Integrating the Qualitative and Quantitative in Education Policy Research

Patricia Burch, Carolyn Heinrich, Caitlyn Farrell, Annalee Good & Molly Stewart

Authors' Note

This research was funded by the Institute for Educational Sciences, PR/Award number: R305A090301, Education Policy, Finance, and Systems Research Program, Goal 3.

Abstract

The term “mixed method research” has become lodged in the lexicon of policy research. However, the “how” of organizing this type of research remains relatively hidden. Beneath the specifics of doing mixed method research lie enduring ontological questions – even if qualitative and quantitative designs can be mixed, should they be? What are the benefits and costs in terms of the rigor and relevance of the research? Based on a rich description of an on-going mixed method study of after-school tutoring in five districts, this study examines organizational strategies for integrating experimental, correlational, and other forms of disciplined inquiry. We argue that these strategies are integral to strengthening the quality and validity of research and information made available to stakeholders.

Keywords: Mixed method research, integrated design, education policy evaluation, supplemental educational services (SES); after-school tutoring

Integrating the Qualitative and Quantitative in Education Policy Research

Many impact evaluations include both quantitative and qualitative investigations of program effects, including exploration of implementation and its relationship to program impacts (Chatterji, 2005; Greene, 2007; Greene & Caracelli, 1997; Greene, Caracelli, & Graham, 1989; Madey, 1982; Tashakkori & Teddlie, 1998, 2003; Weiss, 1998). Mixed method research is defined as "the class of research where the researcher mixes or combines quantitative and qualitative research techniques, methods approaches, concepts, or language into a single study" (Johnson & Onwuegbuzie, 2004, p. 17). The underlying premise of mixed method studies is that qualitative and quantitative methodologies offer distinct yet complementary advantages (Collins, Onwuegbuzie, & Sutton, 2006; Greene et al., 1989; Johnson & Onwuegbuzie, 2004; Johnson & Turner, 2003; Onwuegbuzie & Leech, 2005; Rossman & Wilson, 1985; Tashakkori & Teddlie, 2003). In particular, a key advantage of qualitative studies is the opportunity to explore policy issues in greater levels of depth and detail, understanding the role of place, time, practice, and processes (Denzin & Lincoln, 2011; Marshall & Rossman, 2011), while a key advantage of quantitative research is working with a larger sample of the population, giving the researcher the statistical power to look at effects and empirical associations among intervention variables and outcomes (Campbell & Stanley, 1963; Crotty, 2003; Murnane & Willett, 2011).

Under certain conditions, the use of mixed methods offers education policy researchers distinct advantages (Fry, Chantavanich, & Chantavanich, 1981; Jang, McDougall, Pollon, Herbert, & Russell, 2008; Johnson & Christensen, 2010). However, for those interested in applying or improving the method, there is little guidance for how to think in organizational terms about mixed method research or more specifically about the ways in which the logistics of mixed methods play out at the organizational level. There is an extensive literature on the role of

organizations in education policy reform and education policy research (e.g., see Coburn, 2004; Honig, 2006a; Rowan & Miskel, 1999). The irony is that as a field, we rarely turn the organizational lens on the nature of our own work process. This absence is curious given that a key theme of policy implementation research is the importance of organizational setting and strategy in accomplishing reform goals (Honig, 2006b; McLaughlin, 1987, 2006).

This paper aims to fill this gap in the literature, offering a way of thinking organizationally about mixed method research. We use elements of an organizational lens as a way of framing the elements of what is "unseen" at first glance about this kind of work. As described in more detail below, by an organizational lens on mixed method research, we specifically mean focusing on the environment, structures, boundary spanners, culture and subcultures that researchers knowingly or unknowingly employ in their research. We begin by discussing why policy research is deeply organizational and then discuss the organizational theory of action behind our own mixed method study.

Why an Organizational Perspective?

Following Bardach (2000, p. xiv), we believe that policy research is a social and political activity. Policy researchers rarely work alone; they have clients, audiences, and funders interested in their work. The individuals with a stake in the research will cross sector (e.g., policymakers and practitioners) and disciplinary (e.g., psychology and economics) boundaries. For example, policy researchers write journal articles. However, they also conduct briefings, write op-ed pieces and reports, produce webinars, and so forth. Policy research is deeply organizational in the sense that it is about doing research on contemporary issues that affect large numbers of people; it is about doing research that involves cooperation and inevitably power and conflict. It is about cooperatively setting goals, developing collaborative structures, managing

conflict and ambiguity, learning by and through experience, and using this information to strengthen the organization of the study and its outcomes.

In this paper, we describe our use of mixed methods research in an on-going evaluation of five districts implementing Supplemental Educational Services (SES) under No Child Left Behind.[1] SES is a federal education policy mandating after-school tutoring for schools not making adequate yearly progress on school test scores. We are conducting the SES research in five urban school districts in four states, representing different student demographics: Milwaukee, Wisconsin; Minneapolis, Minnesota; Chicago, Illinois; and Austin and Dallas, Texas.

Rationale for Doing Mixed Methods Research

Our analysis of the existing mixed method research convinced us of its appropriateness in light of our complex study objectives. We wanted to combine analysis of the policy with processes for providing feedback to constituents during implementation. Rather than be purely objective observers, we wanted to be partners in helping the tutoring providers, parents, and government officials obtain access to data needed to improve their program and to act as a broker of information between different stakeholders (e.g., across different districts and/or between providers and districts). However, we did not want this goal to replace or serve as an exchange for disciplined inquiry. To counter this pull, we had to avoid an eclectic, disjointed approach to mixed method research. In practical terms, we needed a clear rationale, logical organization, focused timeline, resources, and the know-how to make it work.

For over two decades, scholars within and outside of education have pursued mixed method studies and offered useful insights on the method (Creswell, 2003; Creswell & Plano Clark, 2007; Greene, 2007; Greene & Caracelli, 1997; Greene et al., 1989; Jick, 1979; Johnson

& Christensen, 2010; Johnson & Onwuegbuzie, 2004; Johnson, Onwuegbuzie, & Turner, 2007; Rossman & Wilson, 1985; Sieber, 1973; Tashakkori & Teddlie, 1998, 2003). Twenty years ago, mixed-methods evaluations were starting to gain popularity, but there was very little scholarly work on mixed-methods design or implementation.

Our article expands on this work in several key ways. First, we incorporate the findings from the subsequent twenty years of publication from a growing field of mixed-methods research, with a multi-disciplinary approach, drawing from education scholarship, mixed-methods research, and evaluation work. Second, rather than draw themes from others' studies, we ground our findings within the context of current empirical work, exploring the critical challenges and benefits that arise in a large-scale mixed methods study. Third, we expand on what it means to integrate mixed-methods components from early design through data collection, analysis, and interpretation in a far more detailed way. Finally, based on our practical experience, we offer a model of organizational strategies that can support integrated mixed methods work.

Identifying Alternative Designs

Mixed methods research is a broad umbrella term for a range of approaches that might be organized in various ways (Benz & Newman, 1998; Teddlie & Tashakkori, 2006). On one end of the continuum is the parallel model. Under the parallel model, qualitative and quantitative methods are run separately and simultaneously (Greene & McClintock, 1985; Louis, 1981; Onwuegbuzie & Leech, 2004; Yin, 2006). For example, in a study of the Rural Experimental Schools programs, qualitative researchers conducted ethnographies at the local sites while the quantitative researchers designed and analyzed survey results completely independently; analysis occurred separately and was used as "secondary data" for the other (Louis, 1981).

In component designs, different methods remain discrete through data collection and analysis, and the mixing of the methods takes place at the level of interpretation and inference (Caracelli & Greene, 1997). Heilig & Darling-Hammond (2008) complemented quantitative student achievement results with findings from interviews with students and staff to understand the high-stakes testing environment in Texas schools. Each kind of data was collected and analyzed separately, while the findings blended the two types of information.

In contrast to both parallel and component designs, fully integrated mixed methods designs such as ours combine methods throughout inquiry -- from research design and data collection to analytic processes and interpretation (McConney, Rudd, & Ayres, 2002; Onwuegbuzie & Leech, 2004). For instance, Day, Sammons, & Gu (2008) designed a four-year integrated mixed methods study on teacher effectiveness that tracked 300 teachers in 100 English schools. Through the conceptual and methodological integration, they were able to arrive at nuanced, robust, and "synergistic" understandings of the connections between the lives of teachers and their perceived and measured effectiveness.

The perceived challenges of mixed-method research map onto the enduring debates existing in the wider research community. One debate concerns the grounding of the data and the nature of inquiry. For example, Sale, Lohfeld, & Brazil (2002) argued qualitative and quantitative methods come from fundamentally different paradigms, with their own assumptions concerning reality (ontology), knowledge of that reality (epistemology), and the particular ways of knowing that reality (methodology). Based on these assumptions, "the two methods do not study the same phenomena" (Sale et al., 2002, p. 44). Each method is weakened or incorrectly employed in a mixed method design if both are used to study the same phenomenon. Further, the authors reported that given that qualitative and quantitative methods cannot study the same

phenomenon, we would expect to see much more variation within the results; the extent of disagreement between qualitative and quantitative results is rarely reported, however (Sale et al., 2002, p. 47). From this perspective, one cannot really know what is represented – more qualitative or more quantitative in the results.

Given the range of approaches to mixed methods research and the challenges outlined above, our rendering of the process needs to be more detailed and microscopic than one would find in the handful of paragraphs in which research design sometimes appears in published articles. This level of detail brings out the complexity of what lies in and beyond mixed method research, the relationships between different tasks, and how our approach was both initially defined and evolved during enactment and interpretation. This detail is also critical in identifying the organizational structures and processes that knitted together rigorous qualitative and quantitative work.

Framework for Integrated Mixed Methods Research

As Figure 1 suggests, the pillars of the framework are three sets of variables that are fundamental to an integrative mixed methods design. We refer to them as boundary spanners because while assuming different forms, they are integral to integrating the qualitative and quantitative approaches.

[INSERT FIGURE 1]

These organizational components, labeled inputs, include (i) the environment of the study; (ii) the structures and functions -- both planned and emergent -- that are believed suitable for accomplishing the purpose of the study; and (iii) the cultures and subcultures of an organization (e.g., what is taken for granted in the study; what the study participants -- what researchers, funders, participants -- say they value as expressed in formal and informal ways; the

conversations that occur within different communities). In addition to the traditional analysis of the impact on policy and practice of the study, our process outcomes also include optimizing the sample, developing sensitive instruments, ensuring construct validity, and working towards stakeholder satisfaction. These organizational inputs and study outcomes span across the four stages of the scientific research process – design, data collection, analysis, and reporting and dissemination. Collectively, Figure 1 represents our basic approach to integrated mixed methods research.

Study Background

Several propositions frame how we conceptualize the core influences in the study design and how we further organize the work within the stages of research: research design, data collection, data analysis, and reporting and dissemination. As researchers, we could and should be reflective on the relationship between our planned objective and actual activities by collecting information about these processes: how we were doing the research, how we were communicating the research, and how we were (or were not) learning from our own research. While we intend to break new ground with our research, we remain cognizant and alert to enduring biases and our own underlying assumptions about how research should be conducted. Finally, rather than a static backdrop, the environment is a central influence on our research – a set of dynamic variables that mediate our research design and which we as researcher "act upon" in conducting our studies.

Political-Legal Environment

The No Child Left Behind (NCLB) Act, signed into public law in 2002 to close the achievement gap in public education, specifically requires public schools that have not made adequate yearly progress for at least two consecutive years to offer parents of children in low-

income families the opportunity to receive extra academic assistance, or supplemental educational services (SES). Consistent with the intent of the law to promote "accountability, flexibility, and choice," SES is implemented at the local level and draws on the private sector to offer eligible students a range of choices for free tutoring outside of regular school hours. Although no new federal monies were allocated along with this mandate, the law lays out criteria and guidelines for state and local educational agencies in approving SES providers, arranging for their services, managing contracts and financial systems, and to an extent, monitoring providers' performance. School districts are obligated to set aside 20% of their Title I funding for SES and to measure provider effectiveness in increasing student achievement.

In theory, accountability for SES is supposed to be realized primarily through the exercise of choice by parents and students who ostensibly use information distributed by local educational agencies and SES providers to identify the best provider to meet the children's needs. Students who become aware of their SES eligibility may choose to register for SES with a specific SES provider, and SES providers invoice the school district for the number of hours of SES students attend, up to a maximum per-student dollar allocation. The service agreement between a district and its SES providers is, effectively, a cost-reimbursement contract, with no performance contingencies. Only state educational agencies have program design authority, for example, to approve SES providers and establish program criteria, such as an acceptable student/tutor ratio for providers to meet.

The conditions of the political-legal environment motivated us to use mixed methods. In particular, the intent of the SES policy was to maintain maximum flexibility in programming, offering providers, tutors, and district administrators little guidance in the "how" of tutoring. Prior to this cross-site study, the co-principal investigators had conducted a mixed method pilot

study of SES in one district. We learned from the pilot data that despite the laissez-faire design of the law, all stakeholders were eager for effects data. We also learned from the pilot study that district administrators wanted concrete guidance on how to address immediate local challenges. These needs required rich descriptive data gathered from multiple district settings. It also required that we connect the dots between quantitative data and organizational processes. For example, through both observations and quantitative data analysis, we saw connections in certain districts between enrollment trends and differing registration processes, such as requiring parents to register online. These types of factors confirmed that we needed both qualitative and quantitative data to respond to the information needs of multiple stakeholders working in diverse settings.

Organizational Components to Mixed Methods Research

In light of our objectives, the co-PIs regularly met to discuss and develop goals, the theory of action, structures, and functions/roles to support our integrated mixed method work. Across the different stages of the research process, we discuss key aspects of each of these organizational components.

Research Design

We created a theory of action for our research design that has two interrelated dimensions. The first dimension is the basic research design within each method. The second dimension is the design for how we integrate qualitative and quantitative methods. The bottom line goal along both dimensions is to enhance the quality and usefulness of our research.

Within-method design. One of the first steps in the qualitative study design was to select the sample of SES providers. Provider characteristics used in our initial sample selection included: high market share; high attendance levels relative to other providers in the same

district; two or more years providing SES in the district; and equal sampling among on-line, in-home, in-school, and community-based tutoring, as well as among for-profit, not-for-profit, district-provided, and, when applicable, faith-based organizations. When possible, we also attempted to include providers that advertised that they targeted English Language Learner (ELL) populations and students with disabilities.

Qualitative data were collected from district, state, and provider records. We also interviewed district and state administrators, directors of tutoring programs, and tutoring staff. We held parent focus groups and observed tutoring practice. Observations of tutoring sessions (n=94) using a classroom observation instrument were designed to capture key features of instructional settings and to more accurately see the nature of the SES intervention. The instrument has the capability of not only providing descriptive information on facilities, instructional materials, and teaching methods in use during sessions but also detecting the relationship between different kinds of format (e.g., online, in-home), resources (e.g., curriculum materials, staffing), and instructional methods with students' observed levels and types of engagement.

For the quantitative component, we estimated the effects of SES as measured by changes in students' standardized math and reading scores by the different levels of SES attended; by district and school type (elementary, middle, or high); and by comparing only SES attendees while controlling for their probability of registration, attendance, and other characteristics. We used propensity score matching methods to account for selective differences between those who register for SES but do not attend and those who attend lower vs. higher levels of SES.

Cross-method design. The second design dimension concerns the integration of the qualitative and quantitative methods. Our orientation to this work is that we seek a balance

between planned and responsive design. Put differently, we developed a blueprint for integration, with the anticipation that this blueprint might evolve as we learned from the work. This blueprint depicts integration as occurring across critical stages of the work -- research design, data collection, analysis, and interpretation -- as displayed in Figure 1. We used the stage model as a heuristic for organizing our integration, defining a specific task and goal for each stage.

The first phase of qualitative study in year one was designed to see "behind the curtain"—what was actually happening on the ground—to help inform what was in the intervention. We documented the different varieties of tutoring approaches in each of the five districts as a means for providing the quantitative team with a lens on potential variables to include in the model for estimating effects. We also expect the quantitative work to inform our fourth year qualitative sample. In year four of the study, we will conduct field research in the settings that the quantitative analysis had identified as promising sites, as measured by provider effects on changes in standardized test scores. The qualitative work in the fifth year will be oriented towards testing and refining quantitative findings, examining the processes beyond variables (e.g., intensity of hours) identified through quantitative estimation as contributing to effects. The qualitative work also will look at the implementation of practices identified by the quantitative research as promising.

Data Collection

Data collection as well as analysis for the study is structured around two distinct work groups. The co-principal investigators each oversee a different work group – the qualitative work group and the quantitative work group – based on their expertise. Each work group meets regularly and collectively designs instrumentation within the group. Roles are differentiated within each work group by site.

Each of the five districts has two site coordinators, one qualitative and one quantitative, although the responsibilities of the qualitative and quantitative coordinator for the same site are distinct based on required research tasks. For example, the Milwaukee site coordinator for the qualitative team is the point person for the interviews, observations, and focus groups with the site. The quantitative Milwaukee coordinator organizes access to student records and other data for the district and is responsible for cleaning and analyzing the data for that site in coordination with other quantitative team members.

In the planning of our research, we designated different individuals and groups with responsibility for core functions. For the PIs, this include (a) initiating the structure of research and the organizational supports, (b) training, (c) monitoring and coordinating the work, (d) supervising and conducting analysis, and (e) disseminating results. Site coordinators noted above act as point people for data collection as well as analysis, reporting to PIs, and they communicate with the other researchers within their work group. The qualitative and quantitative teams are responsible for data collection, development of instrumentation, problem solving around sample selection and data quality, and on-going goal formulation. For example, the qualitative team had extensive conversations to select indicators for the observation instrument that were connected with the body of literature on instructional quality but also appropriate to the distinct instructional setting of SES. The team also worked together to develop a manual with descriptions and rationale for each indicator. The quantitative team drew heavily from their interactions with the qualitative team – specifically, the qualitative documentation and findings on the organization of SES within school districts – to determine an appropriate and consistent way of coding the key treatment measures of the number of hours and types of SES received by students.

Data Analysis

Early on, we realized that integrated data analysis would require certain tools and a culture that encouraged sharing within and across work groups. In our communications, we benefit from the use of intra-net and internet web technology. We use an intra-net database to securely upload and provide immediate but restricted access to different kinds of data from remote locations. For example, curriculum materials collected during observations of tutoring sessions are scanned and uploaded to the database, then linked to coding documents during the analysis process. These documents are accessible to all researchers in the project. We also created a public website that provides an information portal for district partners, providers, and researchers interested in our work.

In our design, we were alert to the fact that despite common research questions, each work group could be expected to shed more or less light on different aspects of the phenomenon. For example, in both the qualitative and quantitative components, we are interested in the quality of instruction. A key anticipated advantage of the quantitative data on student achievement is offering the required statistical power for estimating effects of the SES treatment groups. A key advantage of the qualitative data is in describing the instructional phenomena or treatment as performed by providers in the qualitative sample to answer the question: What is in the treatment being measured?

At the time of design, we did not see these distinct advantages as leading to divergent paths. The organization of integrated tasks by stage helped illuminate the points of departure in our work, as described in greater detail below in the section on constructive challenges. Not every aspect of our work could be fully integrated. There were times when each team needed to pursue its own line of inquiry, freed from the time constraints and operating assumptions of the

"other" paradigm. We began to see these moments of divergence as necessary forks in the road of our integrated path. We had to move apart at times and do our work in depth so that we could come back together again with rich data on different aspects of the intervention. With these findings viewed in combination, we were able to capture the complexity of the phenomenon and its effects.

Reporting and Dissemination

The design for integration extended through data collection and analysis and into reporting and dissemination. We developed a timetable that included common deliverables to the funder as well as formative feedback sessions for districts and providers. Work groups and PIs prepared and presented co-authored paper findings at professional associations and conferences.

As we moved into the research, we recognized the need for formal structures of communication that involved frequent communication between the two PIs leading the qualitative and quantitative components and among site coordinators across different district settings. The next section details the processes that led to the recognition of this need, as well as our various strategies for adequately addressing it.

Constructive Challenges

Next, we discuss several constructive challenges of our work. We identify them as constructive challenges because they represent examples of specific difficulties from which we generated important insight and future directions for our work.

Challenge #1: Sample Selection

Consider the collaboration among the quantitative and qualitative research team when it comes to choosing a sample selection. In qualitative research, sample sizes tend to be small so that research can understand phenomenon in depth and in context. The challenge was how to

select a sample with so many vendors providing services in a district; in one of our sites, vendors numbered in the hundreds. By law, states and districts are responsible for evaluating providers and their compliance. Our district partners were specifically interested in obtaining information on providers serving the most students, that is, those providers receiving the most revenue out of the Title I set-aside. If they had to tighten the weak links, they wanted to start where aggregate costs and benefits were the greatest. In sampling selection, the qualitative team had been working with SES district directors to acquire relevant information (e.g., active providers, tutoring formats). However, the team soon learned that particularly in districts where registration systems had not been automated or at least systematically organized, they lacked accurate and precise information about providers' market share.

The quantitative team was able to step in as a trusted broker of information. Through data sharing agreements, the quantitative team had access to district invoices from vendors providing SES and were using them in their quantitative analyses of SES effects. The quantitative team helped the qualitative team identify providers with market share in a given year, and ultimately, to see provider market shares across multiple years. Even by including only providers who served more than 100 students over multiple years, the potential sample for the qualitative work group remained high. The qualitative team then further narrowed the sample by matching providers with market share along program characteristics of central interest to the design of the study, including the format of providers, legal status (for profit or not for profit), and whether they provided services to students with special needs. Thus, working with the quantitative team, the qualitative team could select a sample that allowed it to take a close look at providers working with large numbers of students in a variety of formats. This information allowed the qualitative team to conserve resources and support more research in fewer sites.

Challenge #2: SES Participation

Nationally, up-take rates in SES have been low, initially hovering at about 16%. Over time, a participation problem has evolved; as more and more students are eligible, a smaller percentage of those eligible can actually take part due to limited district resources. Drawing on the choice literature (e.g., Hochschild & Scovronick, 2003), we might posit that low participation rates are, in large part, a function of differential access to, or quality/relevance of, information.

Integration of qualitative and quantitative methods in our research design suggested that aggregated participation rates belie a more complex story. The design of the law requires that districts provide parents with information on student eligibility. From focus groups with parents, we learned that some districts and providers offer vague information to parents, while others provide detailed information. For example, some districts have specific structures in place to increase enrollment and retention, such as school-based coordinators. Further, some parents highly value the incentives attached to participation such as computers or attendance rewards while others do not; in other sites, state law largely prohibit incentives and thus represent a restriction on providers' enrollment strategies. This information helped inform the design of estimation models for understanding students' selection into the program at different stages of the treatment. The quantitative team hypothesized that selection into SES participation occurs at four stages: awareness of eligibility, registration for the program, participating in at least one program hour (i.e., showing up for the program), and length of program participation (i.e., number of hours).

Working with data on students' eligibility, registration, and attendance of SES, the quantitative team explored patterns of SES participation. They found that rather than following a normal distribution, the number of hours that a student attended SES clustered around particular

thresholds or spikes in the data. With information on the hourly rate charged by providers for SES, the quantitative team was able to test their hypothesis that the spikes in the hours attended were associated with the hourly rate charged by providers. Drawing from interviews with providers, we explored how tutoring rates were set, learning that the hourly rate charged was considered a routine, programmatic task on the part of vendors. These rates were prescribed in advance and written into the state contracts, although not always set in stone, as we found out from some state coordinators.

As the quantitative team worked with program participation data in order to narrow sample selection, we encountered a puzzle. Each student was expected to register with one vendor. In the process of ordering these data, we discovered a significant number of students that had multiple records. In other words, they were recorded as simultaneously participating in tutoring with more than one vendor. The quantitative team used a student ID to link data, and for each ID there were multiple registration records, requiring alternative approaches to organizing the data. Based on tutor and provider interviews, as well as parent focus groups, the qualitative team provided insight into the reasons for multiple records and how they should be handled in data linking and analysis. For example, some parents and students were confused by the registration process and attempted to sign up for more than one vendor. In other instances, lured by incentives such as computers, students were signing up with more than one vendor. And students dissatisfied with their initial provider sometimes transferred from one provider to another in the middle of the school year. This information was important to the quantitative team because it helped them to make appropriate decisions in linking and analyzing data, such as eliminating duplicate records or restructuring the data to allow for analysis of students with multiple registrations and hours attended with multiple providers.

Challenge #3: ELL Participation

A central research question in our study concerned understanding how and whether English language learners (ELLs) and students with disabilities were being served under SES. In this line of investigation, we address some of the problematic aspects of gathering data on ELLs. From the outset, we deliberately selected several districts serving high percentages of English language learners. The classroom observation instrument developed by the qualitative work group was designed to investigate (a) participation by ELL students, (b) the presence or absence of staff/instructors who were bilingual and/or trained to work with ELL students, and (c) the extent to which curriculum and/or instruction was differentiated.

However, we have faced frequent challenges in our data collection, as we were unable to easily identify which students were ELL while doing classroom observations; the category "ELL" only denotes students who have not reached a particular level of proficiency with English and does not refer to all students who are technically learning English. Tutors and directors reported in interviews that they seldom possessed this information in a formal way. In other words, they knew that students with limited English were present in the classroom, but they were not sure whether these students were formally categorized as ELL and whether or not these students needed differentiated instruction. District administrators across several sites also were sometimes unable to retrieve this information. The qualitative principal investigator discussed this problem with the quantitative principal investigator. The quantitative work group had been running data on ELL participation and had discovered that ELL students were more likely to sign up for SES. And if they signed up, they were more likely to attend more than one session. Thus, ELLs were participating but were not easily identifiable to the qualitative team or to instructors.

As a result, the qualitative team discussed ways to improve the sensitivity of our data collection process. The qualitative team moved towards a proactive approach to understanding why this information was often unavailable to tutors, asking questions in interviews about data access. In addition, the qualitative work group used this information to strengthen the sensitivity of its instrumentation in our parent focus groups. For example, our original protocol had the question, "Please tell me about your families' experiences with the tutoring services, if English is not your child's primary language. Given these experiences, what changes if any would you suggest?" The team then added specific follow-up questions such as, "How did you and your family learn about SES? Did your child's tutor know English is not their primary language?"

The school districts participating in our study include information in student records that identifies students as ELL. In some cases, such as in Texas where the proportion of ELLs in the student population is large, the measures existed in finer gradients. In this case, it was important for the quantitative team to ensure that our definitions of ELL as codified in the data were consistent across sites. Obtaining quantitative data on numbers of ELLs participating in SES was a positive development for strengthening instrumentation of the qualitative group. The data ignited an interest in finding better ways to see what SES tutoring for ELL students looked like in action.

Challenge #4: Defining the Invoiced Hour

By design, the quantitative team had identified an invoiced hour of service as a key treatment measure. In other words, we measure the dosage/intensity of the treatment on the basis of the hours of SES for which providers invoiced the district. In light of existing research on after-school programming and given the evidence of significant variation in students' participation, the quantitative research team explored the relationship between invoiced hours per

student and achievement gains. In running early estimations, the quantitative team discerned a pattern first within and then across districts. Positive effects primarily were discernible for students reaching a minimum threshold of hours attended: approximately 40 or more hours of tutoring.

However, the qualitative data suggested that the key treatment measure – an invoiced hour – may be an incomplete or even inaccurate measure. Providers often recorded students as attending a full hour, when in fact, as observed in qualitative research across sites, students often arrived late, left early, or experienced interruptions that caused them to receive less than the invoiced hour of tutoring. At the same time, some providers acknowledged that it was important to allow the students breaks or snacks to help them stay focused and engaged, sometimes after an already long school day. In other words, an invoiced hour may not really be a full hour of student instruction. Additionally, student exits or interruptions in service provision were not necessarily handled the same way by SES providers or across districts in accounting for SES hours attended.

Understanding this issue was essential in the quantitative interpretation of data. Part of the process of working through this puzzle was to understand better what we were measuring; to understand what districts saw as acceptable uses of an invoiced hour; and from an educator's perspective, to try to more accurately convey what fraction of an invoiced hour was actually instructional time.

Challenge #5: Format of Tutoring

As noted above, as part of our original design, we sought to understand whether and how different formats for tutoring were correlated with different kinds of effects. For example, holding other factors constant, was on-line tutoring more effective than home-based tutoring? There were publicly available, self-reported data on provider format: in their state applications,

providers would identify as a home-based, community-based, or school-based provider. They also would indicate whether they provided any on-line services. These data helped both qualitative and quantitative study teams distinguish between different kinds of providers that might otherwise appear very similar (e.g., in terms of market share or instructor qualifications). The quantitative team analyzed these data to begin to explore possible relationships between provider format and achievement effects.

Concurrently, the qualitative team was engaged in corresponding research in the field. Field researchers assigned to each district were conducting repeated observations of four to six providers per district sampled in part on the basis of format (i.e., at least one home-based, school-based, community-based provider per district). The descriptive qualitative data that we gathered from these observations helped us begin to develop a picture of how providers were approaching the work. For instance, were there any similarities in the practices applied among vendors that identified as home-based providers?

We rapidly learned that the reported formats did not always align with actual practice. Several home-based providers actually never went to a family's home or met a child's parents. They offered services in the library and, in a few instances, in schools. The majority of school-based providers did offer services in school settings. However, there were important exceptions, particularly as the program unfolded; a few school-based providers were moved out of school settings and relocated to libraries. Tutors working for home-based providers grew tired of the commute to the home and determined it was more economical to tutor after school in their own classroom.

This rich descriptive information proved useful for the quantitative team as it worked to develop and refine hypotheses and analyses. For example, at the beginning of our study, on-line

providers were gaining market share. Early calculations by the quantitative team suggested that on-line providers also were charging more. Given growing market share and higher than average hourly rates, the question arose as to what was the added value of this format? In the first set of analyses, the quantitative team found that as a group, on-line providers had minimum or negative effects on student achievement, particularly as compared to school-based providers. Drawing on observation and interview data, the qualitative team helped the quantitative team distinguish between substantive on-line providers and those providers who claimed to, but rarely did, use on-line instructional tools. The decision was made to recode these latter providers in estimating on-line provider effects. If an on-line provider was not doing the large majority of tutoring on-line, the provider should not be characterized as so in aggregated estimates of on-line provider effectiveness. In addition, the qualitative team alerted the quantitative team to differences within the on-line format, specifically between "live" on-line tutoring, where students interacted with tutors on-line in real time, and static, software-based tutoring, where students interacted with a software program. This led to plans for more nuanced quantitative analyses that will distinguish effects based on substantively different tutoring within the same general format.

Contributions of the Integrated Approach to These Challenges

From these challenges, we can identify three key areas where the integration advanced our research at the end of Year Two. First, it helped in sample optimization. The quantitative data helped create parameters to guide our selection of SES provider cases, including information on market share that helped to optimize our limited resources in site selection. Qualitative data from the field helped the quantitative team with the issue of multiple records, allowing the team to appropriate handle duplicate student records in the analyses.

Integration of methods also contributed to increased sensitivity and appropriateness of our instruments. With such a large sample size, we needed a standardized measure for gauging the intervention based on student participation. The quantitative team chose provider invoices that recorded student participation in hours as its standardized measure, given the size of the sample and availability of this measure across the districts. The qualitative team spent hours in the field, conducting observations of tutoring practice. The qualitative work sought to understand what was happening in an hour of SES in practice. From this work, we gained valuable perspective on what was "in" an invoiced vs. instructional hour. Behind the invoice lay a much more complicated story – incomplete record keeping, students leaving early or arriving late, tutoring time spent on non-instructional activities, technical/materials difficulties, and other issues. The quantitative research helped the qualitative team strengthen the classroom observation instrument. It confirmed our supposition that English language learners and students with disabilities were participating in the program, providing clarity on this issue when stakeholders often could not.

The integrated nature of our research also helped test and validate emerging interpretations. We designed the study based on the premise that the format of tutoring might be a mediating influence on students' participation rates, experiences in the program, and measured effects. As analysis progressed, the quantitative team provided confirming data on this hypothesis. Across districts, we began to see differences in effect sizes of on-line providers relative to school-based providers. The qualitative work provided an analytic perspective for refining these emerging hypotheses. The interviews and observation data revealed important differences within format. The continuum of on-line providers varied significantly: some self-described on-line providers employed virtually no instructional technology, and/or their

technology did not include an on-line component. This information was critical in supporting the accuracy and validity of our research.

Continuing Organizational Challenges

Our integrated research approach also generated ongoing organizational challenges. The first challenge involves coordination of activities around stakeholders' access to information. Both the qualitative and quantitative work depends on strong relationships with stakeholders, but the nature of these relationships differs. For example, the quantitative team established data sharing agreements with each district and needs to maintain a relationship with evaluation staff in the district, and the districts in turn have provided access to administrative and test score data. The qualitative team has agreements with multiple stakeholders including individual parents, providers, and district and state administrators. These agreements are based on the principle of individual and/or organizational voluntary consent. These differences generate challenges around the goal of sample optimization. At one point in the study, the quantitative team released findings that indicated that on-line providers had minimal or negative effects relative to school-based providers. Releasing this information was important to maintaining good relationships with the districts that were eager for the data.

However, as might be expected, on-line providers were not happy with the "no effects" findings. The providers whose sessions we were observing functioned as businesses and employers. Several on-line providers were worried that the findings would undercut their business. They voiced their concerns to the principal investigators about continuing with the study but ultimately remained in our sample. Maintaining relationships with the on-line providers is critical to the quality of data for qualitative team, which was following a cohort of providers over time.

The second continuing challenge concerns relationships between formal and informal structures. As part of our design of the qualitative work, the qualitative team developed a classroom observation instrument as one of a cluster of strategies for examining tutoring. Other strategies in the cluster included interviews, focus groups, and document analysis. At several junctures, the qualitative team found itself tilting towards the observation instrument as a dominant strategy. For instance, at times we prioritized analysis of this data over other forms of data when pressed by deadlines for developing presentations and disseminating findings. The qualitative team invested in technology that would allow us to digitize the data that we collected so that we could analyze the data rapidly and more easily align it to quantitative findings. We had planned to report our findings in an integrated rather than parallel manner.

However, the qualitative team needed to make sure that in the process, it did not assign lower status to the other important qualitative data, such as interview data. Interview data and field notes take a lot of time both to collect and analyze. Watching the quantitative team run what they referred to as "just-in-time analyses" (that is, data that could come in the night before and be analyzed in preliminary form the next day) made some team members impatient with the wordier data. And yet, it is the words – the voices of tutors, the words used by parents to describe tutoring, and our own unstructured notes on these processes – that the qualitative team returned to again and again when puzzled by the quantitative findings or when seeking to understand differences. The qualitative work group has to be vigilant about not getting so absorbed by the quantitative method that they lose the way with their own. Put simply, the risk of integrating without design and reflection is that methods and findings can become diluted. The quantitative team, alternatively, increasingly realized how important it was to draw from the qualitative team's work before getting too far in establishing cross-site approaches to model specification

and analysis. On many occasions, the quantitative team went back to the data with qualitative team insights in hand to reformulate the parameters for modeling and analysis in addressing the questions of who participates in SES and what do they gain.

Finally, although bound together by a set of common research questions, and committed to integrating methods, there are points where we recognize disparities emerging from multiple cultures and subcultures within the study. The quantitative team measures change in instructional practice primarily in terms of standardized test scores. The qualitative team seeks rich description of the intervention, including differentiation and evidence of culturally relevant and multi-modal practices. The qualitative team also seeks a deeper contextualized understanding of the quality of the intervention as parents, teachers, district administrators, and other stakeholders articulate and experience it.

In the first year of the study, the dominant patterns in our assessment of tutoring quality were much stronger than the patterns discerned in the quantitative data, which showed weak or minimal effects except in a few instances. How to best deal with these pattern differences in an integrated study remains, at the time of writing, an open question. We are currently pushing forward both with separate analyses and products while also continuing to work on integrated findings. The integration process continues to evolve as the qualitative team, requiring more time to synthesize across the various sources of data collected, progresses through their analyses. And rather than trying to reconcile or force connections between all findings, we maintain their differences when appropriate.

Concluding Thoughts

In summary, this paper relates how a group of researchers designed and are implementing an integrated mixed method study, including the principles underlying our work, the strategies

we developed, and how we think these strategies have improved the quality of our research in terms of sample optimization, instrument sensitivity, and verification of findings.

While it is method that we examine, the core themes are organizational. We found numerous insights in the literature on mixed method research, save the one thing that we as education policy researchers constantly conclude is important in our own studies: the organizational strategies behind the work that would allow it to span boundaries and transcend hardened polarities of qualitative and quantitative paradigms. We wanted to do mixed method research on the shoulders of what we already knew about organizations and change. Rather than treat organization as the mysterious backdrop to the lofty method, we decided to reverse the relationship and go into the organizational details of our own processes. This included specifying the theory of action in our work, the roles we assigned, and the structures that enabled cross talk and evidence-based exchange.

At the end of the day, we propose that the problems and possibilities of mixed method policy research are as much organizational as they are ontological. We base this argument on three themes emerging from our own mixed method study. These themes are consistent with the analytic frame on organizations outlined in the earlier part of this paper.

First, policy studies are embedded in organizations, and as such they are open systems. The quality of a study turns in part on the environment (e.g., funders' support for mixed method research) in addition to internal components. The environment not only acts on the study; members of the research group can act on the environment. For example, as the study progressed, we not only drew on existing technology, but we also advanced and transformed the technology. We responded to stakeholders' calls for policy and implementation guidance with studies and reports that reduced ambiguities and guided districts' and providers' policy

decisions. For example, a number of districts made program changes in response to our finding that effects of any significance on student achievement were not observed until students received approximately 40 hours of SES tutoring. In addition, some districts have responded to our findings on the instructional landscape by focusing efforts on facilitating better alignment between the day school and SES curriculum.

Second, in mixed method research, structure and capacity building for critical reflection and feedback is critical. Typically, we think of using information generated by the study to improve outcomes in the field. However, we can also use data and information from our research to improve the rigor, timeliness, and user-friendliness of the research. The frequency of our communication within work groups, the impetus for the meeting (focusing on data rather than management tasks), and the locus of decision-making (flat rather than hierarchical) contribute to a project culture that emphasizes crafting integrated approaches to the research.

These processes also provide an initial footing for us to reassess guiding beliefs and everyday practices in light of our goals. We talk about mixed method work but in a sense the very idea of mixing reinforces the notions that there are two very distinct paradigms – qualitative and quantitative – and the design of research should reflect those paradigms. From our work, we have seen first hand the concerns that unite our work, whether our methods are qualitative or quantitative. In the course of our interactions, we became more aware of our own taken-for-granted realities of separate paradigms. For example, why did we create separate qualitative and quantitative work groups?

Moving forward, we would like to take more seriously the idea that the design of our study does not need to be structured around differences in paradigm but instead rooted in structures, functions, roles, and standards that span methods. Our efforts to be both deliberate

and flexible in our own approaches suggested ways of organizing the work around common tasks, orientations towards organizations, and quality concerns.

That said, we also learned that at some points, the two teams needed to move on to different forks of the road – parallel tracks still united by common research questions and processes, but flexible enough to return to specific forms of inquiry, deduction, and verification specific to our method. Paradoxically, this hybrid approach to mixed method research allows us to more fully integrate on the macro level. Doing so creates the space to innovate within the mixed method approach, without straying too far from vision of research, where qualitative and quantitative researchers move in tandem, supported by structures that feed collaboration and exchange, towards insights that meet the highest principles of discipline inquiry and relevant useable research.

Notes

[1] The project is entitled "Supplemental Educational Services: Integrated Qualitative and Quantitative Study of Implementation and Impact ($SESIQ^2$)." The term, "IQ Squared," was developed in collaboration with Robert Meyer and Elizabeth Graue (see Meyer, Graue, & Author, 2007).

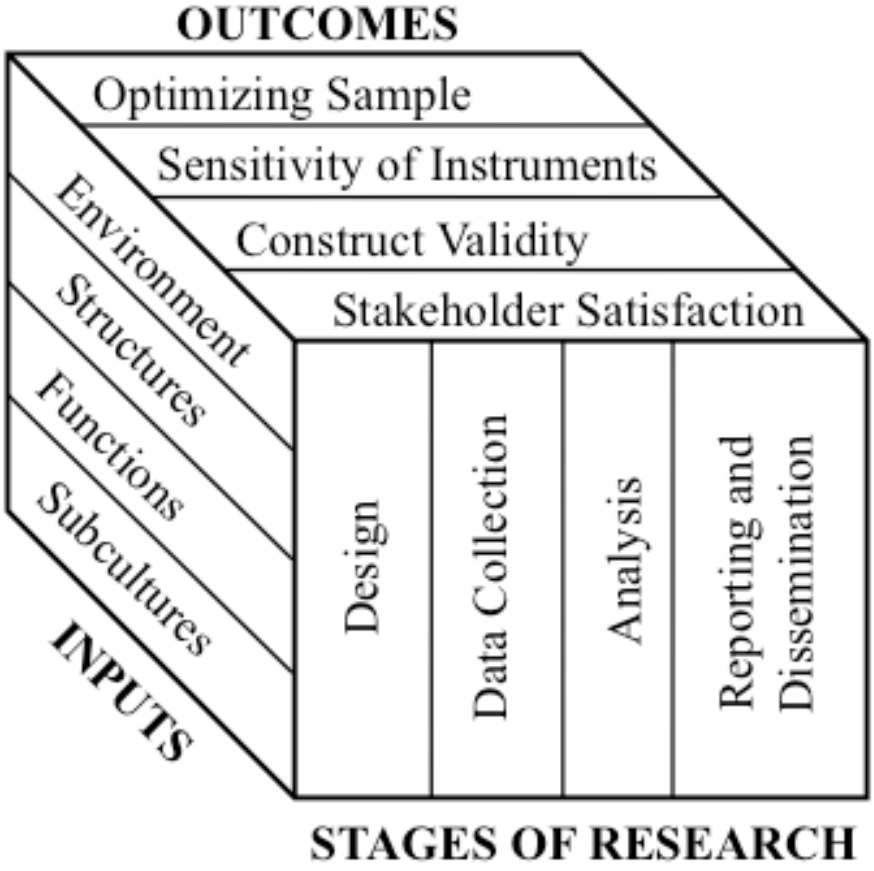

Figure 1. IQ squared: The basic approach to fully integrated mixed methods research.

Works Cited

Bardach, E. (2000). *A practical guide for policy analysis*. New York, NY: Seven Bridges Press.

Benz, C. R., & Newman, I. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Carbondale, IL: Southern Illinois University Press.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Caracelli, V. J., & Greene, J. C. (1997). Crafting mixed method evaluation designs. In J. C. Greene & V. J. Caracelli (Eds.), *Advances in mixed methods evaluation: The challenges and benefits of integrating diverse paradigms*. San Francisco, CA: Jossey-Bass.

Chatterji, M. (2005). Evidence of "what works": An argument for extended-term mixed-methods (ETMM) evaluation designs. *Educational Researcher, 34*(5), 14-24.

Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education, 77*(3), 211-244.

Collins, K. M. T., Onwuegbuzie, A. J., & Sutton, I. L. (2006). A model incorporating the rationale and purpose for conducting mixed methods research in special education and beyond. *Learning Disabilities: A Contemporary Journal, 4*, 67-100.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research.* Thousand Oaks, CA: Sage.

Crotty, M. (2003). *The foundations of social research: Meaning and prespective inthe research process*. Thousand Oaks, CA: Sage.

Day, C., Sammons, P., & Gu, Q. (2008). Combining qualitative and quantitative methodologies from research on teachers' lives, work, and effectiveness: From integration to synergy. *Educational Researcher, 37*(6), 330-342.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The SAGE Handbook of Qualitative Research*. Thousand Oaks, CA: Sage.

Fry, G., Chantavanich, S., & Chantavanich, A. (1981). Merging quantitative and qualitative research techniques: Toward a new research paradigm. *Anthropology & Education Quarterly, 12*(2), 145-158.

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey-Bass.

Greene, J. C., & Caracelli, V. J. (1997). Defining and describing the paradigm issue in mixed-method evaluation. *New Directions for Evaluation, 74*(1), 15-17.

Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274.

Greene, J. C., & McClintock, C. (1985). Triangulation in evaluation: Design and analysis issues. *Evaluation Review, 9*, 523-545.

Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-style: The progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis, 30*(2), 75-110.

Hochschild, J. L., & Scovronick, N. (2003). *The American dream and the public schools*. New York, NY: Oxford University Press.

Honig, M. I. (2006a). Street-level bureacracy revisited: Frontline district central office administrators as boundary spanners in education policy implementation. *Educational Evaluation and Policy Analysis, 28*(4), 357-383.

Honig, M. I. (Ed.). (2006b). *New directions in education policy implementation: Confronting complexity*. Albany, NY: The State University of New York Press.

Jang, E. E., McDougall, D. E., Pollon, D., Herbert, M., & Russell, P. (2008). Integrative mixed methods data analytic strategies in research on school success in challenging circumstances. *Journal of Mixed Methods Research, 2*(3), 221-247.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly, 24*(1), 602-611.

Johnson, R. B., & Christensen, L. B. (2010). *Educational research: Quantitative, qualitative, and mixed approaches* (4th ed.). Thousand Oaks, CA: Sage.

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

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

Johnson, R. B., & Turner, L. A. (2003). Data collection strategies in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 297-319). Thousand Oaks, CA: Sage.

Louis, K. S. (1981). *Policy researcher as sleuth: New approaches to integrating qualitative and quantitative methods*. Paper presented at the American Educational Research Association.

Madey, D. L. (1982). Some benefits of integrating qualitative and quantitative methods in program evaluation, with illustrations. *Educational Evaluation and Policy Analysis, 4*(2), 223-236.

Marshall, C., & Rossman, G. B. (2011). *Designing qualitative research* (5th ed.). Thousand Oaks, CA: Sage.

McConney, A., Rudd, A., & Ayres, R. (2002). Getting to the bottom line: A method for synthesizing findings within mixed-method program evaluations. *American Journal of Evaluation, 23*(2), 121-140.

McLaughlin, M. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis, 9*(2), 171-178.

McLaughlin, M. (2006). Implementation research in education. In M. I. Honig (Ed.), *New directions in education policy implementation: Confronting complexity*. Albany, NY: State University of New York Press.

Meyer, R., Graue, E., & Author. (2007). An integrated qualitative and quantitative study of the SAGE program. Proposal to Wisconsin Department of Public Instruction. Madison, WI.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. New York, NY: Oxford University Press.

Onwuegbuzie, A. J., & Leech, N. L. (2004). Enhancing the interpretation of "significant" findings: The role of mixed methods research. *The Qualitative Report, 9*(4), 770-791.

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

Rossman, G. B., & Wilson, B. L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. *Evaluation Review, 9*, 627-643.

Rowan, B., & Miskel, C. C. (1999). Institutional theory and the study of educational organizations. In J. Murphy & K. S. Louis (Eds.), *Handbook of research on educational administration* (2nd ed., pp. 359-384). San Francisco, CA: Jossey-Bass.

Sale, J. E. M., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality & Quantity, 36*, 43-53.

Sieber, S. D. (1973). The integration of fieldwork and survey methods. *American Journal of Sociology, 73*(1), 1335-1359.

Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology: Combining qualitative and quantitative approaches*. Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of mixed methods in social & behavioral research*. Thousand Oakes, CA: Sage.

Teddlie, C., & Tashakkori, A. (2006). A general typology of research designs featuring mixed methods. *Research in the Schools, 13*(1), 12-28.

Weiss, C. H. (1998). *Evaluation* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Yin, R. K. (2006). Mixed methods research: Are the methods genuinely integrated or merely parallel? *Research in the Schools, 13*(1), 41-47.