Charles Kalish was the PI on NSF award BCS-0425397 ($15,000), “Understanding People as Normative Agents: The Intersection of Morality and Theory of Mind,” funded from September 2004 through August 2006. The meetings that resulted from this award led to several ongoing collaborative projects.

IV. DISSEMINATION

Our goals for this research are to generate transformative concepts in organizational behavior and STEM education, field-test a new approach to the evaluation of education reform efforts, and disseminate our findings to the researcher and practitioner communities across multiple disciplines. Research on knowledge dissemination has identified challenges with traditional researcher-to-practitioner approaches to disseminating results and recommendations from research activities (Hutchinson & Huberman, 1993; National Academy of Education, 1999). To make research useful and accessible to diverse audiences requires particular attention to the medium in which knowledge is presented and delivered and to the intended users of the information. Our plan for achieving broader impacts takes these issues into account. Our intended audiences include the academic leaders at the research sites, all members of the STEM education community, and especially those involved in improvement of postsecondary STEM education, the broader evaluation community, and the researcher communities in organizational behavior, cognitive psychology, and cultural anthropology. The first step in our dissemination plan will be to develop and distribute case study narratives of our findings to administrators and STEM education leaders at each research site. We also disseminate our findings by writing working papers, presenting our findings at professional meetings, and publishing articles in both peer-reviewed and practitioner journals. In particular, during spring 2009 and spring 2010, we will prepare proposals for meetings of the American Educational Research Association, the American Evaluation Association, and the Society for Applied Anthropology. In fall 2009 and fall 2010, we will submit proposals to the NSF Math and Science Partnership Learning Network Conference and other program-specific meetings of STEM faculty and evaluators. Finally, we will establish a project Web site to disseminate our findings and activities to a worldwide audience via the Internet.

V. EVALUATION PLAN

An external advisory board will provide technical and managerial oversight to the project. This advisory board will provide substantive external review through feedback on the project’s research methods and progress, analysis, interpretation, and dissemination. Millar, Hora, and Kalish will consult with each member of the advisory board each summer for an in-depth review of the project. This will allow each advisor substantial time to review the progress of the project and to make detailed assessments and suggestions. Interaction with the board will ensure that the evaluation is distant from the project and objective.

The following individuals will serve on the advisory board: (a) **Dr. Geoffrey Borman**, professor of educational leadership and policy analysis, educational psychology, and educational policy studies at UW-Madison, who is a nationally recognized expert in quantitative methodologies, including multilevel modeling and experimental studies of educational innovations; (b) **Dr. Norbert Otto Ross**, assistant professor of anthropology and psychology at Vanderbilt University, who has published widely in cognitive anthropology and has extensive experience in integrating qualitative and quantitative methods; (c) **Dr. Ann Austin**, professor of educational administration at Michigan State University, who is nationally recognized as an expert in organizational change and faculty cultures in higher education; (d) **Dr. Norman Webb**, senior scientist at WCER, who is a nationally recognized authority on STEM evaluation and survey methodology; (e) **Dr. Elizabeth Lynch**, researcher in the Department of Preventive Medicine at Northwestern University, who is trained as a cognitive psychologist; and (f) **Dr. Sharon Derry**, professor in education psychology at UW-Madison, who is widely recognized as an expert in cognitive aspects of STEM instruction and mental model research. Taken together, the six members of the advisory board have strong expertise in disciplines relevant to our project.

VI. SUMMARY

The proposed study seeks to improve our understanding of the relationship between individual cognition and institutional context and thereby contribute to the body of research on to the weak adoption by departments, colleges, or institutions of demonstrably successful efforts to improve STEM education at the individual instructor or course level. The goals of this study are (a) to integrate cognitive and cultural factors into a mixed-methods evaluation design; (b) to field-test this evaluation design in order to generate new knowledge about these topics while simultaneously assessing the utility of the design for the STEM education evaluation community; and (c) to disseminate our findings by compiling narrative case studies for each research site, preparing working papers and research articles, and presenting findings at professional meetings to a wide audience of evaluators, researchers, practitioners, and policymakers. The proposed research has intellectual merit for several reasons. It combines longitudinal and mixed research methods in a conceptual framework that addresses concerns about the validity of aggregating data on individual cognition to describe the cultural features of larger social units, and it does so in ways that promise to (a) contribute new knowledge about the dynamics that underlie efforts to scale up proven innovations and (b) provide better methods and measures for use in STEM evaluation studies. It presents a carefully designed research plan that will be critiqued and strengthened by a highly qualified group of external advisors and implemented by a highly qualified research team. Last, because we have chosen a research area that not only is important for STEM educators, evaluators, and policy makers, but also addresses foundational issues facing the disciplines of cognitive science, education, organizational behavior, and cultural anthropology, we believe our project findings will be of value to researchers, evaluators, and policymakers in the U.S. and abroad.

References Cited

Acton, W. H., Johnson, P. J., & Goldsmith, T. E. (1994). Structural knowledge assessment: Comparison of referent structures. *Journal of Educational Psychology, 86,* 303–311.

Agar, M. (1996). *The professional stranger: An informal introduction to ethnography*. San Diego, CA: Academic Press.

Anderson, H., & Helms, J. V. (2001). The ideal of standards and the reality of schools: Needed research. *Journal of Research in Science Teaching, 38*(1), 3–16.

Ashkanasy, N. M., Wilderom, C. P. M., & Peterson, M. F. (Eds.). (2000). *Handbook of organizational culture and climate*. Thousand Oaks, CA: Sage.

Argyris, C., & Schon, D. (1978). *Organizational learning*. Reading, MA: Addison Wesley.

Atran, S., Medin, D. L., & Ross, N. O. (2005). The cultural mind: Environmental decision making and cultural modeling within and across populations. *Psychological Review, 112*(4), 744–766.

Austin, A. E. (1996). Institutional and departmental cultures: The relationship between teaching and research. *New Directions for Institutional Research. 84,* 65–82

Baldwin, R. G., & Chronister, J. L. (2001). *Teaching without tenure: Policies and practices for a new era.* Baltimore, MD: The Johns Hopkins University Press.

Bate, S. P. (1997). Whatever happened to organizational anthropology? A review of the field of organizational ethnography and anthropological studies. *Human Relations, 50*(9), 1147–1175.

Bauer, K. W. (2004). Conducting longitudinal studies. *New directions for Institutional Research, 121,* 75–90.

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of discipline*. Philadelphia, PA: Open University Press.

Bernard, H. R. (2002). *Research methods in anthropology: Qualitative and quantitative approaches* (3rd ed.). Lanham, MD: Altamira Press.

Birnbaum, R. (1988). *How colleges work: The cybernetics of academic organization and leadership.* San Francisco: Jossey-Bass.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal, 26,* 473–498.

Brewer, W. F. (1987). Schemas versus mental models in human memory. In P. Morris (Ed.), *Modelling cognition*. New York: Wiley.

Carnegie Foundation for the Advancement of Teaching. (2007). *The Carnegie Classification of Institutions of Higher Education*. Retrieved December 17, 2007, from http://www.carnegiefoundation.org/classifications/

Clariana, R. B., & Wallace, P. E. (in press). A comparison of pair-wise, list-wise, and clustering approaches for eliciting structural knowledge. *International Journal of Instructional Media.*

Clifford, J., & Marcus, G. E. (Eds.). (1986). *Writing culture: The poetics and politics of ethnography*. Berkeley: University of California Press.

Coburn, C. (2001). Collective sensemaking about reading: How teachers mediate reading policy in their professional communities. *Educational Evaluation and Policy Analysis, 23,* 145–170.

Committee on Science, Engineering, and Public Policy. (2006). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: National Academies Press.

Connolly, M. R., & Millar, S. B. (2007). Using educational workshops to improve instruction in STEM courses. *Metropolitan Universities, 17,* 53–65.

Cooke, N. M, Durso, F. T., & Schvaneveldt, R. W. (1986). Recall and measures of memory organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(4), 538–549.

Curtis, M. B., & Davis, M. A. (2003). Assessing knowledge structure in accounting education: An application of Pathfinder Associative Networks. *Journal of Accounting Education, 21*(3), 185–195.

D'Andrade, R. (1995). *The development of cognitive anthropology*. Cambridge, UK: Cambridge University Press.

Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (2nd ed.). New York: Wiley.

Dunkin, M. J. (1995). *Concepts of teaching and teaching excellence in higher education*. *Higher Education Research and Development, 14*(1), 21–33.

Eiseman, J. W., Fairweather, J. S., Rosenblum, S. & Britton, E. (1998). *Evaluation of the Division of Undergraduate Education's Course and Curriculum Development Program*. Arlington, VA, National Science Foundation.

Ertmer, P. A. (1999). Addressing the first- and second-order barriers to change: Strategies for technology integration. *Educational Technology Research and Development, 47*(4), 47–61.

Ethington, C. A. (1997). A hierarchical linear modeling approach to studying college effects. In J. Smart (Ed.), *Higher education handbook of theory and research* (Vol. 12, pp. 165–194). New York: Agathon Press.

Frechtling, J., Sharp, L., & Westat, Inc. (Eds.). (1997). *User-friendly handbook for mixed method evaluations*. Washington, DC: National Science Foundation, Directorate for Education and Human Resources.

Garfield, J., delMas, B., & Zieffler, A. (2007). *Statistics Teaching Inventory (STI)* (Draft version). Minneapolis: University of Minnesota, ARTIST Project.

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

Glick, W. H. (1985). Conceptualizing and measuring organizational and psychological climate: Pitfalls in multilevel research. *Academy of Management Review, 10*(3), 601–616.

Gomez, R. L., Hadfield, O. D., & Housner, L. D. (1996). Conceptual maps and simulated teaching episodes as indicators of competence in teaching elementary mathematics. *Journal of Educational Psychology, 88*(3), 572–585.

Halverson, R. R., & Clifford, M. (2006). Evaluation in the wild: A distributed cognition perspective on teacher assessment. *Educational Administration Quarterly, 42*(4), 578–619.

Handelsman, J., Ebert-May, D., Beichner, R., Bruns, P., Chang, A., DeHaan, R. L., et al. (2004). Scientific teaching. *Science, 304*, 521–522.

Hativa, N., Barak, R., & Simhi, E. (2001). Exemplary university teachers: Knowledge and beliefs regarding effective teaching dimensions and strategies. *The Journal of Higher Education*, *72*(6), 701–729.

Hodgkinson, G. P., & Clarkson, G. P. (2005). What have we learned from almost thirty years of research on causal mapping? Methodological lessons and choices for the information systems and information technology communities. In V. K. Narayanan & D. J. Armstrong (Eds.), *Causal mapping for information systems and technology research: Approaches, advances, and illustrations* (pp. 46–79). Hershey, PA: Idea Group.

Hora, M. (2007). Factors influencing change initiatives to improve K–20 STEM education at California State University, Dominguez Hills: Final case study of SCALE activities. Madison: University of Wisconsin–Madison, Wisconsin Center for Education Research.

Hora, M., & Millar, S. (2006). A preliminary case study of SCALE activities at the University of Wisconsin–Madison: Factors influencing change initiatives in STEM undergraduate education, teacher training, and partnerships with K–12 districts (WCER Working Paper No. 2007-2). Madison: University of Wisconsin–Madison, Wisconsin Center for Education Research. Retrieved December 27, 2007, from http://www.wcer.wisc.edu/publications/workingPapers/papers.php

Housner, L. D., Gomez, R., & Griffey, D. C. (1993). A pathfinder analysis of pedagogical knowledge structures: A follow-up investigation. *Research Quarterly for Exercise and Sport*, *64*(3), 291–299.

Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organization Science, 2*(1), 85–115.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science, 19*, 265–288.

Hutchinson, J., & Huberman, M. (1993). *Knowledge dissemination and utilization in science and mathematics education: A literature review*. Washington, DC: National Science Foundation.

Johnson, D. J. (2000). Levels of success in implementing information technologies. *Innovative Higher Education, 25*, 59–76.
Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26.
Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science: A Multidisciplinary Journal*, *4*(1), 71–115.
Kane, R., Sandretto, S., & Heath, C. (2002). Telling half the story: A critical review of research on the teaching beliefs and practices of university academics. *Review of Educational Research, 72*(2), 177–228.
Kelly, A., & Yin, R. (2007). Strengthening structured abstracts for education research: The need for claim-based structured abstracts. *Educational Researcher, 36*(3), 133–138.
King, G. (1997). *A solution to the ecological inference problem: Reconstructing individual behavior from aggregate data*. Princeton, NJ: Princeton University Press.
Kuh, G. D., & Whitt, E. J. (1988). *The invisible tapestry: Culture in American colleges and universities* (ASHE-ERIC Higher Education Report no. 1). College Station, TX: Association for the Study of Higher Education.
Knight, P. T., & Trowler, P. R. (2000). Department-level cultures and the improvement of learning and teaching. *Studies in Higher Education*, *25*(1), 69–83.
The Knowledge Network Organizing Tool (KNOT) software [Computer software]. Retrieved December 27, 2007, from http://interlinkinc.net/KNOT.html
Lakomski, G. (2001). Organizational change, leadership and learning: culture as cognitive process. *International Journal of Educational Management*, 15(2), 68-77
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
Lawrenz, F., & Huffman, D. (2006). Methodological pluralism: The gold standard of STEM evaluation. *New Directions for Evaluation*, *109*, 19–34.
Lynch, E., & Medin, D. (2006). Explanatory models of illness: A study of within-culture variation. *Cognitive Psychology, 53*(4), 285–309
Martin, J. (2002). *Organizational culture: Mapping the terrain* (Foundations for organizational science). Thousand Oaks, CA: Sage.
McAlpine, L., & Weston, C. (2002). *Reflection: Issues related to improving professors' teaching and students' learning*. In N. Hativa & P. Goodyear (Eds.), Teaching thinking, beliefs, and knowledge in higher education (pp. 59–78). Dordrecht, Netherlands: Kluwer.
Medin, D. L., & Atran, S. (2004). The native mind: Biological categorization and reasoning in development and across cultures. *Psychological Review*, *111*(4), 960–983.
Mertens, D. M., & Hopson, R. K. (2006). Advancing evaluation of STEM efforts through attention to diversity and culture. *New Directions for Evaluation*, *109*, 35–52.
Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.
Millar, S. B. (2003). Effecting faculty change by starting with effective faculty: Characteristics of successful STEM education innovators. In National Research Council, *Improving undergraduate instruction in science, technology, engineering, and mathematics: Report of a workshop* (pp. 101–117). Washington, DC: National Academies Press.
Mintz, S. (1986). Sweetness and power: The place of sugar in modern history. New York: Penguin Books.
Mohammed, S., Klimoski, R., & Rentsch, J. R. (2000). The measurement of team mental models: We have no shared schema. *Organizational Research Methods*, *3*(2), 123–165.
Narayanan, V. K., & Armstrong, D. J. (Eds.). (2005). *Causal mapping for information systems and technology research: Approaches, advances, and illustrations*. Hershey, PA: Idea Group.
National Academy of Education. (1999). *Recommendations regarding research priorities: An advisory report to the National Educational Research Policy and Priorities Board*. Washington, DC: Author.
National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.

National Research Council. (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

Patton, M. (2006, October). *Evidence-based evaluation findings using systems change and complexity science frameworks and ways of thinking*. Paper presented at the MSP Evaluation Summit II, Minneapolis, MN.

Polkinghorne, D. E. (1995). Narrative configuration in qualitative analysis. In J. A. Hatch & R. Wisniewski (Eds.), *Life history and narrative* (pp. 5–23). London: Falmer Press.

Porter, S. R., & Umbach, P. D. (2001). Analyzing faculty workload data using multilevel modeling. *Research in Higher Education, 42*(2), 171–196.

Resnick, L. B. (1991). Shared cognition: Thinking as social practice. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 1–22). Washington, DC: American Psychological Association.

Robinson, W. S. (1950). Ecological correlation and the behavior of individuals. *American Sociological Review*, *15*, 351–357.

Romney A. K., Batchelder W. H., & Weller S. C. (1987). Recent applications of cultural consensus theory. *American Behavioral Scientist, 31*(2), 163–177.

Rumelhart, D. E. (1980). *Schemata: The building blocks of cognition*. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33–58). Hillsdale, NJ: Lawrence Erlbaum.

Ryan, G. W., & Bernard, R. (2003). Techniques to identify themes. *Field Methods, 15*(1), 85–109.

Schank, R. C., & Ableson, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Lawrence Erlbaum.

Schon, D. A. (1990). Educating the reflective practitioner: Toward a new design for teaching and learning in the professions. San Francisco: Jossey-Bass.

Schroeder, C. M. (2001). *Faculty change agents: The individual and organizational factors that enable or impede faculty development participants' involvement in organizational change*. Unpublished doctoral dissertation, University of Wisconsin–Madison.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 249–284). New York: Academic Press.

Seymour, E. (2005). *Partners in Innovation: Teaching Assistants in College Science Courses*. Lanham, MD: Rowman & Littlefield.

Seymour, E., & Hewitt, N. (1997). *Talking about leaving: Why undergraduates leave the sciences.* Boulder, CO: Westview Press.

Shavelson, R. J., Ruiz-Primo, M. A., & Wiley, E. W. (2005). Windows into the mind. *Higher Education, 49*, 413–430.

Shore, B. (1996). *Culture in mind: Cognition, culture, and the problem of meaning*. New York: Oxford University Press.

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

Smircich, L. (1983). Concepts of culture and organizational analysis. *Administrative Science Quarterly, 28*, 339–358.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research, 72*(3), 387–431.

Stake, R. (1994). Case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 236–247). Thousand Oaks, CA: Sage.

Stensaker, B., & Norgard, J. D. (2001). Innovation and isomorphism: A case-study of university identity struggle 1969–1999. *Higher Education, 42*(4), 473–492.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
Strauss, C., & Quinn, N. (1998). *A cognitive theory of cultural meaning*. Cambridge, UK: Cambridge University Press.
Sunal, D. W., Hodges, J., Sunal, C. S., Whitaker, K. W., Freeman, L. M., Edwards, L., et al. (2001). Teaching science in higher education: Faculty professional development and barriers to change. *School Science and Mathematics, 101*(5), 246–257.
Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches* (Applied Social Research Methods Series Vol. 46). Thousand Oaks, CA: Sage.
Thorn, P. M. (2003). *Bridging the gap between what is praised and what is practiced: Supporting the work of change as anatomy and physiology instructors introduce active learning into their undergraduate classroom.* Unpublished doctoral dissertation, University of Texas, Austin.
Tienson, J. L. (1990). An introduction to connectionism. In J. L. Garfield (Ed.), *Foundations of cognitive science: The essential readings*. New York: Paragon House.
Tobias, S. (1992). *Revitalizing undergraduate science. Why some things work and most don't*. Phoenix, AZ: Research Corporation.
Trowler, P. R., & Cooper, A. (2002). Teaching and learning regimes: Implicit theories and recurrent practices in the enhancement of teaching and learning through educational development programs. *Higher Education Research & Development, 21*(3), 221–240.
U. S. Office of Science and Technology Policy. (2006). American Competitiveness Initiative: Leading the world in innovation. Washington, DC: U.S. Government Printing Office.
Umbach, P. D., & Porter, S. R. (2002). How do academic departments impact student satisfaction? Understanding the contextual effects of departments. *Research in Higher Education, 43*(2), 209–234.
Van Maanen, J., & Barley, S. (1985). Cultural organization: Fragments of a theory. In P. J. Frost, L. F. Moore, M. R. Louis, C. C. Lundberg, & J. Martin (Eds.), Organizational culture (pp. 31–53). Beverly Hills, CA: Sage.
Villachia, S.W., Lohr, L.L., Summers, L., Lowell, N., Roberts, S., Javeri, M., Hunt, E., Mahoney, C., & Conn, C. (2001). A cognitive map of human performance technology: a study of domain expertise. *24th Annual proceedings of selected research and development papers presented at the convention of the Association for Educational Communications and Technology,* Atlanta, GA, November 8-12, 2001. ED470122
Walczyk, J. J., Ramsey, L. L., & Zha, P. (2007). Obstacles to instructional innovation according to college science and mathematics faculty. *Journal of Research in Science Teaching, 44*(1), 85–106.
Weller, S. (2007). Cultural consensus theory: Applications and frequently asked questions. *Field Methods, 19*(4), 339–368.
Wolf, E. (1982). *Europe and the people without history*. Berkeley: University of California Press.