

Performance Measurement Challenges in Nonprofit Human Service Organizations

Nonprofit and Voluntary Sector Quarterly
XX(X) 1–19
© The Author(s) 2013
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0899764013508009
nvsq.sagepub.com


Sarah Carnochan¹, Mark Samples¹, Michael Myers², and Michael J. Austin¹

Abstract

This qualitative study examines the experiences of four nonprofit human service organizations engaging in performance measurement processes to satisfy accountability requirements and increase organizational and program effectiveness. Nonprofits are increasingly required to respond to performance measurement mandates issuing from multiple sources. However, many of the recommended strategies have been developed in the for-profit and public sectors, and are less appropriate or feasible for nonprofit organizations. Three central findings emerged from interviews, focus groups, and review of archival data. First, the complexity of human change processes and the variation among individual clients complicate efforts to define client outcomes. Second, staff skills play a critical role in effective utilization of data systems. Third, organizational strategies to support performance measurement include incorporating user perspectives into system design and providing adequate staff access to data.

Keywords

performance measurement, performance management, human service organization

Introduction

Nonprofit human service organizations are increasingly called upon to engage in performance measurement processes aimed at ensuring that the services they deliver are efficient and effective, often as a condition of receiving funding from government and private foundation sources. As Hatry (2002) notes, “The impetus for performance measurement

¹University of California, Berkeley, USA

²Techspierence, Oakland, CA, USA

Corresponding Author:

Sarah Carnochan, Mack Center on Nonprofit and Public Sector Management in the Human Services, School of Social Welfare, University of California, 120 Haviland Hall, Berkeley, CA 94720-7400, USA.
Email: scarnochan@berkeley.edu

has typically come from external funders seeking accountability, not from public managers themselves seeking the information to help them improve their programs” (p. 352). However, performance measurement also offers a strategy for retaining competent staff, addressing outcomes relevant to community and other stakeholders, and informing decisions leading to long-term sustainability (Forbes, 1998). When focused on effectiveness and the outcomes produced by human service organizations, performance measurement may improve outcomes by strengthening evidence-informed practice through the analysis of agency-generated data (Epstein, 2010).

For performance measurement to improve outcomes for clients of nonprofit human service organizations, it is important that organizations draw on the best forms of evidence available to design, implement, and evaluate services. Relevant sources of evidence include both external research as well as data generated within the organization. To use internal data, the organization must engage in effective data mining practices (Epstein, 2010). Data mining depends on both a high-quality data system and staff who are able to define research questions, collect, and analyze data, and utilize these analyses in decision making about clients and programs (Epstein, 2010). However, many nonprofit human service organizations experience challenges related to organizational and staff capacity, and may lack the technology resources needed to collect, store, and analyze the data required for performance measurement (Carman, 2007, 2009; Zimmerman & Stevens, 2006). As a result, as Lynch-Cerullo and Cooney (2011) note in their assessment of the state of performance measurement among nonprofit human service organizations, “the adoption of performance measurement practices at the organizational level appears varied and often superficial” (p. 382).

Despite the increased emphasis on performance measurement on the part of funders and nonprofit human service organizations, empirical research remains limited (Lynch-Cerullo & Cooney, 2011). Moreover, many of the tools and models providing guidance for performance management have been developed in the for-profit and public sectors and nonprofit organizations face multiple challenges when trying to adapt these models (Ospina, Diaz, & O’Sullivan, 2002). This article presents findings from the Performance Measurement Project, a collaborative initiative involving a university-based research center, a regional network of nonprofit human service agencies, and a technology consulting firm. This exploratory study contributes to the emerging knowledge base by examining the performance measurement systems of four nonprofit human service organizations engaged in efforts to design, utilize, and strengthen performance measurement systems and practices. The study was designed to focus on three areas: (a) staff perspectives on the definition of client outcomes, (b) technology resources and limitations, and (c) organizational structures and processes supporting performance measurement. The performance measurement literature broadly identifies the importance of each of these areas.

Literature Review

The terms performance measurement and performance management complement each other, but are sometimes confused or used interchangeably in practice and in the literature (Hatry, 2002; McHargue, 2003). Performance measurement refers to “the regular

collection and reporting of information about the efficiency, quality, and effectiveness of human service programs” (Martin & Kettner, 1996, p. 3). Performance management practices are generally internal organizational processes incorporating logic models, program evaluation, and strategic planning to create “results-oriented systems” (Hatry, 2006; Melkers & Willoughby, 2005; Speckbacher, 2003). Distinguishing between performance measurement and performance management, Speckbacher (2003) describes performance measurement as “a specific definition of the [organization’s] primary objectives and how to measure achievement of these objectives” and performance management as “a specification of the processes that generate performance and, hence, a specification of how management decisions can control performance” (p. 268). This article focuses primarily on performance measurement experiences and challenges, with more limited discussion of performance management practices. Thus, for reasons of economy, the article uses the term *performance measurement* to include both performance measurement and performance management, unless noted otherwise (Lynch-Cerullo & Cooney, 2011).

Sanger (2008) notes that effective performance measurement systems are created using three core strategies: (a) nurturing local stakeholder involvement in the process, (b) creating goals that are specific and logically linked to metrics that measure progress toward those goals, and (c) continually fine-tuning measures and goals that are strategically linked to balancing the needs of federal and state funders with those of clients and local citizens. However, pressure to engage in performance measurement has come primarily from external sources, including contractual obligations to public human service agencies (Hatry, 1997; Lindgren, 2001; McBeath & Meezan, 2006; Poole, Nelson, Carnahan, Chepenik, & Tubiak, 2000) and private foundations demanding financial and program accountability (Benjamin, 2008; Ritchie & Kolodinsky, 2003; Tassie, Murray, Cutt, & Bragg, 1996). Thomson (2010) found that outcome reporting mandates from government funders were associated with increased levels of performance measurement by nonprofit organizations. However, despite the increased emphasis on performance measurement as a means to ensure accountability to funders, research indicates that government monitoring of contracted service providers may not improve performance, and imposes costs that may not be outweighed by benefits (Fernandez, 2009).

While studies have noted that board members may also seek metrics to assess service outcomes (Buckmaster, 1999; Newcomer, 2008), the role of other internal stakeholders in determining specific goals and measures has received less attention. For example, Carman’s (2007, 2009, 2010) important work on the adoption of performance measurement practices among nonprofit human service organizations examines the views of Executive Directors, without addressing the perspectives of managers, supervisors, or line staff. In addition, while significant work has been done in the field to develop outcome measures for multiple human service programs (e.g., Urban Institute Outcome Indicators Project), the utility of these broad measures to inform staff decision making merits further study. This article addresses a limitation of the existing literature by examining staff perspectives on outcome definition.

Human service agencies operate within broad policy frameworks and unique local political economies, subjecting them to limited and fluctuating financial resources

(Hasenfeld, 2010). Nonprofit human service organizations experience resource constraints related to staff, money, and technology needed to develop and manage performance measurement systems (Carman, 2007, 2009; Ebrahim, 2002; Edwards & Hulme, 1995; Uphoff, 1995). While the phenomenon of limited financial and human resources for using technology to support performance measurement among nonprofit human service organizations has been well documented, this study builds upon prior research by exploring the responses of staff within agencies to specific data systems and associated resource constraints and limitations.

Finally, researchers have examined the role that organizational culture plays in supporting performance measurement systems. Lynch-Cerullo and Cooney (2011) summarize the aspects of organizational cultures promoting effective performance measurement: (a) environment emphasizing improving quality rather than “avoiding blame” (Carrilio, Packard, & Clapp, 2003, p. 2), (b) performance measurement conceptualized as an opportunity for learning (Snibbe, 2006), (c) capacity for staff to think evaluatively (Patton, 2004), (d) participatory process to define outcomes (Cairns, Harris, Hutchison, & Tricker, 2005), and (e) regular usage of data (Fisher, 2005). By examining specific organizational structures and processes, this study further elucidates the tangible components of an effective performance measurement culture, where performance measurement involves evidence-informed decision making to achieve successful client outcomes.

Method

Study Background

The Performance Measurement Project is a multiyear study involving seven agencies providing diverse services to children and families, including health and mental health services, case management, recovery programs, emergency shelter and transitional housing, and basic needs (food, clothing). Phase 1 (2008-2009) surveyed nonprofit human service organizations about their experiences developing management information systems to respond to internal and external accountability requirements. In Phase 2 (2009-2010), the project facilitated discussions among senior program, fiscal, and IT managers at the participating agencies to identify strategies for technical capacity building in relationship to performance measurement. Building on this work, agency members of the collaboration requested that the research team examine more closely the experiences and perspectives of agency staff related to performance measurement, focusing specifically on defining client outcomes, developing and utilizing data systems, and organizational structures and processes. This article presents findings from Phase 3 of the project.

Sample

This study was conducted with four medium to large nonprofit organizations providing services to children and families. The organizations are members of a regional consortium that operates as an agency-university research and training partnership. Organizational

size varied, including organizations of medium size operating in a single county along with larger multicounty agencies. Annual agency budgets ranged from US\$5 million to US\$40 million. Services were also diverse: two organizations focused primarily on residential and therapeutic foster care services; one organization provided medical and emergency needs services (food, shelter, clothing); and one organization also offered multiple recreational programs for youth as well as services to immigrant populations.

The organizations also varied with regard to the sophistication of their technology systems and related requirements for staff expertise. The largest organization had contracted with a vendor to develop and support a customized client database system. A database coordinator within the agency oversaw the system, but did not have substantial technological expertise. The second largest organization had designed, built, and maintained its own in-house system, and possessed a team of skilled staff to support the system. The third organization had purchased an off-the-shelf system from a major IT vendor, and relied on an in-house database coordinator. The smallest organization maintained a paper system for tracking client data, and was in the process of moving to an automated client data system.

Data Collection

This qualitative study involved interviews, focus groups, and review of archival records. The use of multiple sources of data allowed for data triangulation and enhanced the internal validity of the findings (Stake, 2006; Yin, 2003). Data collection was carried out by graduate research assistants, and by the director and staff of the technology consulting firm that collaborated on Phases 2 and 3. In-depth interviews of 1 to 2 hours and/or focus groups of 2 to 3 hours were conducted with 8 to 15 staff members at each agency, with representation from multiple levels (senior managers, program managers, supervisors, and line staff) and a selection of programs or divisions. Staff representing different levels was as follows: 16 senior managers (oversight of multiple service programs or a division of the agency), 12 program managers (oversight of a single program or a service site), 4 supervisors (oversight of a unit within a service program), and 14 line staff (e.g., case managers, intake specialists).

The interviews focused on the following domains of the study: (a) definition of client outcomes and logic models, (b) design and utilization of data systems, and (c) organizational structures and processes. Detailed field notes were created for the interviews and focus groups. In addition, the interviews and focus groups were recorded, and recordings were used to develop verbatim transcriptions of material identified as central to the analysis. Finally, agency documents were reviewed, including Mission Statements, Strategic Plans, Annual Reports, Organizational Charts, and Logic Models.

An assessment of technology resources and challenges was conducted through a technology inventory and in-depth interviews with IT and other staff. The inventory domains included (a) applications, (b) network infrastructure, (c) server infrastructure, (d) workstations, (e) Internet connectivity, (f) data integration, and (g) expertise. This component of the study was carried out in collaboration with a consultant who provides IT and data services to nonprofit organizations.

Analysis

Agency documents, interview, and focus group data were analyzed initially to create individual case studies that were then analyzed to identify cross-case comparisons and differences (Stake, 2006; Yin, 2003). The analysis was conducted collaboratively, involving research team members from the university and the technology consulting firm, and leaders from the participating nonprofit human service organizations. The analytical methods included a series of four key steps: (a) iterative coding of transcripts and documents to develop initial concepts and themes; (b) discussions among analyst team members to check and validate concepts and themes; (c) sorting and comparing to develop an “integrated analysis” (Rubin & Rubin, 2005, p. 227); and (d) discussion of findings with individual organizations to validate and further explore themes. The coding strategy utilized an initial set of a priori, descriptive concepts based on the research questions related to defining client outcomes, technology resources and limitations, and organizational structures and processes (Miles & Huberman, 1994). Within these broad categories, an inductive coding strategy was used to identify specific concepts ranging from the micro to the macro level; for example, subcodes related to individual perceptions about client complexity, as well as organizational processes such as system user groups (see, for example, Lofland, 1971).

Findings

The discussion of the findings is organized by three primary domains: (a) challenges related to defining client outcomes, (b) challenges in designing and utilizing data systems for performance measurement, and (c) organizational structures and processes related to performance measurement.

Challenges in Defining Client Outcomes

Participants in the study identified a number of challenges to defining outcomes that would adequately capture the nature of their services and the human experiences of their clients: (a) the dynamic and complex nature of client progress toward goals, (b) tensions between the need for aggregated quantitative data and case-specific data, (c) a lack of systematic processes for defining outcomes, and (d) tensions between funder-mandated measures and staff conceptualizations of client progress.

Complexity of client progress. A central issue raised by participants, particularly line staff, clinical supervisors, and program managers in direct contact with clients, related to the view that certain kinds of data are inherently difficult to track. “People-changing” data were described as hard to capture and difficult to use to assess program and staff performance. For example, in a focus group discussion, an intake coordinator described the challenges her program faced with respect to documenting the progress youth made in forming relationships, and linking this progress to quantifiable outcomes in the areas of suicide prevention, substance abuse, or homelessness:

The fact that these youth are able to create these attachment relationships with the staff and the other youth. If they had never had that relationship we don't know how many people would've committed suicide, how many people would be living on the street dealing drugs . . . There's a big intangible of just having the attachment figure with youth whose parents are no longer around, they don't have any family. You know just being able to connect with people who are solely focused on their success, that's a huge part of that can never be measured. So the stats just seem ridiculous in that respect.

Some program staff described the difficulty of comparing client progress with standard measures or benchmarks because of the varying levels of family or institutional support, unique personal histories and different stages of bio/psycho/social development. One program manager explained that while standardized measures could be difficult to apply, specific client progress could be observed, documented, and interpreted by a case manager or a clinician:

So it becomes very difficult to quantify unless you have a lot of narrative reporting about it. I'm trying to think of what else. I think basically it's just that, and it's also a lot based on youth development model type of stuff so it is relationship building. It's things that you can't really say that "oh we did this, that, and the other thing towards this goal." It's more like you have to . . . [document your work by describing] watching TV together [and] talking about the client's feelings about abortion. [I]t doesn't look like a clinical session . . . but if you really analyze it, it's pretty significant because she was able to articulate how she felt about something.

In some instances, while client progress may be measurable, the measurement system being used is unable to capture the complex progression of improvement. For example, program staff described challenges arising from the time frame of measurement:

So we do pre-FAS and then what's called a mid-FAS and then post-FAS so it is the same tool. It's a scoring system that generates a number saying, look, a family came in at a [level of] -12 and they are leaving with a 0 or 1 or 2. [However], when we do pre-FAS a family may initially say "oh we are doing great, Johnny is in school, mom and dad are getting along, etc." but what we find is that at the mid-FAS we see a decrease in their score . . . [Yet] even though someone says "Hey what's going on? You worked with a family for three weeks and they are worse" . . . in reality, because we have worked with a family, they are finally telling us the truth. So I see strong validity in this tool when we have a family [receiving services] beyond five or six months because we are able to assess them three times, but [for] a family that is only here 2-3 months we only get one pre-FAS in and then we have to do a post-FAS which usually shows a worse score.

In this situation, the validity of the data collected through the assessment tool depends on the accuracy of the information disclosed by the client, which in turn depends on the development of trust between client and staff. Where services are time limited, the tool is unable to capture progress, because the initial baseline data are biased toward the positive, while the second assessment may reflect the true status of the family. When there is insufficient time for a third assessment, progress is difficult to discern or document.

Aggregate versus case-specific data. A tension emerged between a clinician's focus on case-specific data for client-level decision making and a manager's need for aggregate, standardized data for macro-level decisions. In one example, a program manager described a scenario in which a client may be discharged from a program, tracked as a negative outcome at the administrative level, while remaining in contact with program staff and sharing successes related to higher education goals. These positive outcomes may be captured in client case notes, but inadvertently excluded from formal reporting of program outcomes.

This issue was related to a broader phenomenon in which participants tended to define data as "numbers" and "stats." Narrative case record data and case record review data communicated through discussion among colleagues were often described disparagingly, either because these were not deemed to be data, or because of a belief that funders were not interested. For example, one program manager, when asked directly why he or she referred to the things said at group supervision meetings as "anecdotal," explained that it was because he or she had been working on a lot of grant proposals lately, all of which requested descriptions of how proposed outcome objectives would be quantitatively measured. However, narrative data (case plans, progress notes) were seen as potentially valuable by some program managers and line-level staff but difficult to analyze because of their volume and variability. For example, one manager explained,

I think our files are tremendously colorful in the stories that they're telling and there's just no time [to analyze them]. Maybe somebody wants to do a qualitative dissertation [using] content analysis . . . I think we would be wise to offer our material.

Lack of systematic processes. Program managers and line staff did not describe the process of defining outcomes as rigorous or systematic. In some cases, participants described program managers as arbitrarily selecting outcome measures they deemed important. When asked how a particular tool was selected to assess client progress, one program manager reported, "You know, the person who was the director before me, I believe [selected] that particular tool based on using it at another agency." The use of client outcome measures developed with input from multiple stakeholders (e.g., line staff, clients, community members) was rare; it was more common to see programs using outcome measures that were specified and mandated by funders.

Responding to funder mandates. Almost all program managers interviewed in the study referred to funder-mandated outcome measures; often these were the only outcome measures defined for their programs. There was a common tendency among program managers and line staff to believe that funders care more about specific organizational outputs (e.g., number and type of clients served) than client outcomes. Funder requests for documenting client progress were often described as limited or nonexistent, except in the case of some foundation grants. The demand to continue generating accountability reports for funders frequently overshadowed the outcomes that the agency deemed most important. As one participant explained,

I think that the issue isn't going to go away; even if we make a decision internally that we're only going to use this particular outcome measure, we're going to program it into the [data system], everyone is going to get it, we still do need to respond to, "and this contractor wants us to do this other one."

Another study participant described managers' efforts to use the data system to track outcomes, going beyond compliance with funder-mandated reporting:

My impression is that it is, you know, helpful to managers trying to do summary reports, you know monthly or . . . quarterly kinds of things, especially for our funders. So it seems like it's kind of revenue driven . . . I would say that we're trying to make [the data system] work for us a little bit more for . . . outcome measurements and so we have developed some of our own assessments and care plans [that] we are trying to grab and actually incorporate into [the data system] but my sense is that we are all . . . creating our own program specific things.

However, there were some notable exceptions, where staff reported using funder-defined outcomes in useful ways or spoke of internally defined client outcomes that "overlapped" with funder outcomes.

Compliance with funder's performance requirements may conflict with the service provider's views of performance. For example, some participants described funder requirements as overly simplified and lacking the contextual factors needed to assess individual staff performance based on client progress. One case manager described the inadequacy of a client measure, noting,

It doesn't ask how many have maintained a job or education for six months. It says "have they had a job or have they been employed?" So it doesn't measure [duration] and then it also doesn't measure the reasons that they stop [working or going to school]. So it can be family stress, family trauma, you know, why did they stop.

Other participants discussed the distinction between process-related performance (e.g., timely case notes or monthly quotas of service hours) and outcome-related performance (e.g., successfully helping individual clients to achieve permanent change).

Issues in Designing and Utilizing Data Systems for Performance Measurement

The assessment of technology resources and data systems in the participating organizations identified issues and challenges related to (a) the role of staff and processes in an effective client data system, (b) underutilization of the systems, and (c) lack of system integration, including the development of "work-around" systems.

People and processes. A significant finding emerging from the assessment of data systems was that people and processes are as important to effective functioning as the automated client data systems. For example, organizations with at least one staff person

who possessed a detailed, working knowledge of the client data system reported maintaining client data collection and reporting more consistently than organizations lacking this internal staff expertise. A consistent theme emerged whereby information systems that take into consideration real-life work flow processes tend to be resisted less by staff. For example, one agency developed a system that mapped the business processes involved in the flow of client progress notes from initial creation by staff to submission for billing. Finally, although automated data integration processes were observed to increase efficiency, in some cases manual processes were used to achieve data integration, and offered benefits such as providing a method for confirming the accuracy of data transmitted between finance and development departments.

Underutilization of systems. Multiple organizations were underutilizing existing data systems for reasons including limited knowledge of the software applications or insufficient number of software licensing agreements allowing staff to access the database. Regular training and technical support from data system designers and vendors was found to be limited, contributing to the lack of adequate knowledge of the systems within the organization. For example, the database coordinator at one agency reported that the system vendor was unwilling to provide adequate support for the system due to the agency's nonstandard customization of the system. In some cases, individual programs developed separate systems for tracking and analyzing data that were not controlled by senior managers, and were not linked to the organization's integrated system. These "work-around" systems were sometimes created by staff because they did not understand the functionality of the agency's database system, although in other cases, staff were addressing a need for data that the agency system did not provide. Finally, limited staff access to data reports and opportunities to independently generate client/program reports were common in the participating agencies.

System integration. Among the multiple data systems assessed (e.g., human resources, fiscal, and development), the systems related to client/program data represented the greatest challenge for all of the organizations. As outcomes varied considerably across the multiple programs offered by some agencies, due to distinct differences in program goals, populations served, or services, it proved to be challenging to develop an agencywide or integrated client data system. For example, one organization provided a list of 13 separate programs providing services to children, youth, and families (including substance abuse treatment, suicide/crisis services, preventive services for at-risk youth, transitional housing, case management for at-risk families, health and wellness services for girls, and peer youth mentoring). While several common measures were identified across these programs (e.g., General Educational Development [GED] completion, transition to positive living situation, knowledge of personal strengths), there were substantially more unique measures.

There were some exceptions to the challenge posed by the diversity of program outcomes. In one organization, some programs serving youth and adults described their outcomes more universally by viewing families as the unit of analysis instead of individual clients. Two managers in the agency noted that the organization's mission

involves “strengthening community and increasing preventative strengths, based on community resources,” explaining that they were all trying to “reduce risk factors and increase the preventative.”

For several organizations, the continuing use of multiple client data applications is a result of the specialized data that are required by funders. However, several organizations used client data applications that could function as a centralized warehouse for all client information, including core demographic and service data. This solution could be achieved in part through automated, electronic processes, with minimal demand for staff to perform manual tasks, such as transfer of data between systems.

Organizational Structures and Processes to Support Performance Measurement

Organizational structures and processes that were viewed as important to supporting performance measurement included the following: (a) incorporating user perspectives, (b) ensuring appropriate access to data systems, (c) supporting and training people, and (d) rethinking staff roles in relationship to technology.

Incorporating user perspectives. Incorporating the perspectives and experiences of system users into the design of the system emerged as a major theme. The range of methods used to facilitate user input included user groups, focus groups, planning committees, and staff surveys, as well as informal communications. User groups involved regular meetings designed to share user issues with the system developers or managers as well as disseminate information about the system to staff. Focus groups, planning committees, and surveys were used in initial planning stages to identify the priorities of system users. However, these forms of communication were viewed as having more limited value if they were not part of an ongoing process of feedback and negotiation. Individual communications between staff and database coordinators included regular reports as well as specific analyses. Finally, some organizations established online forums within the organization’s data system where individual staff could post suggestions and tips.

Participants described a number of benefits related to user groups and other strategies for obtaining user perspectives. Those who participated in user groups described experiencing a sense of ownership of the agency data system, and in some cases modeled or advocated for use of the system among other staff. Some participants of user groups noted that the experience was helpful in addressing the fear and resistance that can arise when learning a new data system. In addition, obtaining the input of system users was described as improving the system’s design. User input promoted the ease of use for the user interface as well as selection and operationalization of appropriate data elements related to client outcomes.

Participants in the study identified a number of perceived barriers to incorporating user voice into data system design and implementation. Feedback from program staff was sometimes delayed, as they were unable to identify problems in a data system until they had used it for a significant period of time. A second barrier related to underlying

distrust and tension between senior managers and program-level managers, supervisors, and line staff. Some program-level staff members described senior managers as part of a “downtown office” that was disconnected from their work. Specific examples of this disconnect identified by study participants included being informed of changes to data systems “after the fact,” as opposed to being involved in the change process. From the perspective of some senior managers, resistance from program-level staff appeared intentional and deliberate, leading them to use metaphors such as “refusing to get on a train that is leaving the station.”

Determining access to data systems. Barriers to access were a common complaint among system users, many of whom described intentional as well as unintentional barriers. The intentional or formal barriers related primarily to permission or authorization; organizations tended to justify these barriers with reference to confidentiality, limited staff expertise, or relevance of data. The unintentional or informal barriers included inadequate time, inadequate licenses, inadequate skill level, and inadequate communication between system gatekeepers and users. For example, some program-level staff members described contact with their database coordinators to be episodic and focused on reactive problem-solving as opposed to regular, ongoing dialogue. In contrast, staff at other organizations described effective database coordinators who regularly incorporated specific program needs into the agency’s data system.

Supporting and training people. Some participants noted that there is a tendency to view technology, particularly the “new system,” as the solution to problems related to performance measurement for external accountability or internal data-informed decision making. However, the capacity of the people involved in designing, implementing, and using data systems has a substantial impact on the success of specific systems and applications, whether old or new and whether or not deemed to be “user friendly.” The importance of staff capacity was exemplified in organizations with an effective database coordinator, who possessed knowledge of the data system, was responsive to staff inquiries (turnaround and staff satisfaction), and was able to communicate clearly and effectively.

In contrast, insufficient expertise on the part of database coordinators and other key personnel led to inefficient and difficult to manage reporting processes and increased the reliance on “work-around” strategies (e.g., using Excel, word processing applications, or even paper-based strategies to collect, track, and communicate client data). As a result of the growing complexity of the technology involved, staff members at multiple levels of the organizations experience increasing difficulty in understanding and administering or utilizing their data systems. When training is not routine (e.g., all new employees get training) and ongoing (e.g., current employee skills are updated when system changes), the use of data systems is negatively affected. When the knowledge of data systems is housed with specific staff members rather than codified in manuals and taught through formal trainings, it is difficult to distribute expertise across all levels of staff, with negative consequences for system use. More broadly, the need for experts and expertise raises important question about staff roles and structures, as expertise

varies widely within and across agencies. While some agencies rely on in-house expertise, others contract with external consultants.

Human service roles in relationship to technology. A common theme that arose in the interviews and focus groups related to changing roles within the organization as a result of the introduction of new technologies. A number of program-level staff noted that top management plays a role in the speed of acceptance and use of new data systems. Some senior- and program-level managers described the need for executive directors to become champions of performance measurement systems and to find the funds needed to support capacity building. For example, one program manager ascribed the growth of his or her agency data system to the executive director's creative initiatives to find funds to support the development process. In another organization, agency staff members who were championing efforts to increase performance measurement did not receive the support they needed from agency-level administrators and board members, especially when dealing with resistance to change, which requires that the executive director consistently motivate and engage all managers in data system design, implementation, and requests for feedback.

While staff seek guidance from agency leaders (board members and directors) regarding how to understand the need for new or reformed data management systems, these leaders may be hampered by their own lack of expertise in this area. Some program managers described frustration with organizational senior managers for not providing explanations of how existing data systems would be improved by updating them or replacing them with new systems.

The roles of clinicians and other line staff were also affected by the demands of data systems. Clinicians commonly prioritized contact with clients over administrative or analytical tasks, such as data entry or "paperwork." Data entry conducted during client contact also raised concerns about hindering the establishment of rapport with the client. Some program-level managers and line staff explained that laptops or even paper-based assessment tools used in the presence of clients inhibited their ability to create the initial bonds of trust that they viewed as providing the basis for effective interventions. The extent to which participants viewed increased efficiency related to data collection as providing increased time with clients varied across staff and agencies.

Discussion

Summarizing the Findings

The study findings related primarily to three broad themes: (a) challenges related to defining client outcomes, (b) challenges related to designing and utilizing data systems to capture outcomes, and (c) organizational structures and processes related to performance measurement. In addition, findings highlighted a set of factors within the organizational environment that had a substantial influence on performance measurement systems and practices, including funders, educational and training institutions, and IT vendors.

Participants identified a number of challenges to defining the client outcomes that guide the design, implementation, and evaluation of services. One of the most complex challenges involves finding methods for measuring change that are appropriate for a diverse array of clients, and account for the time needed to develop trust between staff and clients. Participants also described the tensions created by the need to respond to funder-defined measures, while seeking to develop outcome measures to inform staff decision making. Where organizational resources are limited, tracking and reporting on funder-mandated outcome measures may divert resources from reporting performance data that are relevant to decision making within the organization.

With regard to designing and utilizing data systems, a key finding related to the critical role played by people and processes in the effective functioning of automated data systems. In this regard, the question of fit between data system and existing clinical and business processes was important. In addition, the study found that data systems in many organizations were underutilized, often related to the lack of staff skills and expertise needed to fully exploit the features of the system. Finally, system integration challenges included the existence of programs within an organization with diverse goals, making it difficult to establish a set of agencywide outcomes.

A number of organizational structures and processes aimed at supporting performance measurement were identified by participants, including the related strategies of incorporating user voices into system design processes and ensuring appropriate data system access. As the use of technology expands in these human service organizations, the roles of leaders and clinicians will need to be reassessed. The ability of educational and training programs to respond to the changing roles of clinicians, managers, and leaders emerged as an area for further study. An additional feature of the organizational environment affecting performance measurement in nonprofit human service organizations related to the role of IT vendors in the design of data systems.

Study Limitations

There are a number of common limitations to qualitative research that should be noted. First, this study was conducted with a small sample of organizations that were not randomly selected, thereby limiting the generalizability of the findings to other organizations. However, the diversity of the sample (e.g., organizational size, service array, and stage of evolution in the development of their data systems) helped to partially address this limitation. Second, interviews and focus groups involved a limited number of agency staff, precluding the inclusion of a complete array of alternative perspectives and experiences within each organization. Finally, the lack of participant observation limited our ability to confirm or question the processes and issues described by study participants. However, the involvement of the technology consultant provided an additional perspective on the organizational phenomena being examined, facilitating investigator triangulation (Yin, 2003) and strengthening internal validity.

Implications for Practice

Based on the common issues that were identified across all the organizations in this exploratory study, several recommendations for nonprofit human service organizations emerge. First, it is important to note that while performance measurement involves “regular measurement of the results (outcomes) and efficiency of services or programs” (Hatry, 2006, p. 3), it is not designed to replace program evaluation that offers a number of strategies to aid organizations in measuring client outcomes and, importantly, determining how these outcomes are being achieved (Newcomer, Hatry, & Wholey, 2010). Performance measurement and program evaluation can be used in ways that are mutually reinforcing; program evaluation can assist in developing measures for continued use, while the data generated by performance measurement systems can be utilized in more rigorous evaluation designs (Poister, 2010). Program evaluations can additionally be used to describe the service delivery processes intended to achieve specific outcomes (formative or implementation evaluation) (Newcomer et al., 2010). For organizations participating in this study, an implementation evaluation would allow them to focus intensively on the experiences of clients throughout participation in a program to develop a detailed picture of client interaction with services and progress across a range of life domains. Such an evaluation could be designed to draw on multiple forms of data, such as client and staff interviews as well as case records, facilitating use of existing, underutilized case notes to inform practice decisions.

Second, organizational development (OD) consultation offers a systematic process that could draw upon the perspectives of stakeholders at all levels of the organization to identify appropriate client outcomes. Through the OD process, organizations can address tensions between funder-mandated measures and data that can be used to support decision making by staff throughout the organization. In addition, such a process would provide an opportunity to assess the organization’s services for consistency with organizational mission, and strengthen the links between mission, services, and client outcomes. Logic models are a useful tool to help staff articulate a program’s theory of change and identify appropriate outcomes to be measured (Savaya & Waysman, 2005). Finally, organizations should support the role of the database coordinator, ensuring the coordinator possesses current and comprehensive knowledge of the system, a working knowledge of the organization’s services, and the ability to communicate with program staff whose technological skills and comfort levels may be limited.

Recommendations related to the role of funders also emerge from the experiences of the organizations participating in the study. First, nonprofit human service organizations would benefit from funder efforts to standardize the performance measurement and reporting requirements that they impose. This would reduce the administrative burden associated with reporting, by streamlining the amount of data organizations need to collect. Second, funders should assess the costs associated with performance measurement and reporting, and adjust overhead rates to adequately fund these processes; if performance measurement is deemed to contribute to program effectiveness,

then it should be funded accordingly (see, for example, Gregory & Howard, 2009). Finally, collaborative initiatives between funders and nonprofit human service organizations to identify appropriate outcome measures can lead to more informed practice by organizations, while continuing to ensure accountability to funders.

Directions for Future Research

The findings from this exploratory study generated several directions for future research on performance measurement in nonprofit human service organizations. First, it would be useful to identify and describe the range of organizational processes and structures that are being used effectively to define client outcomes to provide models for organizations seeking to strengthen their capacity to track outcomes. These models should identify strategies to maximize the benefits associated with including various stakeholders (e.g., funders, program staff, clients, and community members) in the process of defining outcomes and developing performance measurement systems. Related questions involve staff access to data systems and client data, namely (a) What are the benefits and risks associated with increasing access to client data? and (b) How should decisions about access be made to mitigate risk and maximize benefit?

A second area for further research relates to the role of funders in defining client outcomes and developing performance measurement systems. Work to identify, test, and replicate examples of collaborative initiatives involving funders and nonprofit human service organizations aimed at defining measurable client outcomes and providing adequate resources for performance measurement is needed. This work should be informed by research in the field of public administration examining issues of accountability and contract management, as well as studies of performance measurement in the nonprofit literature. While performance measurement holds promise as a strategy to strengthen services and improve client outcomes, multiple challenges persist. Additional research and resources are needed if performance measurement is to fulfill its promise of informing and improving service delivery.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Benjamin, L. M. (2008). Account space: How accountability requirements shape nonprofit practice. *Nonprofit and Voluntary Sector Quarterly*, 37, 201-223.

- Buckmaster, N. (1999). Associations between outcome measurement, accountability and learning for non-profit organizations. *The International Journal of Public Sector Management*, 12, 186-197.
- Cairns, B., Harris, M., Hutchison, R., & Tricker, M. (2005). Improving performance? The adoption and implementation of quality systems in U.K. nonprofits. *Nonprofit Management & Leadership*, 16, 135-151.
- Carman, J. G. (2007). Evaluation practice among community-based organizations: Research into the reality. *American Journal of Evaluation*, 28, 60-75.
- Carman, J. G. (2009). Nonprofits, funders and evaluation: Accountability in action. *American Review of Public Administration*, 39, 374-390.
- Carman, J. G. (2010). The accountability movement: What's wrong with this theory of change? *Nonprofit and Voluntary Sector Quarterly*, 39, 256-274.
- Carrilio, T. E., Packard, T., & Clapp, J. D. (2003). Nothing in—Nothing out: Barriers to the use of performance data in social service programs. *Administration in Social Work*, 27, 61-75.
- Ebrahim, A. (2002). Information struggles: The role of information in the reproduction of NGO-funder relationships. *Nonprofit and Voluntary Sector Quarterly*, 31, 84-114.
- Edwards, M., & Hulme, D. (Eds.). (1995). *Non-governmental organizations: Performance and accountability, beyond the magic bullet*. London, England: Earthscan Publications.
- Epstein, I. (2010). *Clinical data-mining: Integrating practice and research*. New York, NY: Oxford University Press.
- Fernandez, S. (2009). Understanding contracting performance: An empirical analysis. *Administration & Society*, 41, 67-100.
- Fisher, E. A. (2005). Facing the challenges of outcomes measurement: The role of transformational leadership. *Administration in Social Work*, 29, 35-49.
- Forbes, D. P. (1998). Measuring the unmeasurable: Empirical studies of nonprofit organization effectiveness from 1977 to 1997. *Nonprofit and Voluntary Sector Quarterly*, 27, 183-202.
- Gregory, A. G., & Howard, D. (2009, Fall). The nonprofit starvation cycle. *Stanford Social Innovation Review*, 24, 49-63.
- Hasenfeld, Y. (2010). Organizational responses to social policy: The case of welfare reform. *Administration in Social Work*, 34, 148-167.
- Hatry, H. P. (1997). Where the rubber meets the road: Performance measurement for state and local agencies. *New Directions for Evaluation*, 75, 31-44.
- Hatry, H. P. (2002). Performance measurement: Fashions and fallacies. *Public Performance & Management Review*, 25, 352-358.
- Hatry, H. P. (2006). *Performance measurement: Getting results* (2nd. ed.). Washington, DC: The Urban Institute Press.
- Lindgren, L. (2001). The non-profit sector meets the performance-management movement: A programme-theory approach. *Evaluation*, 7, 285-303.
- Lofland, J. (1971). *Analyzing social settings: A guide to qualitative observation and analysis*. Belmont, CA: Wadsworth.
- Lynch-Cerullo, K., & Cooney, K. (2011). Moving from outputs to outcomes: A review of the evolution of performance measurement in the human service nonprofit sector. *Administration in Social Work*, 35, 364-388.
- Martin, L. L., & Kettner, P. M. (1996). *Measuring the performance of human service organizations*. Thousand Oaks, CA: Sage.
- McBeath, B., & Meezan, W. (2006). Nonprofit adaptation to performance-based, managed care contracting in Michigan's foster care system. *Administration in Social Work*, 30, 39-70.

- McHargue, S. K. (2003). Learning for performance in nonprofit organizations. *Advances in Developing Human Resources*, 5, 196-204.
- Melkers, J., & Willoughby, K. (2005). Models of performance measurement use in local governments: Understanding budgeting, communication, and lasting effects. *Public Administration Review*, 65, 180-190.
- Miles, M. B., & Huberman, M. A. (1994). *Qualitative analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.
- Newcomer, K. (2008). Assessing performance in nonprofit service agencies. In P. de Lancer Julnes, F. S. Berry, M. P. Aristigueta, & K. Yang (Eds.), *International handbook of practice-based performance management* (pp. 25-44). Thousand Oaks, CA: Sage.
- Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2010). In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (3rd ed., pp. 5-29). San Francisco, CA: Jossey-Bass.
- Ospina, S., Diaz, W., & O'Sullivan, J. F. (2002). Negotiating accountability: Managerial lessons from identity-based nonprofit organizations. *Nonprofit and Voluntary Sector Quarterly*, 31, 5-31.
- Patton, M. (2004). On evaluation use: Evaluative thinking and process use. *The Evaluation Exchange*, 9, Winter, 4-5.
- Poister, T. (2010). Performance measurement: Monitoring program outcomes. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (3rd ed., pp. 100-124). San Francisco, CA: Jossey-Bass.
- Poole, D. L., Nelson, J., Carnahan, S., Chepenik, N. G., & Tubiak, C. (2000). Evaluating performance measurement systems in nonprofit agencies: The Program Accountability Quality Scale (PAQS). *American Journal of Evaluation*, 21, 15-26.
- Ritchie, W. J., & Kolodinsky, R. W. (2003). Nonprofit organization financial performance measurement: An evaluation of new and existing financial performance measures. *Nonprofit Management & Leadership*, 13, 367-381.
- Rubin, H. J., & Rubin, I. S. (2005). *Qualitative interviewing: The art of hearing data*. Thousand Oaks, CA: Sage.
- Sanger, M. (2008). From measurement to management: Breaking through the barriers to state and local performance. *Public Administration Review*, 68(Suppl. 1), S70-S85.
- Savaya, R., & Waysman, M. (2005). The logic model: A tool for incorporating theory in development and evaluation of programs. *Administration in Social Work*, 29, 85-103.
- Snibbe, A. (2006, Fall). Drowning in data. *Stanford Social Innovation Review*, pp. 38-45.
- Speckbacher, G. (2003). The economics of performance management in nonprofit organizations. *Nonprofit Management & Leadership*, 13, 267-281.
- Stake, R. E. (2006). *Multiple case study analysis*. New York, NY: Guilford.
- Tassie, B., Murray, V., Cutt, J., & Bragg, D. (1996). Rationality and politics: What really goes on when funders evaluate the performance of fundees? *Nonprofit and Voluntary Sector Quarterly*, 25, 347-363.
- Thomson, D. (2010). Exploring the role of funders' performance reporting mandates in nonprofit performance measurement. *Nonprofit and Voluntary Sector Quarterly*, 39, 611-629.
- Uphoff, N. (1995). Why NGOs are not a third sector: A sectoral analysis with some thoughts on accountability, sustainability and evaluation. In M. Edwards & D. Hulme (Eds.), *Non-governmental organizations: Performance and accountability, beyond the magic bullet* (pp. 17-30). London, England: Earthscan Publications.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed.). Thousand Oaks, CA: Sage.

Zimmerman, J. M., & Stevens, B. (2006). The use of performance measurement in South Carolina nonprofits. *Nonprofit Management & Leadership, 16*, 315-327.

Author Biographies

Sarah Carnochan is the research director at the Mack Center on Nonprofit and Public Sector Management in the Human Services (Mack Center). She previously worked in nonprofit legal and human service organizations. Her research interests include performance measurement in human service organizations, interorganizational relations, and qualitative data mining.

Mark Samples has an MSW from the University of California, Berkeley, with publications in human service management and community building. On leave from his doctoral studies at Cal, his current efforts focus on raising his young son and consulting work in nonprofit management practice and planning.

Michael Myers is the director of Techsperience, a technology consulting company providing services to nonprofit organizations for more than 16 years.

Michael Austin is the Milton and Florence Krenz Mack Distinguished Professor of Nonprofit Management, and the director of the Mack Center. His research interests relate to nonprofit management and planning, organizational change, and policy implementation.