Perceptions of Funder-Mandated Performance Metrics on Nonprofit Service Provision

Melissa Iverson

Follow this and additional works at: https://ecommons.luc.edu/luc_diss

Recommended Citation

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 License.
LOYOLA UNIVERSITY CHICAGO

PERCEPTIONS OF FUNDER-MANDATED PERFORMANCE METRICS ON NONPROFIT SERVICE PROVISION

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL IN CANDIDACY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

PROGRAM IN THE SCHOOL OF SOCIAL WORK

BY
MELISSA IVERSON
CHICAGO, IL
DECEMBER 2021
ACKNOWLEDGMENTS

The completion of my doctoral program and this dissertation would not have been possible without the community for which I am so thankful. To my husband, Nick, who has seen me through this process with love, patience, and so much care. Your unwavering support encouraged me to write this next chapter in our story. Our little family made this whole experience possible. To my parents, Kathie and John, thank you for always encouraging, supporting, believing in me, and being in my corner. You both led by example and instilled a strength in me that became the backbone in this journey. To my Grandma Iverson, your love of teaching has encouraged me through all my years of schooling. Thank you to my aunt and uncle, Ralph and Thayae, where I always feel at home and who made it possible to stay connected through the pandemic. Finally, thank you to my extended family for the fun and laughter in moments of stress, the understanding and support when I missed family events, and all the small kindesses along the way.

I would like to thank my dissertation committee and faculty mentors who have generously offered their expertise, time, advice, encouragement, and mentoring; you made this dissertation process such a positive experience. Your collective guidance and support helped me navigate the entire process and were critical to my success. During the pandemic and through the losses, grief, and transition to remote learning, you continually stayed connected and made it possible for me to navigate unprecedented challenges. Dr. Michael Dentato, I could not have
asked for a better advisor and chair. Your teaching, encouragement, and empowerment were consistently accompanied by grace and kindness. There was always laughter, and this made all the difference during the stressful moments along the way. Dr. Nathan Perkins, thank you for teaching me to trust the process and, more importantly, myself. Our meetings to discuss my methods, ideas, and process both shaped my approach to research and reassured me that I could, in fact, “do” research. Dr. Maria Wathen, thank you for all your guidance and expertise. Your insights, generosity in sharing your own experiences, and efforts to create (so many) opportunities to expand my learning have been invaluable. Thank you all for prioritizing this work as well as seeing the value in both my research and my ability to grow as a scholar. I hope to pay it forward someday.

Thank you to the incredible faculty and mentors at Loyola University Chicago and Northeastern Illinois University. These communities of educators consistently create supportive spaces and genuinely invest in students’ success; I have benefited immeasurably from your collective willingness to mentor, network, advise, and guide, and I have learned how to be a great educator through your examples. Thank you to the NEIU Field Team for your support, encouragement, and shared celebration of the joys that our students bring. To Dr. Jackie Anderson, you opened doors to places I knew I wanted to go but for which I did not have the framework. To Dr. John Orwat, thank you for the opportunity to work on an interprofessional research team and for including me in all the parts of the research grant application process. To Dean Stephanie Chapman, thank you for always having an open door, sharing your wisdom, providing the best advice, and for maintaining the best candy bowl in the department.
To my students, you have been the constant inspiration and reminder for why I chose this path. It is a privilege to work alongside each of you, be part of your journeys, and create spaces where we can co-learn, grow, and work. There is such hope for a better world with all of the remarkable change makers and social justice warriors entering the profession. I celebrate you all and am so proud.

Last, to my incredible community of friends—your friendships throughout this journey have meant everything, and I will always cherish them. Megan, my conference-connection-turned-accountability-buddy, you have offered unwavering support and become one of my dearest friends. Mary, your thoughtfulness, humor, and check-ins made all the difference - thank you for always asking what you could do to help. Maggie, our forever scrappy grit constantly has us laughing and learning together. To my Scotland crew, Meghan, Katie, Sara, and Cheryl- no matter the time that has passed, we pick up right where we left off. Finally, to my classmates, Jo, Addie, Siobhan, Sister Hien, Bridget, Whitney, Juli, Diane, Alanna, and Patrick, your friendships have meant the world through this shared journey. I could not imagine going through this process without our support for one another (the study groups! the tea breaks!) or the understanding and grace provided to us all. I look forward to our continued friendships and professional collaborations. All of you were the difference in making these five years an experience of camaraderie and laughter.

My sincerest thanks to you all.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>xiii</td>
</tr>
</tbody>
</table>

## CHAPTER ONE: INTRODUCTION

- Background of Performance Measurement  
  - Significance  
    - Significance to Social Work  
    - Significance to Nonprofit Organizations  
    - Significance to Funders  
  - Definition of Terms  
  - Dissertation Overview  
  - Summary

## CHAPTER TWO: LITERATURE REVIEW

- Historical Overview  
  - Public Confidence and Ambivalence  
  - Accountability Response and Implementation  
  - Government Response  
- Evidence-Based Practice  
- Funder-Mandated Metrics  
- Theories of Nonprofit Performance  
  - Resource Dependency Theory  
  - Institutional Theory  
- Purpose of Performance Metrics  
  - To Evaluate  
  - To Learn, Improve, and Impact Change  
  - To Inform Decision Making  
  - To Budget  
  - To Motivate and Control  
  - To Promote and Celebrate  
- Current Performance Measurement and Evaluation  
  - Theories of Change and Logic Models  
  - Theory of Change  
  - Logic Models  
  - Balanced Scorecards  
  - Benchmarking
<table>
<thead>
<tr>
<th>Social Return on Investment (SROI)</th>
<th>47</th>
</tr>
</thead>
<tbody>
<tr>
<td>Financial Ratios</td>
<td>48</td>
</tr>
<tr>
<td>Barriers to the Implementation of Nonprofit Performance Metrics</td>
<td>50</td>
</tr>
<tr>
<td>Lack of Resources</td>
<td>51</td>
</tr>
<tr>
<td>Lack of Evaluation Literacy and Buy-in</td>
<td>54</td>
</tr>
<tr>
<td>Lack of Organizational or Management Support</td>
<td>55</td>
</tr>
<tr>
<td>Difficulty Defining and Identifying Relevant Metrics</td>
<td>56</td>
</tr>
<tr>
<td>Cultural Shifts and the Impact on Nonprofit Organizations and Clients</td>
<td>64</td>
</tr>
<tr>
<td>Client Impact</td>
<td>65</td>
</tr>
<tr>
<td>Planning and Management</td>
<td>67</td>
</tr>
<tr>
<td>Conclusion</td>
<td>69</td>
</tr>
</tbody>
</table>

**CHAPTER THREE: METHODOLOGY**

<table>
<thead>
<tr>
<th>Grounded Theory</th>
<th>72</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constructivist Grounded Theory Approach</td>
<td>73</td>
</tr>
<tr>
<td>Participants</td>
<td>76</td>
</tr>
<tr>
<td>Inclusion Criteria</td>
<td>77</td>
</tr>
<tr>
<td>Recruitment and Procedures</td>
<td>79</td>
</tr>
<tr>
<td>Details of Confidentiality and Privacy Protections</td>
<td>80</td>
</tr>
<tr>
<td>Protection of Human Subjects</td>
<td>81</td>
</tr>
<tr>
<td>Human Subjects Ethical Implications During COVID-19</td>
<td>82</td>
</tr>
<tr>
<td>Data Collection and Instrumentation</td>
<td>84</td>
</tr>
<tr>
<td>Support for Rigor of Design</td>
<td>85</td>
</tr>
<tr>
<td>Credibility</td>
<td>86</td>
</tr>
<tr>
<td>Transferability</td>
<td>88</td>
</tr>
<tr>
<td>Dependability</td>
<td>89</td>
</tr>
<tr>
<td>Confirmability</td>
<td>89</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>90</td>
</tr>
<tr>
<td>Transcription</td>
<td>90</td>
</tr>
<tr>
<td>Initial Coding</td>
<td>91</td>
</tr>
<tr>
<td>Focused Coding</td>
<td>93</td>
</tr>
<tr>
<td>Theoretical Coding</td>
<td>97</td>
</tr>
<tr>
<td>Peer Debriefing</td>
<td>101</td>
</tr>
<tr>
<td>Reflexive Journaling</td>
<td>103</td>
</tr>
<tr>
<td>Audit Trail</td>
<td>104</td>
</tr>
</tbody>
</table>

**CHAPTER FOUR: FINDINGS**

| Participant Demographics         | 106 |
| Agency Descriptions              | 107 |
| Client Category and Themes       | 109 |
| Definitions of Client Success    | 111 |
| Process and Outcome Evaluation Challenges | 112 |
| Industry Standards Usefulness    | 112 |
Additional Influence of Funder-Mandated Metrics 151
Unintended Influence of Funder-Mandated Metrics 156
Emergent Theory and Connections 162
Agency-Program Connections 163
Agency-Client Connections 165
Client-Program Connections 166
Implications 167
Funders 168
Social Work Practice 171
Social Work Education 173
Future Directions for Research 176
Limitations 180
Conclusion 185

REFERENCE LIST 187

APPENDIX A: EMAIL COMMUNICATION TO POTENTIAL PARTICIPANTS 211
APPENDIX B: CONSENT TO PARTICIPATE IN RESEARCH 214
APPENDIX C: INTERVIEW PROTOCOL AND QUESTIONS 217
APPENDIX D: DEMOGRAPHICS QUESTIONS 220
APPENDIX E: EMAIL COMMUNICATION TO PARTICIPANTS FOR MEMBER CHECK PROCESS 223
LIST OF TABLES

Table 1. Overview of Individual Studies Identifying Barriers to Performance Measurement 51
Table 2. Example of Initial Coding in NVivo 92
Table 3. Example of Client Category and Theme 95
Table 4. Example of Focused Coding in NVivo 95
Table 5. Primary Categories, Themes, and Subthemes 106
Table 6. Participant Pseudonym, Job Title, and Years in Position 107
Table 7. Participant Demographic Information 108
Table 8. Agency Demographics Information 109
LIST OF FIGURES

Figure 1. Three Categories from Focused Coding 96
Figure 2. Example of Concept Map in Focused Coding Phase 98
Figure 3. Example of Concept Map in Theoretical Coding Phase 100
Figure 4. Connections Between Category Themes 163
ABSTRACT

Nonprofit organizations provide essential services to diverse client populations that often address complex and multifaceted personal and social issues. As part of the nonprofit funding process, organizations receive financial support from external sources, such as foundations, government contracts, or corporate giving, and often have to engage in performance evaluation, a process frequently driven by funder priorities or evaluation requirements. As agencies continue to navigate these funder expectations and requirements, understanding nonprofit staff perceptions on the influence of funder-mandated metrics upon service provision becomes more relevant. This qualitative study utilized a grounded theory methodology to explore the perceptions of administrators and staff members within nonprofit organizations regarding the influence of funder-mandated performance metrics on service provision. A purposive sample of 16 direct service providers and administrators from eight nonprofit organizations engaged in semi-structured interviews during the period of March 2020-September 2020. Qualitative analysis revealed that funder-mandated metrics influenced nonprofit organizations across three main themes including the client, program, and agency levels. Overall, mandated metrics fulfill the intended purpose of evaluation systems, such as informing decision-making, guiding service modifications, and informing strategic planning or budgeting processes. Findings also suggest that the influence of funder-mandated metrics can create additional challenges, such as a lack of consideration for client input and personal needs, increased staff time, limited resources to cultivate employee evaluation skill sets and limited resources for program evaluation at the
agency level. Notably, findings suggest that several outcomes unintentionally influence definitions of client success, the client-provider relationship, and client experiences. Funder-mandated metrics requirements can impose subjective and restrictive definitions of success on clients. These metrics can also influence the relationship between direct service providers and clients, influencing the scope of services and time allocated to working with clients. Finally, the influence of metrics can be directly experienced by clients, who often felt less autonomy over their personal goals and decision-making. Implications of the findings for diverse client populations, funders, social work practice and education, social justice, and future directions for research are discussed.
CHAPTER ONE
INTRODUCTION

Nonprofit organizations fill an important societal role, working to address multifaceted social issues that are often large in scale and impervious to change (Carnochan et al., 2019; Head & Alford, 2015; Lee, 2020). The purpose of nonprofit organizations is to provide services that address the complex needs of diverse client populations within the financial and resource constraints placed upon them. To provide high-quality, impactful, and effective services, nonprofit organizations also face the high-stakes challenge of presenting evidence of significant social change to avoid losing critical financial support (Kim et al., 2019; Mosley et al., 2019; Prentice, 2016). Nonprofit leaders are required to provide non-financial indicators to convey to funders that they have used the funding effectively to impact clients’ lives in a successful and positive manner.

Over the last 30 years, government, public, and private donors have demanded increased evaluations to address accountability and transparency concerns (Lee, 2020; Lynch-Cerullo & Cooney, 2011; Worth, 2017). This shift has occurred for several reasons, including the political movement ensuring that those receiving public benefits can demonstrate success, which has led to the stringent oversight and evaluation of government programs. Secondly, there was an increase in public demand for nonprofit accountability after a series of public scandals undermined public confidence in nonprofits. This rise in demand for evaluation metrics has resulted in a challenge to create measures that effectively gauge programmatic impact.
Accountability demands made by funders primarily focus on short-term outcomes, constraining agency learning, or the ability to work toward complex issues with long-term goal setting (Bryan, 2019; Carnochan et al., 2019). The difficulty is finding a balance between accountability and transparency to external funders while meeting the needs of clients and employees and providing services that facilitate positive social change in alignment with the agency’s mission and values (Campbell & Lambright, 2016; Frayne, 2014).

Additionally, nonprofit organizations are often beholden to multiple funders who not only require measurement outcomes as a condition of funding but also mandate that organizations utilize the evaluation measures created by each of numerous external funders (Carman & Fredericks, 2010; Lee & Clerkin, 2017a; Thomson, 2010, 2011). These funder-mandated metrics were developed so that nonprofit organizations that provide similar services would have comparable measurements. However, these unique measurements have the potential to create conflicting expectations from each respective nonprofit and may require additional resource allocations to meet multiple demands for evaluation (Despard, 2017; Frayne, 2014; Moore, 2003). Although metrics can help hold nonprofits accountable for their work and potentially aid them in effectively allocating limited and competitive resources, the expectations for performance measurement may create obstacles for nonprofits. Such obstacles may include socially constructing definitions of performance, effectively defining the impact upon client services, and identifying barriers to implementation of evaluation standards discussed in the following section.
Background of Performance Measurement

To a large extent, when funders mandate evaluation metrics, they often predetermine the goals nonprofit organizations must achieve. Since nonprofit performance is socially constructed, how well an agency performs depends upon who is defining the parameters of client outcomes and success (Kim, 2017; Kim et al., 2019). Research suggests that agencies primarily measure and report only on the outcomes required, which results in using external definitions of success that may or may not align with the organization’s goals and mission (Carman & Fredericks, 2010; Mitchell & Berlan, 2018; Moxham, 2009, 2014). A multitude of unmeasured approaches or interventions can be left without evaluation, resulting in a lack of evidence regarding their impact or potential for replication because funder criteria of success or required outcomes necessary to secure funding are not met (Spitzer, 2007).

The inclusion and use of performance measurement standards continue to be the primary method through which organizations can apply for and ultimately secure financial support at the local, state, and federal levels, as well as through other private and public funding sources. However, attempting to measure effectiveness in a sector that addresses complex and imprecise social issues raises questions about the consequences this approach may have upon the many vulnerable populations and communities served (Bach-Mortensen & Montgomery, 2018; Benjamin & Campbell, 2015; Carman & Fredericks, 2010; Carnochan et al., 2014; Mosley et al., 2019). Concerns have been raised about funders determining client definitions of success, as they may lack the training to know that clients must be understood within the context of their families, communities, and other relational settings and systems (Carnochan et al., 2014; Mosley et al., 2019). The condition of utilizing specific measurements may also require agencies to
engage in specific interventions or treatment models with their clients, resulting in apprehensions about privileging certain types of approaches that may not be best suited for organizations serving unique populations or needs within the community (Benjamin & Campbell, 2015; Mosley et al., 2019). Additionally, the requirement of specific and inflexible performance metrics does not ensure a drive toward the improvement of services, nor does it ensure that measuring data guarantees sound decision-making regarding services (Kim et al., 2019). Several studies raise concerns that evaluation systems limit organizational capacity, allocate program resources ineffectively, and create agency-wide mission drift (Carman, 2010; Lynch-Cerullo & Cooney, 2011; Moxham, 2014).

Additionally, research has extensively captured the barriers to the successful implementation of performance measurement systems. Accountability demands made by funders primarily focus on short-term outcomes, constraining agency learning, or the ability to work toward complex issues with long-term goal setting (Carnochan et al., 2019). Nonprofits also work within fiscal constraints that limit the ability to allocate the necessary resources or training to ensure investment in evaluative systems that capture the required data for funders (Carman & Fredericks, 2010; Frayne, 2014). Lastly, as nonprofits continue to face the requirements of performance measurement systems and tailor their evaluative approaches to those who provide funding, agencies face the challenge of defining outcomes that meet the criteria of imposed evaluation systems (LeRoux & Wright, 2010; Lynch-Cerullo & Cooney, 2011; Moxham, 2014; Thomson, 2010). When definitions of successful outcomes are predetermined, agencies are faced with modifying services to capture the necessary data or, in some instances, struggle with what exactly to measure. As nonprofit organizations continue to battle performance measurement
requirements, they face the tension of being externally accountable while internally maintaining and operating an organization that is client-centered and aligned with its mission. They must struggle to provide individualized client plans and goals based on front-line staff knowledge and clinical wisdom while reporting standardized outcomes that align with funder expectations.

Although the literature raises concerns about the successful and useful implementation of performance measurement systems (Carman, 2011; Carnochan et al., 2014; Lynch-Cerullo & Cooney, 2011; Moxham, 2014), there is still minimal research examining the influence performance metrics have upon the quality and effectiveness of services. The literature explicitly excludes a robust examination of the influence that funder-mandated performance metrics have upon the provision of nonprofit services, and this study seeks to address a significant gap within the literature. The research question guiding the study was, what are the perceptions of administrators and staff members within nonprofit organizations regarding the influence of funder-mandated performance metrics on service provision?

Significance

Researchers continue to discuss the concept of organizational effectiveness as a tool for program assessment (Despard, 2017; Kim et al., 2019; Mosley et al., 2019). Studies also underscore that nonprofit organizations face a number of barriers to successfully implement and sustain performance measurement systems (Bach-Mortensen & Montgomery, 2018; Carman, 2009a; Despard, 2016; Lynch-Cerullo & Cooney, 2011; Mosley et al., 2019). As agencies continue to be obligated to funders and other external stakeholder requirements, the relevance of understanding nonprofit staff perceptions of the influence of funder-mandated metrics upon service provision becomes more relevant. Without a better understanding of these perceptions,
there cannot be an assessment to whether performance metrics information provides value for agencies, the clients they serve, and their funders (D. A. Campbell & Lambright, 2017). The following section discusses the significance of this study in relation to the social work profession, as applied to nonprofit management and leadership, and for funders.

**Significance to Social Work**

Across the social work profession, there is an obligation among service providers and nonprofit leaders to address social problems, challenge social injustice, and advocate for marginalized populations (National Association of Social Workers [NASW], 2017). It is the ethical imperative of those who oversee and carry out program evaluation to ensure that the mission is carried out in a way that best positions the agency to affect true social change. Identifying nonprofit leaders’ and administrators’ perceptions of funder-mandated metrics provides a better understanding of how these metrics impact agency decision-making (Carman, 2009a; Charles & Kim, 2016; Perez Jolles et al., 2017). Since these decisions impact the types of services provided and approach to care, the findings provide insight into the benefits of mandated metrics and identify where they may depart from the agency’s needs.

Additionally, as nonprofit organizations continue to grow and compete for declining resources, there is an increased necessity for social work leaders to produce data-driven results (Carnochan et al., 2014, 2019; Moxham, 2014; Perez Jolles et al., 2017). Identifying perceptions related to funder-mandated metrics in this study documents how socially constructed definitions of organizational effectiveness and performance expectations impact the social work approach of client-centered care. It also identifies the types of external support nonprofit leaders need to continue increasing quality and access to social services.
Significance to Nonprofit Organizations

Several implications relate to the perceptions nonprofit employees have about funder-mandated performance metrics and their influence on service provision. The shift toward accountability created unique opportunities for nonprofit organizations to foster organizational learning and to develop and engage in meaningful evaluation processes over time. Findings from this study can inform social work practice by identifying issues nonprofit organizations face when working to implement funder-mandated performance measurement systems. This could strengthen performance measurement systems for organizations and inform how funders support nonprofit organizational efforts to improve programming through performance measurement tracking. Considering how reliant funders and policymakers are on data from performance measurement, improvements to agency metrics can potentially inform legislation or funding to support improved client services and increase bodies of research supporting interventions for specific populations.

This study also identifies ways for organizations engaging in similar work to identify common outcomes, activities, and performance indicators. Findings identify what types of funders support nonprofit organizations need to capture related to the impact upon clients in a more holistic way rather than through mandated metrics. This type of information can contribute to a larger body of research supporting effective interventions for specific populations (Carman, 2013b; Festen & Philbin, 2010). Such efforts also have the potential to increase evidence-informed practices and identify new, alternative, or culturally relevant models of care. Extending research efforts to better understand nonprofit performance measurements could dramatically impact the social work profession’s interest in addressing the lag between social intervention
research and practitioner implementation. Identifying and addressing nonprofit organizations’ strengths and/or struggles could increase social intervention research, practice-based research and support more collaboration between funders, universities, and agencies alike (Fortune et al., 2010). If funders or academic settings genuinely want to increase such efforts, they must first understand the realities nonprofits face and collectively work to rectify them.

Additionally, how nonprofit leaders perceive the evaluation process affects their decision to conduct performance measurement (or not) and emphasizes their importance on management roles and functions (Frayne, 2014; Perez Jolles et al., 2017). Findings from this study have implications related to understanding management decision making and improving overall program effectiveness. These findings could provide information regarding the value of data being collected, analyzed, and could examine to what extent agencies interpret such data to solely appease funders.

Finally, due to resource dependence upon external funders and stakeholders, nonprofits may be conforming to the demands of these external parties, rather than focusing on the best methods to accomplish their mission (Carman, 2009b; Lee & Clerkin, 2017b; Lynch-Cerullo & Cooney, 2011; Mosley et al., 2019). This research examines how externally imposed metrics influence employee buy-in from nonprofit leaders and its impact on agency support or resistance to the evaluation culture (Spitzer, 2007). The finding could also strengthen relationships with funders and inform their practices, as discussed in the next section.

**Significance to Funders**

Findings from this study create helpful information for funders. Understanding nonprofit leaders’ experiences provides a more accurate picture of organizational performance and
supports efforts to identify what tools best facilitate benchmarking when comparing similar nonprofits (Frayne, 2014; Pennerstorfer & Rutherford, 2019). This information could be utilized to educate both nonprofit leaders and funders about what practices and approaches to measurement can best position the sector for improvements in quality and effectiveness. Additionally, funders could use the information from this study to identify ways to help nonprofit organizations build capacity that measures and captures outcomes in a mutually beneficial way. It can also provide information on how to allocate resources to best support this type of infrastructure growth (Bryan, 2019). Lastly, the findings may assist in minimizing or eliminating funder expectations and behaviors that hinder nonprofits from accomplishing their missions (Despard, 2017; Frayne, 2014). The financial reliance on these funders can limit an agency’s ability to advocate for their needs or receive assistance with metrics that capture data most advantageous to the agency and its clients.

Definition of Terms

Numerous key terms should be defined to outline the appropriate context of this study. These are all terms utilized in the nonprofit sector and are relevant to the subject area researched in this study.

The term nonprofit refers to an organization dedicated to furthering a social cause or advocating for a shared point of view. It uses its revenue surplus to further the organization’s mission rather than distribute income to shareholders, leaders, or members. They are also tax-exempt, meaning they do not pay income tax on the money received to support their organization (Congress et al., 2017; Worth, 2017). Alternative terms also used to describe nonprofit organizations are: nonprofit organizations (NPOs), independent sector, third sector, charitable
sector, and tax-exempt sector. Literature may also use the term *human service organization*, but this term includes entities providing services and making profits. In this study, nonprofit organizations (or their corresponding alternative terms) are included in the research. The term human service organizations was not included as the purpose of this study was to conduct research that examines how metrics influence the development, implementation, and sustainability of performance measurement systems for organizations with financial constraints.

Nonprofit organizations often engage in *strategic planning*, a process that identifies strategies that will best support an agency to advance its mission. Although the strategic planning process can take varying degrees of formality and form, the purpose is to guide and inform decisions for the agency throughout the year (Worth, 2019). Strategic planning supports agency efforts to determine operational and programmatic steps necessary to achieve success and is supported by agency resources allocated to achieve agreed upon goals. *Goal setting* is the process of identifying the direction the agency takes to address the strategic issues that have been identified (Worth, 2019). This results in agencies identifying outcomes, activities, and evaluation tools that are implemented to achieve the agreed upon goals.

*Performance management* is the “process of defining, monitoring, and using objective indicators of the performance of the organization and programs to inform management and decision making on a regular basis” (Poister et al., 2015, p.1). This ongoing process collects and analyzes data to track and monitor the effectiveness of nonprofit programs.

Performance measurement systems and processes are also utilized to produce nonprofit accountability to the public and funders and meet state and federal legal requirements. *Performance measurement* is considered the regular and consistent measurement of activities,
Outcomes, and results that generate reliable data on a program’s effectiveness and efficiency (Bureau of Educational and Cultural Affairs, n.d.). Activities are what a program conducts (e.g., providing job training services or individual counseling). Outcomes are the changes that occur to the individual client or family as a result of their participation in the nonprofit’s programs or activities (Worth, 2019). Outcomes and results can be measured immediately after the client completes the program or over a more extended period (a year or multiple years). Measuring outcomes for nonprofits is often unique to their client base, requiring the establishment of specific measurement systems to capture outcome data.

*Program evaluation* is a method used to determine if specific programs are effective in achieving successful results (Worth, 2019). This application of systematic methods addresses questions about program operations and results (Poister, 2015). A program is defined as “a set of resources and activities directed toward one or more common goals, typically under the direction of a singular manager or management team,” and program evaluation is “an assessment that compares actual with intended inputs, activities, and outputs” (Poister, 2015, p. 34). Inputs are the resources dedicated to a program, which may include money, staff, facilities, equipment, and supplies (Worth, 2019). Outputs are the direct products of the activities (Worth, 2019). Output evaluation determines whether program activities have been implemented as intended (Worth, 2019). The metrics mark success by steps taken in a program to complete a process or accomplish an end goal. Process evaluations focus more on the program service process than on a client’s gain or changes. Outcome evaluations measure the program’s effects on the target population by assessing the progress towards the determined outcomes the program is meant to address (Worth, 2019). Outcome evaluations focus on how the client may have changed,
benefitted, or made strides towards meeting their personal goals. When funder-mandated metrics are implemented, it shapes the types of data to be collected. It informs how client success or progress will be defined within the parameters of the metrics that will measure such progress. The results of evaluation can then be used to make decisions about programming, activities, and any changes necessary to client services.

*Stakeholders* are often individual donors or groups (e.g., foundations, government organizations) that have an interest in ensuring nonprofits fulfill their mission and usually contribute financially to the agency via donations and awarding grants (Worth, 2019). Stakeholders can also be individuals, such as community members, officials, or professionals working in geographical proximity to the agency with a vested interest in the organization’s success (Hatry, 2014; Worth, 2017). For the purposes of this dissertation, funders will refer to foundations, corporate grants, and government grants/contracts, while stakeholders will refer to community members, elective official and others with a vested, non-financial support, in the organization. Accountability to these groups implies an obligation or willingness by the nonprofit organization to explain its actions to the funders and stakeholders (T. H. Poister, 2015; Worth, 2019). This requires a level of transparency from the nonprofit to willingly make critical data about the agency and its programs available (T. H. Poister, 2015; Worth, 2019).

Some nonprofit organizations working with funders who contribute financially to the organization may be tasked with utilizing funder-mandated performance measurement tools as a condition of funding. *Funder-mandated metrics* are defined as externally developed measurement tools that provide definitions of programmatic success, which nonprofit organizations must use to evaluate programs as a condition of financial support or metrics the
funder requires the agency to track and develop their own tools to do so (Carman, 2010; Mensing, 2017; Thomson, 2010, 2011).

Organizational effectiveness is defined in this context as measuring program results accomplished, compared to the results that were set by the organization or other funders (Worth, 2017). For example, a program providing services to address post-traumatic stress disorder might measure the decrease in symptoms among clients after receiving counseling through self-report or a pre-post survey or utilize a standardized scale to assess symptom intensity.

On the other hand, organizational efficiency measures how many resources are used to attain program goals (Worth, 2017). For example, a nonprofit organization may have to show program efficiency through a cost-benefit analysis. This might include identifying how much the agency spends on services per client and weigh that expense against the program’s benefits.

In addition, nonprofit organizations may set up measurement systems through logic modeling to assess program inputs, outputs, activities, and outcomes. When agency programs develop measurement systems, they create goals for each program to achieve. All these terms comprise performance measurement systems and processes that enable organizations to collect data and present results to funders and the public at large. They ensure that programs create social impact, organizations are appropriately allocating resources, and ultimately fulfilling their mission.

Dissertation Overview

This dissertation examines the perceptions that administrators and staff members working in nonprofit organizations within one major metropolitan city located in the Midwestern United States hold regarding funder-mandated performance metrics and their influence on service provision. The following section outlines each of the chapters for this dissertation, including the
literature review and methodology centered upon the framework of grounded theory and discussion chapters.

Chapter two examines the existing literature in relation to the proposed research study. As nonprofit organizations must increasingly utilize performance measurement systems to demonstrate program and client impact and progress, numerous approaches have been proposed as best practice methods. For many nonprofits, one of these approaches has been funders requiring the use of metrics (often developed by the funder) in return for nonprofit financial support. Resource dependency theory and institutional theory have been used throughout the literature to examine the power of funders as influencers over nonprofit outcomes. These theories also provide a framework to examine how metrics influence the agency’s ability to measure client success to best align with their mission, which may differ from the funders’ outcomes. Although the intended purpose of nonprofit performance measurement systems is to evaluate, learn and impact change, inform decision making, budget, control/motivate, promote and celebrate (Behn, 2003; Lee & Clerkin, 2017a), concerns are often raised about the unintended consequences of metric systems. The conclusion of the second chapter focuses on the barriers nonprofit organizations face in the implementation of performance measurement systems. Barriers include a lack of necessary resources to develop and sustain measurement systems, difficulty defining and identifying relevant metrics to satisfy funder requirements, and various challenges related to time, resources, staffing, and other areas.

The methodology section in Chapter three centers upon the use of grounded theory to answer the following research question: what are the perceptions of administrators and staff members within nonprofit organizations regarding the influence of funder-mandated
The constructivist grounded theory method aims to generate knowledge and theory with explanatory power associated with the researched topic. The chapter includes the approach to sampling, description of the inclusion criteria for participants’ working at nonprofits, and explanation of recruitment procedures. A detailed description of the data collection and interview process is also provided, including a thorough explanation of the steps taken to increase the study’s rigor. The chapter concludes by providing information regarding the data analysis process that followed procedures developed by Charmaz (2014). This includes three phases of coding, concept mapping and the use of peer debriefing, reflexive journaling, and audit trail development.

Chapter four presents the findings associated with the data analysis of the 16 interviews, in addition to information related to the participants and agencies in which they were employed. The findings identified three primary categories influenced by funder-mandated performance metrics on service provision that include (1) clients; (2) nonprofit programs; and (3) agencies.

The first category, clients, provides findings on the influence of funder-mandated performance metrics, with three themes that describe the influence on (1) definitions of client success; (2) the client-provider relationship; and (3) client experiences and outcomes. The second category, programs, provides results on the influence of funder-mandated performance metrics through three themes and two subthemes. The themes include the program influence on (1) time and staff capacity; (2) program assessment; (3) service modifications. The subthemes explore the program influence on (1) the limitations of evaluation tools and metrics; and (2) how funder relationships influence performance measurement and service provision. The third category, agency, provides results on the influence of funder-mandated performance metrics with
four themes which describe the agency influence on (1) data learning; (2) agency planning; (3) agency financials; and (4) metrics systems development and sustainability.

Chapter five contains a summary of the study, a discussion of the findings, and emergent theoretical connections. The dissertation concludes by exploring the implications of the study on social work practice, education, social justice, and directions for future research and the study limitations.

Summary

This research study examined nonprofit administrators and employee perceptions of funder-mandated performance measurement influence on service provision. As nonprofit organizations continue to incorporate funder-mandated metric systems into their program evaluation, as a condition of financial support from funders, an understanding of how such metrics influence agency functionality and client services is crucial. Increased knowledge of how nonprofit staff perceive funder-mandated metrics on service provision can provide an understanding of organizational and program decision making based on financial constraints. Additionally, it illuminates how these metrics may influence client success definitions, assessment of client progress or lack thereof, and agency decision making. Finally, this research contributes to a body of literature identifying strengths and limitations to the funder metrics approach that can inform nonprofit agencies and external funders alike.

In the next section, a review of the literature on performance measurement in nonprofit organizations is presented. The chapter includes a historical overview of the progression of performance measurement requirements in nonprofits, a review of the intended purpose of performance metrics, and examining the influence funder-mandated metrics have on service
delivery and program accountability. The gaps in the literature and research are identified, underscoring the need for this research study.
CHAPTER TWO

LITERATURE REVIEW

With more than 1.4 million 501(c)3 tax-exempt organizations in the United States, it is essential to examine and understand how such organizations adopt performance measurement practices, develop evaluative systems, and assess how performance metrics influence the quality and effectiveness of service delivery (The Charitable Sector, 2018; McKeever, 2015). Nonprofit organizations are almost exclusively required to engage in performance measurement processes to demonstrate quality, effective, and efficient service delivery as a condition to receive and maintain funding (Carlson et al., 2010; Lynch-Cerullo & Cooney, 2011). This requires organizations to not only establish measurable outputs, but also to develop systems to track, collect, analyze, and report on these measures. However, there is a large gap in the literature on how to best define and measure the performance of nonprofit organizations (Boateng et al., 2016; Carman, 2011). Despite a rise in accountability requirements from funders and pressure to demonstrate that programs are complying with grant and contract requirements, limited research examines how nonprofit organizations develop systems, collect data, and assess if performance metrics truly evaluate the level of success in program delivery and positive impact upon clients (Carman & Fredericks, 2010; Carnochan et al., 2014). To date, little consensus has emerged as to what truly establishes best practices in nonprofit performance systems, and researchers have not determined how to consistently define and measure the concepts of nonprofit performance (Boateng et al., 2016; Glassman & Spahn, 2012; Sowa et al., 2004).
Nonprofit organizations have a wide range of multifaceted missions, serve a diverse clientele, provide varying types of services, and have a broad range of organizational capacity to carry out their missions (Boateng et al., 2016; Hallock, 2002; Speckbacher, 2003). As a result, researchers have argued that nonprofit organizations may face challenges to operationalize services into metrics, and program success may be reduced to a unified or single performance measure to evaluate services and client impact (Boateng et al., 2016; Frumkin & Keating, 2001; Hallock, 2002; Jobome, 2006; Speckbacher, 2003).

Through a review of the existing literature, Chapter two will contextualize the current study within the historical, theoretical, and more recent body of research. This chapter includes a historical overview of the evolution of performance measurement requirements within nonprofit organizations and a review of the implementation of funder-mandated metrics systems. Included is an examination of relevant theories, including Resource Dependency Theory and Institutional Theory. Such theoretical lenses assist in understanding the impact funders have on nonprofit performance and evaluation systems. The purpose of nonprofit performance measurement, its intended impact on service delivery and accountability, as well as limitations are also reviewed. The current types of metric tools and their usefulness in program evaluation will also be discussed. The chapter concludes with a discussion of relevant strengths and barriers to performance measurements, including a review of the literature examining the impact of performance measurement systems and requirements on clients and program management, and how these systems often dictate client service delivery.

Historical Overview

Over the past thirty years, nonprofit organizations have increasingly adopted performance measurement systems in an effort to address pressures to demonstrate greater public
accountability for public and private funding (Fisher, 2005; Lynch-Cerullo & Cooney, 2011; Worth, 2017). The following section will explore various events that have impacted public perceptions of nonprofit organizations, as well as examine how the public, governmental, and academic communities responded and created strong momentum to implement and monitor performance metrics as a method of accountability.

**Public Confidence and Ambivalence**

The initiative to measure the impact of agency-based programming on a larger scale can be traced back to the Johnson administration’s War on Poverty that began in the 1960s (Jones, 2003; Matthews, 2014; Somers & Block, 2005). Johnson revealed his ambitious War on Poverty plan during his first State of the Union address in response to the nearly 20 percent of Americans living in poverty (Jones, 2003; Matthews, 2014). The administration believed the nation’s poverty levels were a national disgrace and warranted a national response to create a more just society (Jones, 2003; Matthews, 2014)). As a result, there was a radical change in government poverty initiatives in the United States, with the passage of the Food Stamp Act of 1964, Economic Opportunity Act of 1964, and the Social Security Act of 1965, along with the creation of Medicare and Medicaid (DiNitto & Johnson, 2016). Rolling out these progressive social policies, in addition to others such as Head Start and the Jobs Corps, came with sharp criticism from liberal and conservative political parties alike (DiNitto & Johnson, 2016). Despite best efforts from the government to focus on such a massive redistribution of wealth, there was a lack of clarity with regard to defining success, and no clear measures were established to capture progress or failure (Jones, 2003; LeRoux & Wright, 2010).

As government programs increased child welfare services and Medicaid expanded to provide mental health services and support for people with developmental disabilities (Jones,
the question of how to define success and measurement outcomes of poverty policies and their impact continued to create challenges over the next thirty years. As growing criticism of the “welfare state” grew, there was a political and ideological move towards reducing federal aid to those in need throughout the 1980s and 1990s (DiNitto & Johnson, 2016; Sheffield & Rector, 2014). During the Clinton administration in the 1990s, major legislation was passed in the Personal Responsibility and Work Opportunity Act (PRWORA), focusing on ending the current welfare benefits (Melnick, 2002; Somers & Block, 2005). The Act focused on personal responsibility and increased efforts to “reassert American’s work ethic” (Melnick, 2002; Somers & Block, 2005). The new legislation increased restrictions on the availability of cash support to families experiencing poverty and imposed more stringent requirements for individuals to provide “proof” of their need and to show their efforts and progress in achieving financial self-sufficiency (DiNitto & Johnson, 2016). These policy changes were monitored and measured through the use of performance metrics developed by the government in an effort to identify programmatic and individual success of the PRWORA and its initiatives.

Additionally, during the 1990s, a public scandal created a profound loss of public confidence in the nonprofit sector and raised concerns about public accountability. In 1992, a widely publicized misuse of national United Way funding by the organization’s president made news headlines (Carman, 2010; Lynch-Cerullo & Cooney, 2011; Worth, 2017). The United Way President of twenty-two years, William Aramony, along with two other individuals, was charged in a fifty-three count federal indictment for defrauding the organization of $1.2 million dollars (Cipalla, 1992; Scala, 1992). A federal investigation brought charges of fiscal mismanagement in addition to charges of using organizational funds for personal luxury travel and financing a
second home for Aramony’s mistress (Barringer, 1992; Shepard, 1992). As a result of the United Way scandal, public outcry focused on the expectation that nonprofits operate with the highest level of ethics to carry out their missions and serve communities in need (Carnochan et al., 2014; LeRoux & Wright, 2010).

Public trust waivered once again in 2001 when another high-profile scandal raised questions regarding nonprofit financial practices. The American Red Cross came under public scrutiny surrounding “Liberty” funds that were donated in the wake of the 9/11 terrorist attacks. In the aftermath of the 9/11 attacks, $967 million dollars were donated to the Liberty Disaster Fund set up by the Red Cross (Jones, 2003; LeRoux & Wright, 2010). Donors were dismayed to learn that the charity did not intend to use all of the donor-directed funds on behalf of the victims, but rather would reallocate more than $200 million into a reserve that would fund programs unrelated to the 9/11 attacks (Henriques & Barstow, 2001).

Until that time, the practice of raising funds for specific disaster response and redistributing them towards a variety of services and programs had been common practice for the American Red Cross (Barbagallo, 2002; Henriques & Barstow, 2001). The organization determines the amount of need required for a specific event, and any additional funds are reallocated as the organization deems appropriate (Barbagallo, 2002). This information was not conveyed to donors, and in the instance of the 9/11 fund-raising effort, many first-time donors believed their donation would go directly towards supporting families impacted by the attack. Public response raised criticism that the Red Cross misled people making donations, ignored the intended allocation of funds, and did not provide evidence of the impact of the donated funds (Barbagallo, 2002; Henriques & Barstow, 2001).
This was also not the first time the organization had been accused of financial misallocation of funds, with a 1998 legal case brought by the State of Minnesota, accusing the Red Cross of withholding more than $4 million of financial relief for victims of the Red River Valley flooding (Barbagallo, 2002). These high-profile scandals eroded public trust and confidence in nonprofit organizations, prompting agencies to increase effort for public transparency and accountability (Fitzgibbon, 1997; Spitzer, 2007). This heightened the profile of nonprofit organizations, with an emphasis on ethical and responsible fund management in fulfilling a public mission, in exchange for the privilege of tax-exempt status (LeRoux & Wright, 2010). In turn, funders and the public increased pressure on nonprofits to engage in a more formalized method of program assessment and evaluation (LeRoux & Wright, 2010; Worth, 2017). Additionally, a number of self-appointed watchdog organizations, such as GuideStar and Charity Navigator, emerged with the intended purpose of providing information about fundraising efficiency and to monitor the fiscal health of nonprofit organizations through 990 data and public tax records and to increase fiscal transparency (Charles & Kim, 2016; Cordery & Sinclair, 2013; Liket & Maas, 2015). Since these organizations primarily rely on financial markers as indicators of effectiveness, this also increased public concern about nonprofits targeting their funds towards direct services by limiting or decreasing administrative, fundraising, and general expenses (Jones, 2003). As a result, the financial watchdogs increased expectations to hold nonprofits financially accountable and contributed to the increased demand to transparently conceptualize the effectiveness of services. This included demonstrating program impact through outcomes measured to capture effectiveness (Liket & Maas, 2015).

Due to increased public demand for nonprofit responsibility, transparency, and accountability, organizations such as the United Way and the American Red Cross led the charge
to increase efforts to assess nonprofit program performance. They focused on measurable outcomes in an effort to connect financial health with organizational effectiveness (Carman, 2009b; Cooney & Lynch-Cerullo, 2014).

**Accountability Response and Implementation**

Prior to the 1990s, nonprofit charity and funding entities placed minimal emphasis on measuring performance, and organizations were typically held accountable by the communities that financially supported their services and programs (Carman, 2011; LeRoux & Wright, 2010). However, creating metrics to illustrate programmatic impact and performance created a pathway to restoring public trust and confidence in nonprofits (Carman, 2009b; Knapp et al., 2010). As the nonprofit sector continued to professionalize and expand access to various types of donors, funding entities, and stakeholders alike, an emphasis on nonprofit accountability and performance measurement became increasingly important (Moxham, 2014). As accountability and performance measurement requirements grew, the nonprofit sector shifted best practices and standards in an effort to align with those that funding entities required (Lynch-Cerullo & Cooney, 2011). Nonprofit organizations soon developed measurable goals and outcomes for their programming, identified systems to track and collect data on these outcomes, and reported findings to funders. Such findings provide evidence of effectiveness, efficiency, and quality service delivery (Carnochan et al., 2014).

**Government Response**

The trend to improve oversight through measurement took hold in the government as well, with the U.S. Congress holding hearings to discuss new legislation regulating the nonprofit sector. Lawmakers and government administrators sought a way to gather information in an understandable format to inform decisions with regard to organizations and programs funded at
the federal level (Kautz et al., 1997). This led to the passage of the Government Performance and Results Act (GPRA) of 1993 (Carman, 2009; GPRA, 1993; Kautz et al., 1997; LeRoux & Wright, 2010).

The GPRA created an intentional way to link nonprofit performance to funding and budgetary requirements. The act required any organization receiving federal funds to develop strategic plans that outlined goals and objectives which could be measured quantitatively and reported annually, and also inform the budget planning process (Kautz et al., 1997; LeRoux & Wright, 2010). For some programs, mandates required that all federal grant recipients collect and report performance data using standardized, federally approved measurement tools (GPRA, 1993; Lynch-Cerullo & Cooney, 2011).

The GPRA also mandated strategic plans, the development of an annual performance plan, and the production of an annual performance report (Carman, 2009a; GPRA, 1993; Kautz et al., 1997; LeRoux & Wright, 2010). The federal Office of Management and Budget required that strategic plans submitted for funding consideration include (1) goals and objectives for the majority of operations of the agency; (2) specific strategies for achieving the goals and objectives; (3) detailed descriptions of the program evaluation tools used in the development of goals and objectives; (4) goals and objectives written in quantifiable and measurable formats; and (5) a clear way for programs to compare actual program results with the initial goals and objectives (Kautz et al., 1997).

This federal legislation resulted in a widespread response, with many states passing similar legislation at the local level, extending mandates to nonprofit organizations receiving funding through both state and federal funding, such as the U.S. Department of Health and Human Services (LeRoux & Wright, 2010; Lynch-Cerullo & Cooney, 2011). Government
contract partners and larger nonprofit organizations such as the United Way followed suit, requiring organizations to follow these new requirements. However, other organizations and funding entities went a step further and started to require organizations to utilize evidence-based tools and principles for public accountability (Worth, 2017).

**Evidence-Based Practice**

During this period of increased public accountability and the shifting of government funding requirements arose the emergence of the evidence-based practice (EBP) movement in social work and related professions (Fisher, 2005; Lynch-Cerullo & Cooney, 2011). Evidence-based practice can be defined as a “research-informed, clinical decision-making process” to support providers making decisions about the types of interventions or treatment approaches to use with clients (Gray et al., 2012, p. 158). It was regarded as a way to “reaffirm a profession’s commitment to the scientific knowledge base in general and, more specifically, an expectation that practice decisions be based on evidence from scientific research” (Rosen, 2003, p. 197). The EBP movement was fueled by a growing realization that practitioners were not using the best available research and evidence for their clients and practice decisions (Dodd & Epstein, 2012; Rosen, 2003). Evidence-based practice initially encouraged the uniform use of rigorous interventions and empirically tested measurement tools for many types of client treatments (Dodd & Epstein, 2012). These interventions utilized standardized measurement tools that had predetermined performance and outcome goals and came with predefined markers to assess treatment effectiveness (Rosen, 2003; Sheldon, 2001; Webb, 2001).

However, a number of concerns were raised about the EBP approach to client services. For example, EBP could limit the role of social worker practice wisdom and rather emphasize systematic interventions and client evaluation processes (Epstein, 2011; Harris, 1998; Rosen,
2003). Critics further argued that the EBP model eliminated room for flexibility in service delivery, and applications of practice wisdom when working with a specific population or in circumstances where the uniform guide could not be applied (Okpych & Yu, 2014; Witkin, 2012). Concerns were also raised regarding the assumption that EBP methods identified the best model of practice or intervention to utilize, while often not accounting for cultural diversity or the complexity of client needs (Harris, 1998; Okpych & Yu, 2014; Rosen, 2003; Witkin, 2012). Additionally, critics pointed out that EBP privileged empirical research, focusing on the systematic reduction of symptoms or creating behavior change and often created a uniform step-by-step guide to client treatment (Okpych & Yu, 2014; Witkin, 2012). It has been argued that such “cookie-cutter” approaches often reduce the client to a symptom or behavior (Witkin, 2012). The approaches do not examine the whole client, disregard their unique experiences, and view the client as the problem instead of examining a client’s needs within the context of the environment (Witkin, 2012). Finally, EBP created an increased concern for cost-effectiveness rather than program effectiveness based on the outcomes measurement model (Harris, 1998; Sheldon, 2001). These issues related to EBP contributed to a growing concern that this type of service provision could limit the scope of work with a client and limit an agency’s ability to implement creative solutions to meet complex needs.

Despite these concerns, the idea of standardized procedures and measures was very appealing to funding entities and stakeholders attempting to gauge the success of nonprofit programming (Epstein, 2011; Rosen, 2003). Evidence-based practice would ultimately provide research and outcomes to identify best practices for client services, to offer a standardized way to compare programming goals to actual outcomes, and to more easily compare nonprofit organizations providing similar services (Harris, 1998; Webb, 2001). These metrics also had the
potential to help prioritize budgetary decisions and funding allocations based on the information collected and reported to funders and donors alike (Harris, 1998; Webb, 2001). For some funding entities, this led to the implementation and requirement for recipients to measure programs based on the funder’s definitions of program success and utilize a standardized tool developed and provided by the funder as well.

**Funder-Mandated Metrics**

As nonprofit organizations moved towards developing and implementing measurable performance metrics, funders sought to increase their use across nonprofit organizations (Thomson, 2010). Requiring nonprofits to measure outcomes as a condition of funding created an impetus for agencies to develop systems, overcome barriers, and utilize metrics (Thomson, 2010). While many funding entities left the development of performance measurement to the discretion of the nonprofit, some funders moved towards a process of developing external definitions of nonprofit programming success, identified the metrics useful to the funder to make financial support decision, and developed the tools that organizations would utilize for measurement (Carman, 2010; Mensing, 2017; Thomson, 2010, 2011). These types of funder-mandated metrics were created with a similar intention as the government’s GPRA tool—to support agency development of outcomes directed services, providing a rationale for the types of service delivery and a quantifiable way to assess services, while providing a way for funders to compare similar types of organizations (Mensing, 2017; Thomson, 2010). It was also another way to tie performance outcomes to financial support, helping funders make decisions about where their financial priorities were needed while assessing where programs were succeeding and/or failing (Carnochan et al., 2014; Mensing, 2017; Thomson, 2010).
All of the aforementioned efforts to increase public accountability and transparency continue to propel the nonprofit community into the sustained performance measurement culture to better gauge their overall effectiveness and quality of services. Independent agencies have also emerged to rate human service organizational effectiveness, such as GuideStar and Charity Navigator, focusing primarily on numeric metrics and financial ratios to demonstrate impact (Liket & Maas, 2015). Despite all of these changes, organizations still report that performance metrics utilized to convey impact to funders often do not accurately capture the actual changes many providers witness among their client populations (Carman, 2010; Lee, 2021; Lee & Clerkin, 2017b). Such service metrics often do not fully capture nor articulate the impact of personalized client plans (Benjamin & Campbell, 2015).

The required changes for nonprofits to receive and maintain funding over the past 25 years has resulted in the implementation and testing of many different evaluative systems and tools. For example, different approaches to management and leadership styles have been examined (Austin et al., 2011; de Waal et al., 2011; Moxham, 2009), as well as the organizational culture (Glisson, 2015) and types of organizational business approaches (Bakker, 2010) that more effectively implement metrics systems. However, there are few, if any, large-scale empirical studies that examine the evaluation practices of nonprofits (Carman & Fredericks, 2010; Thomson, 2010). Theoretical frameworks and a variety of methods have been proposed, and some empirically tested (Carman & Fredericks, 2010; Thomson, 2010). For example, multiple constituency theory has been tested as a means to improve program performance (Campbell & Lambright, 2016), the; additionally, the goal attainment model, systems resource model, resource dependency theory and systems theory have been explored as concepts for improving performance measurement systems (Forbes, 1998; Liket & Maas, 2015;
Medina-Borja & Triantis, 2007; Mitchell, 2013). However, the continued theme throughout these studies and the literature is the difficulty of determining what the purpose of performance evaluation is, what systems demonstrate effectiveness, and if the metric systems contribute to the growth and support of a nonprofit’s mission (Carnochan et al., 2014; Frayne, 2014). Although the intentionality of funder-mandated metrics is to create a way to compare organizations and drive funding decisions, there is minimal literature to examine how these systems impact programming and client services.

The next section discusses theories that demonstrate how nonprofit organizations are dependent upon external resources to operate successfully. The theories also provide a framework for understanding how external influence can influence nonprofit autonomy and pressure agencies to conform to external performance measurement standards and norms.

**Theories of Nonprofit Performance**

Nonprofit organizations are social systems that rely on external resources to function effectively. These external sources of support can influence and challenge the autonomy of a nonprofit (Frayne, 2014). Agencies may feel pressure to conform to external expectations and standards of programmatic success, definitions of effectiveness, impact, and performance (Frayne, 2014; Worth, 2017). Resource dependency theory and institutional theory may help explain the relationship between resources provided by external entities and how organizations dependent on such resources respond to demands placed upon them (Thomson, 2011). Both theories examine the potential power of external funders as drivers of industry standards and norms that can affect nonprofit performance measurement (Thomson, 2010).
Resource Dependency Theory

Pfeffer and Salancik’s (1978) resource dependency theory explains the “behavior of organizations in terms of their dependence on external constituencies” (Worth, 2017, p. 62). The fundamental premise of resource dependency theory is that agencies must acquire resources from external sources in order to survive (Carman, 2011; Hillman et al., 2009; Worth, 2017). The core ideas of resource dependency are (1) organizations depend on resources; (2) the environment, to a considerable extent, contains other organizations; (3) the resources one organization needs are often in the hand of other organizations; (4) resources are a basis of power; (5) power and resource dependence are directly linked; and (6) power is thus relational, situational and potentially mutual (Hillman et al., 2009; Pfeffer & Salancik, 1978, 2003). According to this theory, the choices and behaviors of nonprofit organizations can be understood by studying the interdependence agencies have with external organizations (Worth, 2017). Examining the power external funders or stakeholders have over the agency by virtue of the resources provided can offer insight into the organization’s operational decisions (Carman, 2011; Hillman et al., 2009; Worth, 2017).

Resource dependency theory provides a framework for understanding the power of relationships between funders and nonprofit organizations, and the role power plays in developing agency performance metrics and outputs. It examines the relationship between resources and organizational control. According to resource dependency theory, nonprofit organizations would deliberately engage in program measurement and evaluation because this would support their efforts to secure financial and other types of resources (Carman, 2011). The pressure to think through program outcomes and activities provides the opportunity for agencies to make changes to their programming that results in improved services. Resource dependency
theory can also explain the effort that nonprofits take to be more accountable through performance measurement. Foundations, state, and federal funders all have the ability to impose performance standards on nonprofits, set achievement standards for programs, and package these requirements as conditions for future financial support (Amirkhanyan, 2009; Lynch-Cerullo & Cooney, 2011; Worth, 2017).

A potential limitation of resource dependency is goal displacement, or the “actions taken by a nonprofit to alter its goals and activities to satisfy the contributor of the funds” (Worth, 2017, p. 62). At times, goal displacement may occur as a result of funder requirements. The influence of resource dependency on an organization is that is their primary source of revenue can, over time, create circumstances where the nonprofit modifies its internal structures, processes, and performance measurements to be in alignment with the primary funder (Carman, 2011; Worth, 2017). The efforts made by nonprofits to align with funders can result in agencies that experience mission drift. It can also lead agencies to measure outcomes that do not reflect the actual quality or impact of their services and, instead, concentrate their energy on the funders’ definitions of successful goal achievement.

A study by Amirkhanyan (2009) that surveyed 69 nonprofit organizations found that the odds of incorporating a funder’s input into their programming decisions are twelve times more likely for agencies that were more financially dependent on a funder. However, nonprofits with more diverse revenue and funding sources were able to maintain more control over their services and programming activities. Notably, even while organizations are completely constrained by the requirements of external funders, resource dependency theory underscores the issues that challenge a nonprofit’s autonomy based on their reliance upon external resources and support (Worth, 2017).
Institutional theory examines the rules, norms, and ideologies imposed on and created within an organization that shape the organizational environment’s expected characteristics (Worth, 2017). The premise of institutional theory is that pressures from the environment in which an organization exists will prompt them to adopt certain processes or structures in an effort to gain legitimacy or be perceived as more successful (Carman, 2011; Worth, 2017). These expectations can be associated with policies, laws, funder mandates, industry standards, or professional norms (Carman, 2011). According to institutional theory, nonprofit organizations would be expected to “engage in program evaluation as a result of coercive or normative pressures in the environment” (Carman, 2011, p.355). Such influences surrounding performance measurements often come from external funders, accrediting organizations, national associations, or other affiliations (Carman, 2011; Thomson, 2010). The practice of program measurement and evaluation would be perceived as compliance with industry standards, a process that legitimate agencies would complete (Carman, 2010, 2011; Worth, 2017). Additionally, when an organization does not comply with these expectations, the organization is viewed as deviant and less likely to receive resources (Oliver, 1991).

Institutional theory provides a framework to understand how external pressure from funders may influence a nonprofit organization’s compliance with mandated performance metrics and how agencies address any disconnect between funder priorities and their programs (Thomson, 2010). Institutional environments in which performance measurement, outcome development, and program evaluation is the norm for accountability to external funders may cause nonprofits to comply quickly with their requirements (Thomson, 2010). A rush to comply may cause organizations to disregard the barriers they face or to fail to fully consider
programming changes that may need to take place to comply with these expectations (Thomson, 2010). When examining agencies through the framework of institutional theory, it is important to understand the environment in which an agency exists and how strongly embedded they are within the larger environment. Similarly, it is essential to assess the power that a nonprofit has to impose its own values and norms and the extent to which the agency’s activities and services are the norms for a given population/community. For example, institutional theory could be used to explain how nonprofit organizations conduct evaluation for symbolic purposes because their funders have little influence on the agencies’ programming decisions (Carman, 2011; Worth, 2017).

Although institutional theory can be useful in explaining performance measurement practices, some scholars claim that the theory does not offer a “complete explanation of the performance measurement process in nonprofit organizations” (Campbell & Lambright, 2016, p. 45). Concerns have been raised that the theory does not take into account the impact that multiple funders may have on influencing the performance measurement process (Benjamin, 2008; Benjamin & Campbell, 2015; Campbell & Lambright, 2016; Carman, 2011). Another critique is that this theory does not consider the perspective of the nonprofit agency or its clients, but only explains performance measurement in terms of external funding sources, imposing norms and limiting opportunities for agencies to share and define outcomes and successes (Campbell & Lambright, 2016). Lastly, institutional theory may explain what might prevent nonprofits from gathering information internally to inform decision making if the norms are being set by outside environmental sources (Campbell & Lambright, 2016; Moxham, 2014). These issues also raise the question of whether or not the implementation of performance
measurement systems designed by outside environments increases overall nonprofit program effectiveness, quality, or efficiency.

Resource dependency theory and institutional theory both examine the power of funders as influencers of environmental norms that affect nonprofit performance measurement and identify what factors may also decrease an agency’s power (Thomson, 2010). These theories also provide a framework to better understand how nonprofit organizations use measurement tools and to what extent they do so effectively. However, they are less useful when examining the extent to which multiple funders or additional factors have an impact on environmental norms or values. When nonprofit organizations have multiple funding entities placing conflicting reporting demands on them, it is more difficult to identify the nonprofits’ strategic responses in relation to any one external organization (Lynch-Cerullo & Cooney, 2011; Thomson, 2010). Based on these theories, it can also be difficult to examine the complexity of relationships with multiple funders and identify what pressures are more likely to influence organizational decision making and service provision (Carman, 2011; Lynch-Cerullo & Cooney, 2011; Thomson, 2010).

**Purpose of Performance Metrics**

The idea that performance measurement improves the effectiveness and efficiency of programs is generally agreed upon by nonprofits and private and public funding entities (Frayne, 2014). However, there is little agreement on the definition and measurement of outcomes. Many concepts exist for how to define and measure performance, including conceptual frameworks, and even standardized tools for completing evaluation. However, there is no consistent agreement in the literature regarding any of these topics (Carman, 2011, 2013a; Carman & Fredericks, 2010; Carnochan et al., 2014; Cordery & Sinclair, 2013; Dawson, 2010; LeRoux & Wright, 2010; Lynch-Cerullo & Cooney, 2011; Moxham, 2009). While there is general
disagreement about the approach to performance metric implementation, the literature does provide content outlining the intended purpose and benefits of performance measurement. For example, the purpose of performance measurement can be (1) to evaluate; (2) to learn and impact change; (3) to inform decision making; (4) to budget; (5) to control/motivate; and (6) to promote and celebrate (Behn, 2003; Lee & Clerkin, 2017a). The following sections expand upon these six key areas.

**To Evaluate**

The move into the accountability era has created opportunities for nonprofit organizations to develop and engage in meaningful evaluation. Evaluation provides the opportunity to compare past and present performance, predict future performance, and assess performance related to similar agencies or against professional standards (Behn, 2003; Cairns et al., 2005; Dawson, 2010; Forbes, 1998; Moxham, 2009; Willems et al., 2014). To compare actual performance criteria, nonprofit organizations design outcomes that capture information regarding the effectiveness, efficiency, impact, best practices, or satisfaction associated with programming (Behn, 2003; Dawson, 2010; Willems et al., 2014). Effectiveness focuses on the organization’s ability to accomplish the results they set out to achieve (Behn, 2003; Dawson, 2010; Willems et al., 2014). Willems (2014) describe effectiveness as focusing on the balance of inputs and outputs “achieved through a combination of processes, projects, and programs implemented by the nonprofit organization to reach its predefined goals” (p. 1650).

Efficiency metrics primarily examine whether results were reached in a cost-effective way (Behn, 2003; Greiling & Stötzer, 2015; Lefroy & Tsarenko, 2014; Moxham, 2014). They gauge the use of financial and non-financial resources, such as time and labor used to achieve predetermined goals (Ramadan & Borgonovi, 2015). For many nonprofits, there is an
expectation to report the costs dedicated to the program as related to the number of clients served or the number of services/products provided to achieve programmatic goals (Morley et al., 2001; Ramadan & Borgonovi, 2015).

Impact evaluation focuses on accomplishments, utilizing data to show the difference between actual outcomes compared to outcomes that would have occurred had an intervention or service not taken place (Behn, 2003; Dawson, 2010; Moxham, 2009). Outcomes are often quantified and can be used to report the cost of services and number of clients served, or to provide data on client outcomes (Cordery & Sinclair, 2013). Nonprofits also often have to provide evidence regarding impact in alignment with the agency’s mission statement (Macedo et al., 2016).

Metric systems assess how the operations compare to evidence-based practices or interventions that are known to be the more effective (Bach-Mortensen & Montgomery, 2018; Behn, 2003; Liket & Maas, 2015; Willems et al., 2014). Performance measurement tools can capture data for comparison with best practice models in a specified field, against evidence-based research, or compare to metrics provided by a government contractor or funding entity (Amirkhanyan, 2009). Lastly, satisfaction evaluation focuses on client and consumer feedback with regard to service delivery (Boateng et al., 2016; D. Campbell, 2002; Carnochan et al., 2014). Such evaluative tools can be used to gauge how consumers view all aspects of services provided by nonprofits.

**To Learn, Improve, and Impact Change**

Performance measurement systems have the potential to collect timely data that improve response times to meet clients’ needs and can provide meaningful information that informs nonprofit decision making (Moxham, 2014). As noted by (Behn, 2003), performance metrics can
answer several key questions for nonprofits such as: What is the agency doing well? Where is the agency not doing well? What does an agency need to do in order to improve? Performance metrics systems have the potential to not only measure program success but to also help identify how the organization could work towards achieving program change (Behn, 2003; Dawson, 2010). Obtaining a better understanding of how an agency works, what influences progress, and how to identify what internal connections impact desired outcomes provide critical data to inform programmatic changes or to add or remove specific services (Behn, 2003). Lessons learned from programming can fuel innovative responses to meet specific community needs and support efforts to provide excellent care. Strengthening performance measurement systems creates opportunities for organizations engaged in similar work to collaboratively identify common performance indicators. This can result in a larger body of data collection to support interventions for specific populations (Carman, 2013b; Festen & Philbin, 2010). These efforts can improve client outcomes, strengthen evidence-informed practices or provide data that inform public policy decisions (Carnochan et al., 2014).

Increasing research efforts to understand nonprofit performance measurements could also have a dramatic impact on the social work professions’ interest in addressing the lag between social intervention research and practitioner implementation (Bellamy et al., 2013; Gray et al., 2012). Identifying and addressing nonprofit organizations’ infrastructure and resource needs could increase social intervention research, practice-based research, and support more collaboration between funders and agencies (Fortune et al., 2010). If professional associations or academic settings truly want to increase these efforts, they must first understand the limitations nonprofits face and work to rectify them.
To Inform Decision Making

Information generated by performance measurements can provide a purposeful way for nonprofit leaders to manage programs, as well as allocate funds and resources to improve client services. Data have the potential to provide insight into nonprofit strengths and weaknesses that could guide strategic planning decisions (LeRoux & Wright, 2010). Decisions about program reorganization, goal changes, and capacity building can all be directed with useful information provided by performance measurement. Metric data can also provide information to executive leadership that can support more effective staff management and monitor resources more effectively. The data can also provide information to make a stronger case for funding and help efforts to build strong relationships with the community, funders, and other stakeholders (Carman, 2013a).

To Budget

Performance measures can also be useful in nonprofit budgetary processes. Data can be utilized to justify programming or additional needs in funding applications (Willems et al., 2014). Metrics are also a useful way to track direct and indirect costs of programming, as well as capturing cost per client, or to identify where resources need to be allocated (Behn, 2003; Cook, 2014; Willems et al., 2014). Performance measurement data can provide a way to assist organizations’ efforts to become more resilient, build financial capacity, and secure resources necessary to accomplish its mission (Behn, 2003). The information derived from these metrics also supports decision making by public officials and in policy funding decisions (Behn, 2003; Carman, 2013b). The data can be used to assess if programs are successful in reaching their objectives or failing to create an impact. Using financial support as a motivating factor for nonprofits to secure funds has the potential to create a reward system for well-performing
agencies, but it also raises concerns that underperforming nonprofits may be punished with a reduction in financial support (Behn, 2003; Carman, 2013a). However, such struggling agencies may benefit from financial support to improve or identify causes for poor performance. All of these efforts to utilize metrics to inform budgeting also point to the larger issue of creating visible public accountability and to justifying financial decisions to community, donors, and other stakeholders.

**To Motivate and Control**

Creating goals to achieve outcomes and providing measures to guide the direction of nonprofit programming can create a structure and provide a sense of accomplishment for staff if outcomes are achieved (Behn, 2003; Cook, 2014). Performance measurement goals can help staff focus their thinking, productivity, and work efforts towards accomplishing a common goal. Collecting data also provides a way for managers to gather and distribute progress markers in a timely way, creating an opportunity to change or modify services (Behn, 2003). Additionally, and often less discussed, performance measurement also creates a way to control nonprofits by providing measures that may predetermine success (Behn, 2003; Ramadan & Borgonovi, 2015). Performance measurements dictate the specific actions that need to take place. In cases of funder-mandated metrics, funders can control the types of services or interventions provided, and create expectations regarding client behavior changes (Behn, 2003; Dawson, 2010).

**To Promote and Celebrate**

Finally, performance measurement can be used to promote and celebrate nonprofit organizational successes. Highlighting successful programming is a way for agencies to celebrate accomplishments, recognize milestones, and build staff loyalty (Behn, 2003; Carman, 2011; Cook, 2014). The same evaluation information can also be used to advertise and promote
programming activities and outcomes. This information is a way to attract donors, validate success to funders, build a marketing brand, and create recognition outside of the organization (Behn, 2003).

As evidenced by these six key areas, the intended purpose and benefits of performance measurement offer many potential ways to support programmatic and organizational improvements and change, all in the effort to more effectively serve communities in need. Performance measurement can address the public’s concerns for accountability and transparency, while also providing a justification for the agency’s existence and importance of their mission driven work. Many types of performance measurements have been developed or continue to be implemented within nonprofit organizations. Metrics systems and tools supporting these efforts are discussed in the following section.

**Current Performance Measurement and Evaluation**

Part of the lack of agreement on how to define performance measurements lies with the socially constructed nature of “effectiveness.” It cannot be accurate to assume there is a “single objective reality” (Renz & Herman, 2016, p. 270). Despite any cautions raised about performance measurement purposes and definitions, the requirement for their use is not likely to disappear (Worth, 2019). Nonprofit administrators and staff members have to select or develop tools to capture and measure outcomes that will appease funder requirements and also meet the needs of the organization. Numerous performance measurement techniques and tools have been developed in an effort to assist nonprofits in creating metrics systems that can be implemented within their programs and agencies.

Performance measurement typically occurs in two different ways: through process evaluation and outcome evaluation. As described by Maschi and de Saussure (2017), process
evaluation occurs over time, examining program service delivery, program operations, and implementation. This type of evaluation also focuses on how consistent the program services are with the original program plan and includes the collection of feedback from clients. The purpose of process evaluation is to gather performance measurement data and use it to improve the program’s ability to reach projected outcomes (Maschi & de Saussue, 2017). Outcome evaluation (also referred to as summative evaluation) provides information focusing on the extent to which change occurred with program clients and within the targeted community. Outcomes are often tracked and measured in achievable increments, whether over the course of one year, three years, or five-plus years and represent key programmatic milestones in the client change process, such as a reduction in community violence (Maschi & de Saussue, 2017). These targeted results are often important to the nonprofit and to funders as well. Nonprofit organizations use various approaches and tools to capture and assess performance. The next sections will examine the most commonly utilized performance measurement and evaluation types including: logic models, benchmarking, balanced scorecards, social return on investment, and financial ratios.

**Theories of Change and Logic Models**

The outcomes approach to measurement has been largely accepted across nonprofits in the United States, in part, due to the early efforts of the United Way and its response to increase operational transparency and improve relations with donors (Worth, 2019). The outcomes approach was their primary method and has since been adopted by many foundations and government agencies as a tool for grantees to measure program effectiveness (Worth, 2019). Outcome measurement utilizes logic models, which include inputs, outputs, and outcomes (e.g., short, mid, and long-term) as performance predictors and measurements. Determining that the
outcomes of a program are related to a program’s specific objectives requires a framework that links all of the services in a step-by-step process from inputs to outcomes (Worth, 2019). Guided by theories of change, logic models are used to determine if a program has made an impact on clients and if the program results are connected to a program’s specific objectives (Bozzo, 2000; Ramadan & Borgonovi, 2015).

Theory of Change

A theory of change is a “description of the theoretical basis for a program and the assumed causal mechanisms that are needed to produce desired ends” (Hatry, 2014, p.12). It provides a theoretical explanation that links all of the steps of a program from inputs to outcomes to illustrate how and why the desired change is expected to happen based on the services provided and how the program promotes social change (Boyce, 2017; Carman, 2010; Cordery & Sinclair, 2013; Maschi & de Saussue, 2017; Worth, 2017). Nonprofits often use theory of change as a basis from which they can articulate why their program is expected to work. Applying a theory of change approach to performance measurement tools assumes that programs are built on “explicit and implicit theories about how and why the program will work” (Weiss, 1995, p.72). There are many possible theories of change an organization can utilize depending on the various social issues the agency is addressing (Carman, 2010; Mayne, 2015).

This approach has been embraced by external funders and nonprofit managers as it provides a number of benefits including (1) helping external funders reach consensus on what they are trying to impact and why; (2) generating a hypothesis about what makes the program work; (3) pinpointing problems in the program plan, and; (4) focusing performance measurement resources on key aspects of a program or activity (Carman, 2010). A theory of change approach is what the majority of nonprofit organizations utilize as a foundation for their program.
development, and in turn, their program evaluation (Boyce, 2017; Maschi & de Saussue, 2017; Worth, 2017). Nonprofit organizations conceptualize specific theories of change for their specific interventions and programs.

**Logic Models**

Logic models are a practical way to take a theory of change and break it down into component parts for a program. A logic model is a framework that illustrates what a program’s intended goals are, explains how the program will work under certain environmental conditions to solve identified issues for a target population, and identifies the performance measurement tools that will be used to capture the data (Lynch-Cerullo & Cooney, 2011; Maschi & de Saussue, 2017; Worth, 2019). The benefit of logic models is that they can be applied to nonprofits with limited financial resources, can be built into more complex data collection systems, and are generally preferred by funding entities (Lynch-Cerullo & Cooney, 2011). Logic models also provide a sensible roadmap and visual diagram for nonprofits, which outline plans for program change and impact (Bliss, 2007).

However, logic models also have drawbacks. First, the model should be based on the mission of the organization. If the mission is too broad or vague, it can be difficult to measure impact when there are so many factors that can influence client behaviors or attitudes, or when it is difficult to define organizational vision or mission. For example, if an agency is focused on improving the dignity and self-worth of a client, how are such concepts operationalized or measured? Second, this approach does not necessarily provide a way to incorporate or illustrate the financial components of programming. A nonprofit may be providing effective services, but they may not be financially sustainable nor reflect whether the program is running efficiently (Worth, 2019). Third, logic models are not a practical tool to compare nonprofits among other
similar organizations. Organizations with varying missions develop logic models to fit their specific performance measurement needs best and may find it challenging to compare common indicators of success (Worth, 2019). This approach does not provide a comprehensive tool to obtain a more balanced perspective of the nonprofit organization as a whole, unlike the intentions of the balanced scorecard approach to performance measurement.

**Balanced Scorecards**

Balanced scorecards were developed to capture a more comprehensive approach to assessing nonprofit organizations as a whole. It is regarded as a modern tool that helps nonprofits create and convey the legitimacy of their overall services and mission (Moxham, 2014). The concept was developed by Robert Kaplan and David Norton in 1992 as a way for businesses to develop a holistic perspective on performance by combining four key areas of assessment (Murray, 2010; Worth, 2019). First, the scorecard looks at the financial perspective, such as fundraising growth and funding distribution through the use of financial indicators (Ramadan & Borgonovi, 2015; Worth, 2019). Second, it measures client and funder satisfaction. Third, it measures the internal business perspective that gauges operational efficiency and quality, along with volunteering development and information communication (Ramadan & Borgonovi, 2015; Worth, 2019). Fourth, the scorecard measures innovation and learning perspective. This focuses on the nonprofit’s ability to adjust to changes within the environment and create innovative programming (Ramadan & Borgonovi, 2015; Worth, 2019).

The balanced scorecard approach and similar variations focus on assessing whether different activities, services, management, and programs broadly achieve the intended results (Worth, 2019). However, the approach also has limitations. Speckbacher (2003), Grieling (2010), and Moxham (2014) raise questions about the replicability of this approach, as successful
studies have only been documented by the originating authors, Kaplan and Normaton. There are limited empirical studies from other researchers and only a small number of single-case studies. Bozzo (2000) also found that a survey of twenty nonprofit organizations utilizing scorecards reported seventy percent had difficulty drawing generalizable conclusions about their nonprofit performance and suggested this approach may not be entirely suitable.

**Benchmarking**

Benchmarking is another way for nonprofits to capture and assess their progress, specifically in comparison to other nonprofit organizations that are similar in their mission, size, and location (Worth, 2019). Poister et al. (2015) define benchmarking as the “comparison among organizations, either at the macro (whole organization) or at the micro (program function level), to see how its performance stacks up against counterpart programs elsewhere” (Poister et al., 2015, p. 382). Benchmarking can be used to compare performance measurements in an effort to identify and highlight programmatic strengths and weaknesses (Worth, 2019). Benchmarking can also be used to focus on how a program compares to best practices. Identifying a practice approach or type of service delivery that has proven to be the most effective and efficient method for service provision creates a standard against which a nonprofit program may compare, as well as to set future goals (Poister et al., 2015).

Although benchmarking can be a way to identify how nonprofit programs compare to similar service providers, it does have its limitations. This type of performance measurement requires a substantially large investment of time and effort by nonprofit staff and administrators to implement systems and to track, collect, and analyze data (Worth, 2019). Additionally, it can often be challenging to identify what specific indicators to use to compare with best practices, which also raises concerns about the reliability of comparable data (Poister et al., 2015; Worth,
Finally, benchmarking focuses on a specific program or subsection of services, making it difficult to know if the program or service being provided is related to the overall effectiveness of the nonprofit organization and its mission (Worth, 2019). When focusing on the indicators of nonprofit mission performance, the benchmarking approach has gained wide acceptance.

**Social Return on Investment (SROI)**

Balance scorecards attempted to look at nonprofit performance from more than one perspective, which is similar to the development of the social return on investment (SROI) approach that attempts to look at nonprofit financial performance and the social impact the programming has upon its clients. Developed in 1996, the concept of SROI was a way to monetize the social value created by a nonprofit (Cooney & Lynch-Cerullo, 2014; Worth, 2019). This approach applies a cost-benefit analysis to estimate the value of the benefit created by a program (program outcomes) and place them over the cost of the initiative (program inputs) (Cooney & Lynch-Cerullo, 2014). This approach offers several benefits. It provides a nonprofit with a broad way to demonstrate the social value created by its services through financial language (Cooney & Lynch-Cerullo, 2014). It also has the potential to help nonprofits identify how to allocate their resources best. In addition, there is the potential to increase philanthropic funds to high-performance organizations that need capital investments to increase the scale of their social impact (Cooney & Lynch-Cerullo, 2014; Moxham, 2009).

Despite the potential of SROI, concerns have been raised about monetizing the impact of nonprofit organizations on client outcomes. While this approach can look at the financial value of social benefits at some level, it can be difficult to measure some client outcomes, especially in financial terms (Worth, 2019). For example, SROI can measure the dollar value in a domestic violence therapy program, but it can be hard to monetize the real benefits, such as reduced fear,
improved self-esteem, increased feelings of safety, or reduction of post-traumatic stress symptoms among clients. Additionally, costs that are spent to support current programming may not yield the benefits until far into the future (Worth, 2019).

**Financial Ratios**

Financial measures are a commonly used way to measure a nonprofit’s performance and health (Worth, 2019). Understanding financial measures have been identified as a critical way for funding entities to understand nonprofits and has several advantages (Kim, 2017). First, the data is objective, easily accessible, and often available via Form 990, providing a way to compare the data with other nonprofits (Worth, 2019). For nonprofits, there are several measures that may be applied. A study by Fremont-Smith and Cordes (2004) identified the following applications (1) comparing the ratio of program expenses to contributed income; (2) examining how much it costs a nonprofit to fundraise; (3) examining the percentage of total expenditures that are being applied to programs or activities; (4) comparing the percentage of total expenditures being applied to overhead (fundraising efforts and administrative costs); and (5) evaluating how much cash and assets are in reserves in relation to the nonprofits operating budget.

The underlying assumption with the use of financial ratio measurements is that “financially healthy organizations will perform better because they are better able to cope with unexpected fiscal challenges while continuously supporting program improvement” (Kim, 2017, p. 526). However, there are very few studies that examine whether (and to what extent) financial metrics capture and predict nonprofit program success (Kim, 2017; Worth, 2019). A recent study by Kim (2017) examined the financial data of 5,000 nonprofits over a period of eight years. Findings suggest that the effort nonprofit organizations take to develop and maintain fiscal
strength may be made at the expense of program effectiveness and efficiency. Organizations in this study were more likely to show adverse performance outcomes as they increased their financial donations. The findings also suggested that nonprofits lost autonomy over their programs as donor expectations were imposed upon the programming while diminishing the agency’s ability to maintain its mission driven programming. Critics argue that this method may inadvertently be encouraging nonprofits to value thrift over excellence by trying to maintain low costs while also increasing clientele (Worth, 2019). This approach also focuses on basic financial measures to reflect the overall success of a nonprofit and can be limiting (Kim, 2017; Worth, 2019). There is an underlying assumption that nonprofit organizations that are fiscally healthy are better able to cope with unexpected fiscal challenges, while also continuously supporting program improvement (Kim, 2017). However, it an untested assumption that there is a correlation between fiscal health and program effectiveness. Although, a connection between an organization’s capacity to provide services and their fiscal health may be intuitive, empirical studies are needed to substantiate the relevant of fiscal measures for program outcomes. Financial measures are useful and impactful only to the extent that they can help understand organizational capacity to achieve their mission (Kim, 2017; Worth, 2019).

All of these approaches offer a variety of performance measurements that may work for differing parts of nonprofit programming and operations. While some bring a structured method to develop systems and examine data, studies suggest that measuring performance in nonprofits creates many challenges for agencies. Some of the approaches are viewed as too complicated for nonprofits’ capacity, and others raise concerns regarding the increased allocation of staff time and resources to collect and process data (Lampkin et al., 2006). Finally, there are concerns about the actual use of the data compiled by nonprofits, which raises questions about the
usefulness of the data, and whether such data that will help service providers improve their services, or if there is a best practice model to measure program and/or organizational effectiveness (Carman, 2013b; Lynch-Cerullo & Cooney, 2011; Worth, 2019).

**Barriers to the Implementation of Nonprofit Performance Metrics**

This next section examines research studies identifying the barriers and struggles nonprofits face when developing and implementing performance measurement systems. The continued external pressure from funders, foundations, and government agencies has required organizations to place a significant emphasis on performance measurement systems to assess program effectiveness and has become a standard requirement to receive financial support (Carlson et al., 2010; Carman, 2013b; Lynch-Cerullo & Cooney, 2011). The need for more complex performance measurement systems has required nonprofit organizations to invest resources into developing systems that track, collect, and analyze measurable program outcomes. Directors and managers use these systems to improve program effectiveness and efficiency, improve strategic decision making, and increase public accountability (LeRoux & Wright, 2010; Thomson, 2011). However, research studies examining the use of performance measurement systems within nonprofit organizations suggest that agencies have not achieved the desired results metrics systems were intended for and rather, identify a number of barriers to successful implementation (Lynch-Cerullo & Cooney, 2011; Moxham, 2014). Studies identified that the main barriers to successful implementation were related to four key areas including (1) lack of resources; (2) lack of evaluation literacy and staff buy-in; (3) lack of organizational or management support; and (4) difficulty defining and identifying relevant outcomes or evaluation systems (LeRoux & Wright, 2010; Lynch-Cerullo & Cooney, 2011; Moxham, 2014; Thomson, 2010). Table 1 illustrates how the themes correspond to the related studies.
Table 1. Overview of Individual Studies Identifying Barriers to Performance Measurement

<table>
<thead>
<tr>
<th>Study</th>
<th>Study methodology</th>
<th>Sample size</th>
<th>Lack of resources (time, money, technology)</th>
<th>Lack of evaluation literacy and buy-in</th>
<th>Lack of organizational support and management</th>
<th>Difficulty defining and identifying relevant outcomes or evaluation systems</th>
<th>Metric issues related to funder-mandated requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akitobi et al. (2012)</td>
<td>Interviews</td>
<td>24</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Bach-Mortensen &amp; Montgomery (2018)</td>
<td>Systematic review</td>
<td>24</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Carman &amp; Fredericks (2010)</td>
<td>Mix-methods</td>
<td>189</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Carnochan et al. (2014)</td>
<td>Interviews &amp; focus groups</td>
<td>4 (36 interviews)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Hoefer (2000)</td>
<td>Survey</td>
<td>91</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Kegeles et al. (2005)</td>
<td>Interviews</td>
<td>4</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Minich et al.</td>
<td>Interviews</td>
<td>24</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Michelle &amp; Berlan (2016)</td>
<td>Survey</td>
<td>311</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Moxham (2009)</td>
<td>Case study</td>
<td>18 orgs (24 interviews)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Moxham (2013)</td>
<td>Systematic literature review</td>
<td>55 papers</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Newcomer et al. (2013)</td>
<td>Interviews</td>
<td>12</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Williams-Gray (2016)</td>
<td>Survey</td>
<td>265</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Thomson (2011)</td>
<td>Interviews</td>
<td>18</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

The lack of resources, evaluation literacy, and support all create difficulties for nonprofits to successfully implement performance measurements and may result in a limited capacity to increase the quality or effectiveness of services. However, the proposed research study focuses primarily on the perceptions of impacts resulting from funder-mandated performance metrics. It provides a more in-depth assessment of the issues related to the difficulty of defining and identifying relevant metrics, as well as assess the impact mandated requirements specifically.
have upon a nonprofit organization’s ability to increase the quality and effectiveness of services. The following sections will summarize and discuss studies examining the first three barriers (lack of resources, lack of evaluation literacy and staff buy-in, lack of organizational or management support), followed by a more in-depth examination of the fourth barrier areas (e.g., difficulty defining and identifying relevant outcomes or evaluation systems, highlighting metric issues related to funder-mandated requirements), as evidenced within a number of seminal studies.

**Lack of Resources**

The inadequacy or lack of resources is often identified as one of the largest obstacles to the successful implementation of performance measurement (Lynch-Cerullo & Cooney, 2011). Specifically, the lack of financial resources, information technology, and employee time are primarily identified as the largest barriers (Bach-Mortensen & Montgomery, 2018; Moxham, 2014; Williams-Gray, 2016). Nonprofits often reported that funding problems had a substantial impact on their ability to develop, implement, conduct, and analyze the data from performance measurements (Bach-Mortensen & Montgomery, 2018; Carman & Fredericks, 2010; Hoefer, 2000; Kegeles et al., 2005; Moxham, 2014; Williams-Gray, 2016). This has the potential to curtail an agency’s ability to gather information regarding program impact and to thwart organizational learning to improve client services (Carman, 2010; LeRoux & Wright, 2010; Moxham, 2014). The lack of financial stability can also hinder a nonprofit’s ability to provide data necessary to funders or to apply for new funding sources that require more complex data. The lack of financial resources also has implications with regard to meeting an organization’s technology needs.
An inability to invest in information technology, computers, and evaluation systems hampers an organization’s ability to implement efficient performance measurement and evaluation systems (Carman & Fredericks, 2010). Many organizations struggle to find a system to best meet their needs while also being affordable, or they often lack the financial resources and knowledge with which to develop their own systems (Akintobi et al., 2012; Bach-Mortensen & Montgomery, 2018; Carman & Fredericks, 2010; Williams-Gray, 2016). The inability to financially support an in-house expert or to contract out for necessary technology services leaves nonprofits grappling with how to collect data in a way that meets their needs within limited financial circumstances (Carman & Fredericks, 2010; Kegeles et al., 2005; Moxham, 2009). This lack of technology infrastructure can leave nonprofits with piecemeal systems that each program develops on their own, resulting in the limited ability to convey impact outside of the specific service and towards the overall mission of the organization (Carman & Fredericks, 2010; Williams-Gray, 2016).

Studies also suggest that this lack of technology investment adds additional strain on staff members’ time and ability to track, collect, and report data (Carman, 2010; Williams-Gray, 2016). Without the necessary technology or expertise, nonprofit managers and staff report that they spend an exponential amount of time conceptualizing and designing performance measurement systems (Carman & Fredericks, 2010; Mitchell, 2013; Williams-Gray, 2016). Employees often have to carve out time to research and learn about the types of services available. Once an evaluation system is employed, they must identify ways to effectively gather, track, analyze, report and store the data they are collecting (Carman & Fredericks, 2010; Carnochan et al., 2014).
Lack of Evaluation Literacy and Buy-in

Numerous studies have reported that many nonprofits did not have staff with the proper knowledge to evaluate their programs and lacked resources to train them in evaluation methods effectively (Carnochan et al., 2014; Hoefer, 2000; Kegeles et al., 2005). Additionally, employees do not always understand why it is necessary to collect specific funder identified data or how engaging in performance measurement is related to their job description (Cairns et al., 2005; Carnochan et al., 2014). Untrained or unsupported nonprofit staff who lack evaluation literacy and are charged with such additional job requirements often do not display motivation to participate in the metric systems creation process (Carman & Fredericks, 2010; Cooney & Lynch-Cerullo, 2014). The other issue reported in multiple studies suggests that when systems are put in place and staff are not comfortable or resistant to using them, increased “workaround strategies” and more informal processes such as using Excel, Word, or paper-based strategies to collect information were common (Carnochan et al., 2014; Moxham, 2014; Newcomer et al., 2013).

When nonprofit organizations do not have the capacity or do not invest in staff training in computer skills and performance measurement systems, and fail to communicate expectations to utilize the systems, they often appear very disconnected from a provider’s direct service role. As a result, findings from Carnochan et al. (2014) suggest an increase in staff resistance to engage in the metrics process. Some staff developed feelings of opposition to utilizing the data systems, while others felt resistant to engage in such agency processes due to their lack of knowledge and experience. These employee responses were also often perceived by senior managers as insubordination or intentional refusal to complete this portion of their job, creating issues between management and direct service providers (Carnochan et al., 2014; Kegeles et al., 2005).
However, the issues of evaluative literacy and a reluctance to engage in a data culture can also be held by upper management and administrators, which creates an additional set of barriers for nonprofit organizations.

**Lack of Organizational or Management Support**

An agency’s ability to successfully implement performance measurements can be greatly influenced by the overall organizational culture (Fisher, 2005). Poor leadership, a lack of board support, and/or a low prioritization by administrators can drastically impact organizational culture with regard to program evaluation (Bach-Mortensen & Montgomery, 2018; Carman & Fredericks, 2010). When administrators fail to provide direction and support to staff, research suggests a decline in overall dedication to collecting data within programs (Bach-Mortensen & Montgomery, 2018; Carman & Fredericks, 2010). The primary contributing cause to directionless leadership of performance measurement systems is often related to the lack of resources within some nonprofit organizations. A study by Mitchell and Berlan (2018) suggests that the more resources allocated to the training of professional staff, the more positive an agency’s culture regarding evaluation. However, as noted above, a lack of resources is consistently cited as a primary barrier to conducting performance measurement. When there are limited resources, metrics planning and decision making can often remain at the management level, without input or inclusion from direct service providers.

Additionally, when organizations struggle with effective leadership, there is an increase in employee turnover rates, which can impede the data collection process. It can be difficult for organizations to effectively evaluate programs without consistent staff and difficult to provide incentives related to adding evaluation responsibilities to their jobs - typically already exceeding workload capacity (Kegeles et al., 2005). A lack of training can also limit the sharing of
information across the organization, muddy the clarity with regard to the necessity for performance measurement, and create a negative culture towards evaluation overall (Mitchell & Berlan, 2018; Williams-Gray, 2016). Kegeles, Rebchook and Tebbetts (2005) also found that people in leadership positions feared evaluation might reflect poorly on their programming or result in judgment (by funders, board members, staff, etc.) that they are ineffective leaders. A number of organizations reported that the fear of perceived failure created resistance to documenting programmatic data (Kegeles et al., 2005). Finally, managers and leaders reported that the lack of support for performance measurement was due to difficulty defining and identifying relevant metrics, in addition to frustrations imposed by funder-mandated program goals and metric systems (Bach-Mortensen & Montgomery, 2018; Carman & Fredericks, 2010; Carnochan et al., 2014).

**Difficulty Defining and Identifying Relevant Metrics**

Another primary barrier to developing and implementing successful performance measurement systems was often related to the difficulties that organizations faced in defining useful and meaningful client outcomes, as well as identifying ways to accurately capture and measure these outcomes (Carman & Fredericks, 2010; Carnochan et al., 2014; Mitchell & Berlan, 2018; Moxham, 2009; Williams-Gray, 2016). Such challenges are critical to the current study, as research has examined the issues with defining and identifying relevant metrics and what barriers nonprofits can face. However, there is a gap in the literature examining whether nonprofit perceptions of imposed funder-mandated performance measurement outcomes and metrics systems result in or contribute to the increasing quality or effectiveness of their services.

The first study is a qualitative study that examined the experiences of four nonprofit organizations engaging in performance measurement processes to satisfy accountability
requirements and increase program and organizational effectiveness (Carnochan et al., 2014). The study was conducted over three years and included nonprofit agencies providing diverse services, including health, mental health, case management, recovery services, transitional housing, and basics needs services (e.g., food, clothing). The study was conducted in three phases to allow for data triangulation. The first phase included a survey sent to staff at the organizations asking about their “experiences developing management information systems [in] response to internal and external accountability requirements” (Carnochan et al., 2014, p. 1017). Phases 2 and 3 included focus groups that were conducted for two to three hours or individual interviews for one to two hours with 8 to 15 people at each agency. There were a variety of different personnel positions represented across the agencies, including 16 program managers, 12 program directors, four supervisors, and 14 line staff. The focus groups and interviews focused on the participants’ definitions of client outcomes, how they designed and utilized data systems, and what the organizational processes were regarding performance measurement. The study also conducted a review of archival records and an assessment of technology resources and challenges. These three approaches were analyzed to create individual case studies and then analyzed for cross-case comparison.

The findings identified three primary domains (1) challenges related to defining client outcomes; (2) challenges agencies faced in designing and implementing database systems; and (3) issues regarding organizational structures and processes related to performance measurement. The findings related to challenges in defining client outcomes raise a concern about accurately capturing the complexities of the natural progression of client progress toward meeting goals. Participants in the study were concerned that there were certain kinds of data that are difficult to track because it is hard to capture and difficult to assess success when people are at various
stages of change (Carnochan et al., 2014). Concerns were also raised regarding the difficulty of comparing client progress with predetermined benchmarks or standardized measures due to the varying family and environmental circumstances, making their complex and unique circumstances hard to measure and capture client progress that does not take place in a linear process.

The study also found that there was a trend among program managers to only measure funder-mandated outcomes, believing that funders care more about “specific organizational outputs (e.g., number and type of clients served) than client outcomes” (Carnochan et al., 2014, p. 1021). Funder requests for data regarding client progress was limited or nonexistent and overshadowed the outcomes the organization identified as the most important. These findings suggest that nonprofit organizations face challenges defining outcomes that would adequately capture the nature of their programs and client experiences, lacked the training to support the development of rigorous outcomes, and often solely measured outcomes related to funder requirements. The study also provides a more in-depth look at the realities nonprofit organizations face when required to meet imposed funder mandates. Although the study had a small sample size limiting transferability, the agencies were diverse in size and service provision. Overall, the study provides an in-depth perspective regarding the realities nonprofit organizations face when addressing funder measurement mandates.

A mixed-methods study by Carman and Fredericks (2010) also explored the evaluation capacity of nonprofit organizations and identified barriers to implementation. The study examined the extent to which nonprofit agencies need assistance with evaluation and what types of assistance would be most useful. Utilizing the Indiana State 2004 IRC Form 990 database maintained by the National Center for Charitable Statistics, nonprofit organizations were placed
into three categories (1) agencies providing social services; (2) agencies providing services to people with disabilities; and (3) agencies providing housing and community development services (Carman & Fredericks, 2010). A sample of 340 nonprofit organizations was identified, and a six-page mail survey was sent to each. The survey had 10 independent variables and 19 dependent variables. Of the 340 surveys, 189 were returned, with a 56-percent response rate. Follow-up interviews were also conducted with executive or associate directors at 26 organizations. Interviews lasted between 45 to 60 minutes and were conducted using 10 open-ended questions.

Multiple approaches were utilized for data analysis. First, simple frequencies were run to identify the types of barriers nonprofits experienced, with a total of 19 identified. Second, Chi-square tests were conducted to see if organizations providing different services faced unique challenges. Third, crosstabs were analyzed to determine if some of the challenges identified might be related to one another. Fourth, a multivariate analysis using binary logistic regression was used to identify the extent that certain organizational characteristics explained the variability in their experiences (Carman & Fredericks, 2010). Finally, the interviews were coded using an inductive approach to identify recurring concepts and then sorted into clusters based on similarities and differences between each cluster.

Findings identified 19 barriers that nonprofits noted and broke the agencies into three clusters based on the number of challenges they experienced. Among these, a number of the challenges centered upon difficulty defining outcomes, identifying proper metrics to utilize, and frustrations related to capturing funder-mandated outcomes. Participants reported that their evaluation processes met the needs of the funders, but did not adequately capture client impact or change (Carman & Fredericks, 2010). For some agencies, the outcomes did not align with the
program goals, and the imposed measurement systems were too rigid and prescriptive. Other organizations had difficulty developing measurement tools or identifying appropriate outcomes to meet the funders’ expectations. Nonprofit organizations also reported frustration with the specific reporting requirements imposed by funders, raising concerns that the required metrics did not align with the needs of the agency (Carman & Fredericks, 2010).

First, the approach to identifying clusters is subjective and may not be sensitive to outliers. The study also relied upon self-reported data and had the potential for respondents to give socially desirable answers. Finally, the study focused on the differences of who is being served by the organizations, but there could be additional ways to examine the topic, such as the fiscal size of the organization. The study also did not examine any information about specific funder-mandated evaluation requirements, which could have assisted in explaining the variability in the three clusters.

Finally, Moxham (2009) examined performance measurement practices in six nonprofit organizations and their associated funders. The study posed three research questions including (1) Why measure nonprofit performance; (2) How is the performance measurement criteria developed; and (3) How is the nonprofit evaluated. Implementing a case study approach, six agencies that met the criteria of receipt of external funding and requirement to provide data on performance to funders were found. Additionally, to reflect the diversity of the nonprofit sector and improve validity, two local, two national, and two international nonprofit organizations were identified.

A total of 24 managers were interviewed using a semi-structured format, and data were analyzed in three stages. First, coding was used to examine the transcripts of the interviews and flow charts were developed to present a “story” of performance measurement for each agency.
Next, pattern coding was utilized to identify themes and trends, followed by a cross-case, comparative analysis to compare the different drivers of performance measurements for each agency.

Findings identified four primary drivers for measuring nonprofit performance: to report financials, to demonstrate achievement to funders, to demonstrate operational control to funders, and to facilitate organizational improvements (Moxham, 2009). Metrics were used to prove financial accountability and responsibility to funders as a way to maintain funding revenue. The performance measurements were also used to demonstrate the quality of their programs, but often also as a way to secure funding or seek additional financial support. The data produced from these metrics were used to support organizational growth, but often as outlined by the funder. All of the funder metrics had been developed and required by funders as a result of government policy, or as developed by the funder, with some agencies also developing their own outcomes and metrics. However, the majority of the agencies that had funder stipulations on performance metrics did not know how the criteria they had to report on (or against predetermined standards) were developed (Moxham, 2009). The study was not without limitations in that it had a small sample size, potentially limiting generalizability. It also could have been designed to examine nonprofits all from the same service sector for a more consistent context regarding performance measurement. Additional criteria could have been added (such as charitable status, budget size, or age of organization) to create a more comparable set of organizations for cross-case comparison.

In addition to these three studies, two systematic reviews of the literature have been conducted to examine nonprofit barriers to performance measurement systems. The first systematic review from Bach-Mortensen and Montgomery (2018) sought to identify what
barriers and facilitators nonprofit agencies identify when evaluating the services, they provide. The systematic review followed the PRISMA guidelines, searching ten interdisciplinary databases, and included any study examining barriers and facilitators to the evaluation process compared to the quality checklist developed by researchers. Twenty-four studies were included in the final analysis to identify barriers and facilitators to nonprofit evaluation and were coded to identify themes.

The findings reflect the barriers discussed in this section, with issues of resources, training, and organizational support being reported. In addition, of the 24 studies examined, 19 identified challenges in using evaluation systems or identifying indicators for outcomes, and 13 studies also reported that they faced barriers developing and implementing systems from funder-mandated requirements (Bach-Mortensen & Montgomery, 2018). These findings support the discussion regarding issues nonprofits face defining outcomes and implementing systems successfully.

The study discusses several limitations, including the limited number of databases, which resulted in a number of publications being excluded even when meeting the review criteria. Additionally, the inclusion criteria did not differentiate between grey literature and peer-reviewed articles, raising the question of quality and rigor. Finally, the different organizations included in the study were likely subject to different factors creating barriers or facilitators to performance measurement, which limited the generalizability of the study (Bach-Mortensen & Montgomery, 2018).

The second systematic review aimed to examine the nonprofit literature regarding performance measurement systems designs and identify key themes (Moxham, 2014). Utilizing a search string approach an initial 110 papers were identified and 55 met the three criteria, with
peer-reviewed articles related to human service organizations and a focus on systems for organizational performance measurement.

Results found that the studies fell into three main research categories including (1) Why do nonprofit organizations measure performance; (2) How do nonprofits measure performance; and (3) What are the key considerations for nonprofit performance measurement systems design (Moxham, 2014). The study discussed the topics covered in this section with regard to limited resources, lack of staff training, the need to demonstrate accountability to funders, and identified barriers to developing measurement systems, as well as finding difficulty in identifying appropriate measurement definitions and outcomes. These difficulties included measurement systems that detracted from agency programming, tools that were developed in isolation and did not provide the agency with integrated systems, outcomes that only focused on short-term goals, and the fact that organizations often did not share the same ideas regarding program improvement as the funder. Based on the findings, Moxham (2014) determined there was a lack of clarity around the design of performance measurement systems and suggested that questions should be raised about whether nonprofit organizations should be using the current types of performance measurements at all. There is limited detail provided in this study regarding the systematic review process, specifically with regard to the string approach taken to identify relevant works. Inclusion and exclusion criteria were lacking, as were the details of the analytic processes. The findings did not examine the context of different agencies (e.g., budget size, age of the agency), nor did it examine any studies that tested the theory on this topic (Moxham, 2014).

In summary, much of the existing literature has examined the intended purposes of performance measurement; types of systems designs; and challenges, barriers, and limitations to
successfully implementing systems within nonprofit organizations. Specifically, the overarching issue agencies face is identifying outcome definitions and selecting appropriate metric tools. Additionally, agencies face challenges to integrate or prioritize funder-mandated outcomes and metrics into systems designs in an effort to demonstrate accountability, effectiveness, and secure or maintain funding. Although the literature has raised these issues, performance measurement continues to be the primary practice to assess nonprofit organizations, a model that has become standard practice for both funders and agencies alike.

The move toward the current standards of performance measurements is not without issue, with nonprofits raising a number of concerns about the challenges they face in their efforts to successfully define client outcomes and effectively develop or implement systems that capture client impact and change. These changes have not only modified the process for measuring nonprofit programming but have had consequences in the way nonprofits implement services as well as and have shifted the way clients are viewed as agents of change within these systems. The next section will discuss the cultural shifts and implications performance measurements have on nonprofits and the clients they serve.

**Cultural Shifts and the Impact on Nonprofit Organizations and Clients**

The intended purpose of performance metrics was to support nonprofit organization efforts to evaluate services, assess and monitor client changes, inform decision making, support the budgeting process, motivate staff, and promote the success of services. Research on performance metrics to date has primarily focused on examining what barriers agencies are experiencing during their efforts to develop and implement performance measurement systems (Carman, 2009a; Carman & Fredericks, 2010; Carnochan et al., 2014; Cordery & Sinclair, 2013; Lynch-Cerullo & Cooney, 2011; Mensing, 2017). Considerably less literature is available that
examines the impact performance measurement systems have upon the vulnerable individuals and communities these organizations serve. The following sections discuss research findings with regard to the impact of performance measurement systems on clients, organizational culture, program planning, and staff management.

**Client Impact**

With the changes that nonprofit organizations have faced to meet the requirements of funders, limited research has focused on how such changes in evaluating programs and services has actually impacted clients over the short or long term. Some studies focusing on identification of performance measurement barriers have also found that nonprofit staff members are concerned about decisions being made related to client outcomes by funders who may not have an implicit understanding of the individual needs of their clients, environmental factors impacting their circumstances, or the most effective treatment approaches to care (Carman & Fredericks, 2010; Carnochan et al., 2014). Performance measurements may be able to capture some level of client process, but the measurement systems are unable to truly “capture the complex progression of improvement” over time (Carnochan et al., 2014, p.7).

Benjamin and Campbell (2015) argue that the reason this complex progression of client goals, needs, or outcomes are not fully captured is due to a lack of consideration for essential aspects of frontline staff members’ direct practice with their clients. “Clients are active agents whose desires, attitudes, needs, and situational constraint play key roles on the change process” (Benjamin & Campbell, 2015, p. 989). Direct service providers have to create relationships with clients, collaborate on identifying their desired goals, and create strategies to support the client. This nuanced individual work with clients can be difficult to capture with an outcomes-based model to performance measurement (Benjamin & Campbell, 2015; Smith, 2010). This approach
can also miss important causal mechanisms that contribute to client change over time and overlook nonprogrammatic outcomes that support client change efforts (Benjamin & Campbell, 2015; Carman & Fredericks, 2010).

Additionally, a client’s pathway is not linear, and any setbacks or challenges can be significant to their eventual progress or success. The challenge remains that a client’s pathway may be perceived as failing from a programmatic perspective, specifically if the results of their behaviors or decisions do not conform to a program logic model (Benjamin & Campbell, 2015; Carnochan et al., 2014). The imposed structure of a logic model has also impacted how some service providers approach their work with clients, reporting that they set goals for clients rather than collaborating with them to do so—while others take on the role of completing a task for a client to successfully meet requirements (Benjamin, 2008; Benjamin & Campbell, 2015). Some direct service providers have even reported allocating their energy towards clients they deemed more likely to succeed in order to improve their program’s outcome data (Benjamin & Campbell, 2015; Carman & Fredericks, 2010; Thomson, 2013). This raises important concerns about the approach to care and potential conflicts with the social work value of respecting a client’s right to self-determination. Performance measurement systems have the potential to impose predetermined definitions of success onto clients, minimizing the client’s agency in generating their own definition of success (Benjamin & Campbell, 2015).

In other circumstances, nonprofit staff felt that mandated metrics placed the full responsibility of change upon the client(s) within a specified time period and failed to recognize the role time can have on client growth (Carman, 2010; Lynch-Cerullo & Cooney, 2011; Thomson, 2010). Programs have focused on immediate outputs rather than long-term outcomes in response to funders incentivizing financial support for more immediate change results (Bach-
Mortensen & Montgomery, 2018). Outcome models also often focus on results with a fiscal year, or specified time frame that is not necessarily reflective of a client’s entry or exit into a program (Bach-Mortensen & Montgomery, 2018; Benjamin & Campbell, 2015). It also does not reflect a multitude of environmental factors that may impact the pace and ability to meet benchmarks or outcomes within a specified timeframe. A key tenet of social work is to ‘meet clients where they are’ while supporting their individual strengths to build on their own goals and self-define success. When the pressure to meet reporting requirements compromises the interactions that providers have with their clients, it is a disservice to the individual and places programmatic self-interests over client-specific needs.

**Planning and Management**

Performance measurements have also impacted nonprofit strategic and programmatic planning and management. The results of performance metrics should support organizations in their program evaluation, decision making processes, and strategic planning efforts. Outcomes development, goal setting, and workload expectations should be informed by evidence and data produced by performance measurement systems. However, there is mixed research to support this theory (LeRoux & Wright, 2010; Moxham, 2014). Several studies found that nonprofit leadership does appear to apply the data generated from performance measurements to inform strategic planning and programmatic decision making (Cairns et al., 2005; Campbell & Lambright, 2016; LeRoux & Wright, 2010; Thomson, 2011). Agencies have developed a greater reliance on metrics to inform decision making for their clients’ long term goals, to increase the effectiveness of specific types of program activities, and to determine service priorities necessary to meet client needs (Campbell & Lambright, 2016; LeRoux & Wright, 2010; Thomson, 2011).
They are also used to make decisions about allocating staff time and resources and necessary financial resources and identifying services that may need modifications (Thomson, 2011).

Although these evaluative systems have increased their use in organizational learning and decision making, several other studies raise concerns about how such systems impact the decision-making culture. Several studies suggest that when funder-mandated tools are used in programs, staff only measure the outcomes that have been prescribed to their services (Carnochan et al., 2014; Charles & Kim, 2016; Kim, 2017). Staff also reported believing that funders only care about the superficial mandated outputs identified rather than other client outcomes (Carnochan et al., 2014; Kim, 2017; Thomson, 2010). In an effort to meet the requirements of funders, results from several studies suggested that there is pressure on managers to compromise the way services are delivered in an effort to meet funder requirements, often impacting quality of care or moving away from the agency mission (Bach-Mortensen & Montgomery, 2018; Carnochan et al., 2014). This also raised the concern that organizations lose autonomy over programming in an effort to meet donor expectations while diminishing a nonprofit’s ability to maintain mission-driven programming (Charles & Kim, 2016; Kim, 2017). Ultimately nonprofit organizations have become tasked with producing data that demonstrates accountability to the communities they serve by ensuring all of their available and allotted resources are allocated to programs. They must also provide data that programs are achieving something measurable and, therefore can be deemed worthwhile of funding, support, and investment (Hoefer, 2000; Moxham, 2014). However, the literature is limited in this area and future research should continue to examine the impact performance measurement has on strategic planning and decision making. Specifically, such research should examine whether
performance measurements (and the changes they intend) are ultimately resulting in quality or effectiveness of services.

**Conclusion**

The emphasis on performance measurement has created an extremely entrenched focus on how funders and policymakers think conceptually about program design, implementation, evaluation, and client services. It has become standard practice for nonprofit organizations to demonstrate impact and quantify how individuals, families and/or communities have been strengthened through service provision, thus creating a case that their funding base has made a positive social return on their investment (Cordery & Sinclair, 2013; Lynch-Cerullo & Cooney, 2011). In some instances, nonprofit organizations are able to develop their outcomes and measurable goals, while others are faced with implementing metrics systems developed by the funders as a requirement of funding (Carman, 2010; Mensing, 2017; Thomson, 2010, 2011).

As a result of this resource dependency, nonprofit organizations are often faced with conforming to the socially constructed and imposed performance measurement system and outcome expectations of external funders, potentially compromising service provision and achievement of goals and agency missions (Carman, 2011; Frayne, 2014; Hillman et al., 2009). Such funder-mandated metrics result in funders defining client markers for success and controlling what measures are used to identify success or failure, with systems often lacking a holistic view of the organizational impact (Carman, 2011; Frayne, 2014; Thomson, 2010).

Although performance metrics provide a structure for evaluation and can help inform organizational decision making, the literature identifies some barriers to implementation (Behn, 2003). The external demand for accountability and measurement is complicated by an extensive body of literature identifying a vast number of barriers that nonprofit organizations face in
successfully implementing metrics systems (Bach-Mortensen & Montgomery, 2018; Carman & Fredericks, 2010; Hoefer, 2000; Kegeles et al., 2005; Lynch-Cerullo & Cooney, 2011; Moxham, 2014; Williams-Gray, 2016). Nonprofits are faced with inadequate resources, a lack of staff training, and minimal organizational support that can create barriers to implementation (Carman & Fredericks, 2010; Cooney & Lynch-Cerullo, 2014).

Additionally, research suggests that an additional barrier to developing and implementing successful performance measurement systems is defining useful and meaningful client outcomes, as well as identifying ways to accurately capture and measure such outcomes (Carman & Fredericks, 2010; Carnochan et al., 2014; Mitchell & Berlan, 2018; Moxham, 2009; Williams-Gray, 2016). Such issues examined throughout the literature raise concern about the impact that funder-mandated metrics have on clients, mission driven program planning, and service implementation (Lynch-Cerullo & Cooney, 2011).

Although the literature has started to examine the implications facing nonprofits when utilizing performance measurement systems, through this review of the research, an important gap in the literature has been identified. As nonprofits continue to use performance measurement systems, the literature has not examined its intended purpose of improving the quality and effectiveness of organizational activities and programs. Research has not adequately examined nonprofit administrator and employee perceptions of how well metric systems support the improvement of services or improve organizational effectiveness in supporting client services. In a performance measurement driven era, research to understand the perceptions of nonprofit leadership and direct service providers regarding funder-mandated performance metrics and the impact on service provision is vital. The gap necessitates a need for continued research on this topic, which will be examined in the present study.
The following chapter focuses on the methodology of this research study. The chapter explains the sampling method, recruitment procedures, data collection process, approaches to increase rigor, and the three phases of coding. It concludes with a description of the use of peer debriefing, reflexive journaling, and audit trail development.
CHAPTER THREE

METHODOLOGY

This qualitative study examines the perceptions of administrators and staff members working within nonprofit organizations of the influence of funder-mandated performance metrics on service provision. The study is based on constructivist grounded theory. This chapter first explores the history and background of grounded theory, followed by participant information and recruitment procedures, discussion of rigor, data collection, and analysis.

Grounded Theory

Grounded theory consists of “systematic yet flexible guidelines for collecting and analyzing qualitative data to construct theories from the data themselves” (Charmaz, 2014, p. 1). A relatively new approach to research, grounded theory was originally defined in the 1967 publication of The Discovery of Grounded Theory by Barney Glaser and Anselm Strauss (Charmaz, 2014; Egan, 2002; O’Connor et al., 2008). This new approach focused on the discovery of theory derived from the data, taking an inductive approach to data analysis (Egan, 2002). As Charmaz (2000) discusses, this approach was revolutionary at the time for several reasons. First, grounded theory challenged the arbitrary divisions between theory and research. Second, it challenged the perspective that qualitative research was largely a precursor to more rigorous quantitative methods. Third, it challenged the assumptions that qualitative research could only produce descriptive data rather than theory development. Since the initial publication
by Glaser and Strauss, grounded theory has continued to evolve along several different lines of use since its first application.

By the 1990s, scholars started to move grounded theory away from the earlier positivist versions of the method after mounting criticism that grounded theory was clinging to an outdated epistemology (Birks & Mills, 2015; Charmaz, 2014; O’Connor et al., 2008). Kathy Charmaz developed the constructivist grounded theory approach after bringing relativity and subjectivity into epistemological discussions around grounded theory. The constructivist approach disavows notions of a neutral observer or a value-free expert and acknowledges subjectivity and the researcher’s role in constructing and interpreting the data (Charmaz, 2008, 2014). It recognizes that the “categories, concepts and theoretical level of analysis emerge from the researcher’s interaction within the field and questions about the data” (Charmaz, 2000, p. 523). In addition, the constructivist approach assumes the “relativism of multiple social realities, recognizes the mutual creation of knowledge by the viewer and the viewed and aims toward interpretive understanding of the subjects’ meaning” (Charmaz, 2000, p. 510). These advances in grounded theory continue to utilize the original methods, taking an inductive, comparative, emergent, and opened-ended approach while working within a more open-ended practice that stresses emergent, constructivist elements (Charmaz, 2000, 2008, 2014). The changes also led to the development of practicing researcher reflexivity and the relativity of the researcher’s perspective throughout the process.

**Constructivist Grounded Theory Approach**

The constructivist grounded theory approach was the most suitable method to examine the research question guiding this study for several reasons. First, there was limited research and
literature examining the perceptions of nonprofit administrators and staff members regarding the influence of funder-mandated metrics on service provision. This resulted in an exploratory study and approach to data collection that aims to identify new information. Grounded theory is utilized for creating theory and widely used in areas that are considered more exploratory or where little is known about the area of study (Birks & Mills, 2015; Charmaz, 2014).

Grounded theory is also the preferred method when the intent is to generate new theory with explanatory power regarding the topic (Birks & Mills, 2015; Charmaz, 2014). The goal of this study was to generate a theory related to the perceptions of participants utilizing funder-mandated metrics and their influence on service provision. Data collection, coding rationale, analysis, integration of categories, and construction of theory were guided by the grounded theory process. The grounded theory process provides flexible guidelines for collecting data and making sense of the data. This flexible approach is useful when making meaning of the participants’ perceptions. The incorporation of new ideas into the original topic could occur during data collection while moving through continued stages of data analysis. This autonomy permitted me to follow any leads that emerged and to keep an open mind, allowing the ideas to emerge directly from the data (Charmaz, 2014).

By taking a constructivist grounded theory approach, the study acknowledged subjectivity and my involvement in the construction and interpretation of the data. The interpretive method required the researcher to enter the participants’ world, and the constructivist approach focused on preserving the participants’ dignity by making concerted efforts to learn about their views and perspectives (Charmaz, 2014). Grounded theory does not force preconceived ideas and theories onto the data. Rather, the iterative analysis process follows leads
that are defined in the data. The grounded theory approach also included reflexivity practice, which assisted in addressing or identifying any bias or perceptions that could impact analysis. While the researcher does not necessarily adopt or reproduce the participants’ views as their own, the interactions between them should not unwittingly reproduce the researcher’s assumptions. Utilizing a grounded theory approach provided a process for collecting data that prioritized theory derived from new information that emerged organically through analysis.

Additionally, grounded theory methodology has several key aspects that support a rigorous methodological approach. First, grounded theory has strategies that guide the researcher through the data collection and analysis processes. When adopting these systematic guidelines, the study can be clearly directed, efficiently managed, and data collection process streamlined to construct an original analysis of the data (Charmaz, 2008, 2014; Hussein et al., 2014). This study utilized the process developed by Charmaz (2014), which takes a constructivist approach. The process of continuously analyzing data during the collection and analytic phase allows the theory to be developed from the data and not a prescriptive formulaic process. This provided rigor and ensured the trustworthiness of the emerging theory while also creating enough evidence to support claims (Charmaz, 2014; Hussein et al., 2014).

Second, the grounded theory approach includes strategies that are self-corrective in nature (Charmaz, 2000). Grounded theory logic requires the researcher to go back and forth between the data to refine emerging theoretical frameworks, providing a fresh look at concept or category creation (Hussein et al., 2014). This constant comparison process supports the emergence of theory and provides a heightened level of rigor when compared to other qualitative approaches (Charmaz, 2008). Engaging in constant comparison and practicing reflexivity,
provided a way to address researcher bias and to refrain from ascribing motives or personal beliefs onto participant responses (Charmaz, 2014).

Third, grounded theory methods produce data that has depth and richness which provides a researcher with “a concrete and dense fabric to construct a thorough analysis of the data in addition to aiding the researcher to go beneath the surface of the participants’ social and subjective life” (Hussein et al., 2014, p. 5) To develop rich data, researchers engage in journaling and create an audit trail that result in dense descriptions of the process and procedures. These tools guided the meaning making process and refined it to produce insight regarding participant experiences (Charmaz, 2008; Hussein et al., 2014). Last, the grounded theory process focuses on data analysis that supports the construction of a theory that arises from the data. The theory is an “explanatory scheme that comprises a set of concepts related to each other through logical patterns of connectivity” (Charmaz, 2014, p. 108) and explains what was discovered about the research question while posing a theory to explain the phenomenon examined.

Participants

The grounded theory approach to sampling in this study began with a purposive selective sample. The selective sampling process included identifying the populations and settings prior to the data collection and served as the “theoretical jumping off point” from which to begin theory development (Draucker et al., 2007, p.1138). This approach maximized the amount of information collected for data analysis (Birks & Mills, 2015; Charmaz, 2014).

A purposive selective sample of 16 direct service providers and administrators from eight nonprofit organizations was included in the study. The rationale to include two individuals per agency was to provide diverse perspectives on how funder-mandated performance metrics
influence service provision. Participants were recruited from multiple agencies within one major metropolitan city located in the Midwestern United States. Detailed information about the participants is included in chapter four.

**Inclusion Criteria**

In this study, maximum variation purposive sampling was used based on the following criteria. Given the scope of the study, participating agencies had to be classified as a not-for-profit organization with 501c3 status. For-profit human service organizations were not included since the absence of reliance on funders may not create the same issues in developing, implementing, or sustaining performance measurement systems. Selected agencies for this study were required to have mandated performance measurement metrics as required by an external funder.

Financial constraints and access to resources also differ between organizational budget size. Following the financial categories of the National Center for Charitable Statistics and Urban Institute National Survey criteria, two nonprofit budget categories were chosen for inclusion criteria in this study. The categories’ ranges included budgets of: $1 million to $4.99 million and $5 million or more (Pettijohn et al., 2013). Four agencies were included in the $1 million to $4.99 million and four agencies were included in the $5 million or more category.

Nonprofit organizations with budgets less than $100,000 were not included in this study, as they receive the smallest amount of grant funding and were least likely to have funder-mandated requirements. Nonprofits with budgets between $100,000 to $1 million were not included due to accessibility and capacity issues faced during the COVID-19 global pandemic. Smaller agencies often did not have the staff required for the study, had limited capacity due to
the increased demands on the agency to provide services remotely and online, or due to laying off staff members because of budget cuts and temporary program and/or agency closures.

In addition to these financial categories, the nonprofit could not receive funding from one primary source (more than 49 percent of the annual budget) such as United Way or via a government contract. This was done intentionally to learn more about multiple metric systems (both public and private) that nonprofits may have to implement. Each of the agencies also had to have a minimum of four programs or services offered at the agency to ensure they had metrics to support more than one program.

The nonprofit organizations likewise had to provide services that fall under the Human Service category (Pettijohn et al., 2013). This category is the most inclusive of nonprofit organizations that provide direct client services and provides the most extensive sampling frame. Agencies additionally had to be located within a major metropolitan area in the Midwest, underscoring a focus on urban organizations.

Lastly, with regard to sampling for this study, within each of these organizations, two participants were interviewed, one individual who held an administrative leadership position (e.g., chief executive officer, executive director) and the second individual who was a direct service provider in a leadership position (e.g., program director, program coordinator). These participants expressed their interest in being interviewed via email and I then corresponded with each person to identify their position at the agency and ensure one person from each category was met. Interviewing individuals from different employee groups in each organization provided insight as to how perceptions of funder-mandated performance metrics vary related to the impact upon service provision.
Recruitment and Procedures

Nonprofit organizations were identified through existing relationships with agencies in the community. I am employed by two universities to teach and provide administrative support within their social work departments. Both universities provided their social work internship site list to assist with recruitment. To ensure that the nonprofit organizations met the inclusion criteria, I cross-referenced agency annual reports and 990 forms. After identifying agencies that met all criteria, an email was sent out to key personnel asking them to forward the study recruitment information and email (Appendix A) to any staff members who may have an interest in participating. Additionally, emails containing information about my research were sent out to colleagues who work with various internship sites across the city.

Potential participants were asked if they wished to participate in a research study that would explore the influence of funder-mandated metrics on nonprofit service provisions. Two individuals per agency were identified—one individual in an administrative position and the other in a direct service leadership position. Once two individuals from the agency reached out to express their interest in being interviewed, both employees were provided with the consent form (Appendix B) and assured that their participation would in no way be identifiable. Participants were then informed that the interviews would last between one to two hours and that should they wish to terminate their participation or interview at any time, their information would be removed from the study and destroyed. All participants were informed that their interviews would be confidential and that interview transcriptions would be deidentified and assigned random pseudonyms to protect their identity. Prior to scheduling the interview, consent forms and a brief demographics survey were completed and returned via email or mail.
I was the only person with access to the list of participant names and corresponding pseudonyms used for analysis, which were kept on a secure and locked computer. Initial in-person interviews were conducted at a location of the participant’s choosing to enhance their comfort with the interview. The three in-person interviews all took place at the individual’s place of work. Due to a change in study procedures in response to agency closures and lockdowns associated with the COVID-19 global pandemic, subsequent virtual interviews were conducted over the phone or via Zoom. All participants who participated via phone or Zoom connected while working from home during the COVID-19 pandemic mandated stay-at-home orders. All participants were compensated for completing an interview and received a $10 Starbucks gift card via email.

Details of Confidentiality and Privacy Protections

The consent form (Appendix B) provides a detailed explanation of the research project and procedures, including information regarding protection of confidentiality. All of the participants were informed that the interview content would be kept confidential and that names and other identifiable details regarding the agency and participants would be removed or de-identified prior to the publication of this dissertation and any subsequent peer-reviewed manuscripts or conference presentations. This increased the intentions toward ensuring confidentiality, but complete confidentiality cannot be fully guaranteed. With the permission of each participant, an audio recording was created, transcribed, and used for data analysis. All data, including participant contact information, were stored on a password-protected computer, and recordings were uploaded through a secure site for transcription. Data will be stored for a
period of seven years in a secure and locked location with password-protected encryption per the expectations set by the Institutional Review Board (IRB) of Loyola University Chicago.

**Protection of Human Subjects**

This dissertation adhered to all research ethics guidelines of the Loyola University Chicago, Institutional Review Board (IRB). This study involved minimal risk to individuals participating in the research interviews. Participants were provided written information regarding the study prior to the interview to ensure they understood the topic and purpose of the study. Participants who agreed to an interview were required to provide written consent prior to participation in the study. Each individual was informed that their participation was voluntary and that they could withdraw from the study at any time, without consequences. Participants were also informed that if they chose to withdraw from the study, at any time, all information, including interview audio recordings, would be removed from the study and destroyed. However, none of the participants preemptively withdrew from the study.

Participants were also informed that interviews would be audio-recorded on a digital recorder for in-person interviews and via Zoom for virtual interviews. Additionally, they were informed that they had the option to decline being recorded at any point in the interview process, with an alternative option to complete the interview with the researcher taking notes. However, none of the participants asked not to be audio-recorded, nor did they request recordings stop at any point during the interview process.

The potential benefits of this research included contributing to a better understanding of the influence of funder-mandated metrics on nonprofit service provision. Participants in this study may also have benefitted from knowing that the narratives provided would contribute to
the larger knowledge base in this area and provide insight into the unique relationship dynamics between funders and nonprofit organizations and their staff members. Findings from this study may identify additional ways that funders can support nonprofit organizations and how nonprofits can successfully navigate and manage funder relationships. Furthermore, each participant received a $10 gift card as gratitude for their participation in the study and as compensation for their time.

**Human Subjects Ethical Implications During COVID-19**

Participant interviews began in March of 2020, with 18 in-person meetings scheduled throughout the month. However, due to the COVID-19 global health pandemic, the governor of the state and mayor of the city enacted a mandatory stay-at-home order effective mid-March which lasted through May 30, 2020. This order required all nonessential employees to work and stay-at-home, except for essential needs.

Due to the stay-at-home order, I submitted an amendment to the IRB to expand research procedures and interview participants virtually. I emailed the remaining 15 participants and asked them to reschedule at a later date once the IRB amendment was approved. The IRB amendment included language modifications to the informed consent letter to adjust for virtual interviews over the phone or via Zoom. For interviews conducted via Zoom, only the audio files were saved to a password-protected computer and destroyed after the study was complete. The interview protocol and questions (Appendix C) were also amended to ensure additional verbal consent was given at the start of the virtual interview.

The amendment was modified to ensure demographics and consent forms could be provided via mail or email. Upon receipt of emailed documents, participants were asked to sign
and return the consent form to my university email address. Emailed consent forms were printed and stored in a locked file cabinet. The original email was deleted for security. Mailed forms were sent to the participant’s work address and included a self-addressed and stamped envelope for return. The amendment also specified that gift cards could be emailed or mailed to participant work addresses.

In addition, the majority of the participants worked at nonprofit organizations that similarly found themselves mandated with COVID-19 pandemic compliance efforts and safety protocols for the delivery of essential services (Goolsbee et al., 2020). Many personnel within the organizations had to make decisions to close their programs and agencies, identify ways to offer remote services, and best serve clients during the stay-at-home orders and during a series of evolving Centers for Disease Control and Prevention (CDC) guidelines and mandates (Maher et al., 2020). For all nonessential personnel, nonprofit organizations had to identify ways for employees to work remotely and manage family responsibilities, such as children without childcare or attending school remotely (Maher et al., 2020).

Many of the participants and the agencies participating in this study found themselves in crisis response mode. After emailing all participants to inform them of the need to reschedule pending IRB amendments, I consulted my dissertation committee members to discuss the best ethical approach to reaching back out to study participants while asking them to reschedule interviews during very stressful and unprecedented circumstances. With guidance from my committee members, an email was sent to all study participants acknowledging their difficult circumstances and informing them that I would contact them a month later to discuss the possibility of rescheduling. Several participants responded to the email, stating that they were
willing to reschedule within the next month. Others were open to being in contact in a month, and a few others asked to be contacted at a later predetermined date. Over the next six months, virtual interviews were rescheduled with 13 of the original 15 remaining participants, and two participants withdrew from the study.

**Data Collection and Instrumentation**

Once an agency was identified and two participants meeting the specified study inclusion criteria expressed interest in being interviewed, consent forms were provided. Interviews were scheduled with each individual after they expressed consent and completed the form. A demographics form was also provided to participants before the interview (Appendix D). In-depth interviews were then conducted, lasting between 60 and 90 minutes in total. The interviews utilized a semi-structured approach to maximize the breadth of interactions between me and the participant (Birks & Mills, 2015; Charmaz, 2008; Denzin & Lincoln, 1994). In alignment with a constructivist approach to interviewing, the approach was dialogical, allowing participants to share their perspectives through conveying stories and lived experiences (Patton, 2015). Interviews were facilitated by asking open-ended questions, using planned probes (Appendix C), and unplanned follow-up questions dependent on each interview (Patton, 2015). Participants interpreted the questions within the context of their own experiences, and I asked for specific examples or more details when necessary.

Interviews for each agency were scheduled and conducted on separate days to allow time to engage in reflective journaling and the constant comparison process. Journaling took place after each interview to track ideas, biases, and thoughts related to each participant’s story, or any additional information that might impact the data. Each interview was audio-recorded and
downloaded onto an MP4 file immediately after the completion of the interview. Within 24 hours of the interview, files were uploaded for transcription, and an online service, Cabbage Tree Solutions, was used to transcribe audio recordings into transcripts. Transcribed interviews were completed within 72 hours of the interview. I then listened to the audio recordings while reviewing the transcripts to ensure the interview was accurately transcribed, while correcting or clarifying any mistakes. The transcribed data was then uploaded into NVivo for coding and analysis.

**Support for Rigor of Design**

Qualitative research models need to ensure the rigor of the design, credibility of the findings, and interpretation by taking steps to establish trustworthiness. Lincoln and Guba (1985) developed a model for assessing trustworthiness and identified four aspects of trustworthiness that include (a) truth value, (b) applicability, (c) consistency, and (d) neutrality. Krefting (1991) and Lincoln and Guba (1985) established a detailed description of each of the four aspects. Truth value asks if the researcher has established confidence in the truth of the findings for the participants and the context in which the study was conducted. This established how confident the research is with the truth of the findings based on the design. Applicability refers to the extent to which the study’s findings can be applied to other contexts and settings or with other groups. Consistency focuses on whether the findings are consistent if the inquiry were replicated with the same participants or in a similar context. The fourth criterion is trustworthiness, neutrality, which refers to the degree to which the study’s findings are determined by the participants and not by the researcher’s biases, motivations, interests, or perspectives.
Addressing these aspects was an essential part of the study design to increase the study’s rigor and for readers to use to assess the value of the findings. Lincoln and Guba (1985) identified four criteria to address the different aspects of trustworthiness: credibility, transferability, dependability, and confirmability. In addition, they identified strategies for each criterion that can be employed in research design. The following four sections discuss the key strategies of trustworthiness utilized in this study.

**Credibility.** The qualitative approach to address trust value is credibility. Credibility is defined as the degree of fit between respondents’ views and the researcher’s description and interpretations (Lincoln & Guba, 1985; Padgett, 2008). Three approaches were undertaken in this study to increase credibility.

First, I focused on taking steps to ensure credibility during the interview process. The reframing or expansion of questions on different occasions was employed (Krefting, 1991; May, 1989). These techniques were applied during the interview process and support the interview’s internal consistency (Krefting, 1991). Utilizing the constant comparison method between interviews, I was also able to reflect on my interviewing techniques, adjust any questions, or identify additional probe questions to add to the interview process.

Second, I engaged in peer debriefing with colleagues at different points of the data collection and analysis process. Lincoln and Guba (1985) suggested that peer debriefing is a way of keeping the researcher honest by discussing biases and supporting awareness of the researcher’s posture and process throughout the study. This can reduce researcher bias and provides the ability to practice reflectivity (Padgett, 2008). Likewise, it provides an opportunity to discuss code or theories that may emerge through the grounded theory process (Creswell,
Finally, member checking is a technique that determines the accuracy of the findings, which consists of taking the data, analytic categories, interpretations, or conclusions back to the study participants to determine whether they feel the data are accurate (Creswell, 2014; Krefting, 1991; Lincoln & Guba, 1985). Member checking can also address threats of reactivity, researcher bias, and respondent bias (Padgett, 2008). Reviewing parts of the data provides an opportunity for participants to assess intentionality, correct errors of fact, check for adequate representation of the participants’ realities, and confirm the accuracy of the investigator’s interpretations (Creswell, 2014; Lincoln & Guba, 1985). During the consent process at the start of the interview, participants were asked to be a reviewer during the final stages of the themes and theory development process. Participants were emailed (Appendix E) at the work email address provided and asked if they would like to participate in the member checking process. I asked that participants review the findings and identify whether they felt their responses were represented in the theme and theory description. Additionally, they were asked to reflect upon their contributions and identify if they aligned with the proposed findings. All 16 participants were emailed individually, and 7 email addresses bounced back. From the automated responses, I was informed that these individuals were no longer working for the organization due to planned retirements, new jobs, layoffs, family obligations, or unknown circumstances. Of the remaining 9 participants, three did not respond to my email, two individuals responded that they did not feel it necessary to review the findings, and 4 individuals asked to review the findings.
The chapter four findings were then emailed, a timeframe for review was agreed upon, and an offer to meet with them via phone or zoom was reiterated from the initial email. The four individuals all responded via email to share that they had found no discrepancies, felt their contributions to be represented and aligned with our interviews. One participant noted that while some of the findings did not always align with her agency experience, she recognized that other organizations might have had those experiences. She also added that from her other work history, she felt they aligned with those experiences and stories she heard from peers working in the nonprofit industry.

Transferability. Transferability addresses the applicability criterion of trustworthiness and the issue of applicability of the study to other populations or similar groups. (Krefting, 1991; Lincoln & Guba, 1985; Patton, 2015). I utilized three strategies to increase transferability. First, dense descriptions were provided about the eight nonprofit organizations in which the 16 participants worked. Providing a dense description allows for researchers to assess the finding’s transferability. Lincoln and Guba (1985) noted that it is not the researcher’s job to provide an index of transferability, but it is their responsibility to allow transferability judgments to be made by others. The dense descriptions provided earlier in this chapter ensured conveyance of necessary information to enable someone interested in the study to conclude whether the study has transferability.

In addition, I engaged in reflexive journaling and documented a detailed audit trail, while journaling via NVivo during the recruitment and rescheduling process and after each interview for a total of 26 journal entries. The audit trail was a running document that started when developing protocols to submit for IRB, and notes were entered through the end of data analysis.
The audit trail noted steps taken, as well as the decision-making process throughout the study. Both enhance reproducibility and transferability (Lincoln & Guba, 1985; Padgett, 2008).

**Dependability.** Dependability addresses the consistency criterion of trustworthiness and focuses on the level at which a study’s procedures are documented, traceable, and have logical sense to others (Krefting, 1991; Lincoln & Guba, 1985; Padgett, 2008). I also utilized reflexive journaling and documentation for the audit trail to increase the dependability of the study. The journaling served to document decision-making processes during interviews, coding ideas, emerging themes, thought processes, and information pertaining to the interpretation of the interviews. The audit trail recorded the steps taken throughout the entire research process, making it easier for others to understand the logic that led to my conclusions.

**Confirmability.** Confirmability is defined as the level to which the study’s findings are linked firmly to the data (Lincoln & Guba, 1985; Padgett, 2008). As described in the previous section, I created an audit trail throughout the entire research process. It was used to enhance the confirmability of the findings and provided a detailed explanation of the process to make meaning out of the interviews and during analysis.

The final technique, reflexive journaling, had a broad range of applications for all four areas and was maintained during the data collection and analysis (Krefting, 1991; Lincoln & Guba, 1985; Padgett, 2008). As discussed, reflexive journaling took place during the recruitment process and after each interview. It likewise contributed to the creation of an audit trail, documenting decision-making processes. I could use the journal to identify key information to include in the audit trail, enhancing the ability to verify the findings.
As a macro practitioner who has worked in leadership positions within nonprofit organizations, it was critical for me to engage in journaling to practice reflectivity, keep audit trail notes, and engage in peer debriefing. Engaging in both reflexivity and documentation increased the study’s rigor and trustworthiness and reduced the threat of researcher bias (Padgett, 2008).

In addition, I was able to use my journals and audit trail notes during peer consultation. While working from a constructivist grounded theory approach, the peer review process was utilized to provide support throughout the research process. It was essential to discuss and reflect on my own bias, which aided in obtaining strong objectivity in the research process (Lincoln & Guba, 1985). The journaling and audit trail process are discussed in more detail in the following data analysis section.

Data Analysis

Data analysis followed the procedures developed by Charmaz (2014), which delineates a three-phase process. After completing transcription, these phases focus on initial, intermediate, and advanced coding procedures that correlate with and feed into low, medium, and high-level conceptual development (Birks & Mills, 2015). I began the analysis process after all interviews had been completed and transcribed. The following section details the process of transcription, coding, reflective journaling, and the development of an audit trail.

Transcription

Once each interview was completed, I uploaded the audio recorded MP4 file to the secure online transcription website service, Cabbage Tree Solutions. Interviews were transcribed and returned within 72 hours. I then listened to each interview again to cross-check for accuracy and
clarify mistakes or jargon, continuing to journal while listening to the audio recordings to capture any questions or ideas that arose. Notes were taken on key sections or participant quotes for further examination during the constant comparison process between interviews. The continued journaling during transcription editing provided an additional opportunity to see how each interview built upon the previous one. I made a note of any patterns in the transcription review process for later review during analysis. After each transcribed interview had been reviewed and cleaned, I uploaded the files to NVivo, and created a coded system to deidentify each participant’s file in the database.

**Initial Coding**

Initial coding focused on low-level conceptual analysis of the data, and Charmaz (2014) discussed the need to ask these questions of the data (1) What is this data a study of? (2) What do the data suggest? (3) From whose point of view? and (4) What theoretical category does this specific datum indicate? This process included line-by-line coding, which meant naming each line or segment of the written data. The line-by-line approach was helpful in the beginning stages of analysis to help gain a sense of conceptual control over the data and identify otherwise undetected patterns (Birks & Mills, 2015; Charmaz, 2014). Two other steps were taken during the initial coding process, which included reflexive journaling, and the constant comparison method.

The first round of coding took place over six weeks. Before I started to code an interview, transcripts were reread, and corresponding journal entries were reviewed. This process enabled me to become familiar with the interview and become immersed in the data once again. This additionally promoted a closeness to the data that is effective during coding and analysis (Birks
& Mills, 2015; Charmaz, 2008, 2014). Using NVivo, I moved through each line of the data, applying codes to individual lines or segments. The initial codes often used direct participant language when creating a new code to preserve the participants’ meaning and start to identify common phrases and language used by participants. An example of initial coding using participants’ language is included in Table 2.

Table 2. Example of Initial Coding in NVivo

<table>
<thead>
<tr>
<th>Interview excerpt</th>
<th>Initial code</th>
</tr>
</thead>
<tbody>
<tr>
<td>“We tend to base a lot of our metrics on <strong>industry standards</strong> and looking at national research and our peers’ standards to see what good performance measurements are.”</td>
<td>Industry standard of performance</td>
</tr>
<tr>
<td>“Well, now our department, it recently became the department and thinking about how are we measuring resiliency? <strong>How do we measure these things that are pretty intangible</strong>, but funders want to know what are you doing, right? So, kind of like helping us think some of those things through.”</td>
<td>Difficulty defining performance</td>
</tr>
</tbody>
</table>

I coded the first four interviews using a combination of line-by-line coding and coding by phrase or complete thought. This approach captured the participants’ examples or explanations, which were often several lines long. After completing coding of the first four interviews, I met with doctoral committee members to discuss my process, code development, organize the codes in NVivo, receive feedback and adjust, and then to continue to code the remaining interviews.

The initial phase of coding required constant reflection on the early analysis decisions being made (Birks & Mills, 2015). Reviewing my notes and engaging in constant comparison between interviews enabled me to ensure the fit and relevance of analysis by the emergence of topics or patterns emerging from the data. For example, when I noticed participants discussing
similar topics but with different language, a journal entry would be made to note thoughts on the possible connection or similarity. This also provided a documented process to utilize during the constant comparative process. The process of comparing data between interviews allowed me to start to identify similarities and differences (Charmaz, 2014). It was making sequential comparisons that aided in code development, helped me identify gaps in the data collection, and determined if theoretical saturation had been met. Saturation is the conceptualization of the data, which yields different pattern properties until no new patterns emerge (Charmaz, 2014). At the end of the initial coding of all 16 interviews, there were 127 codes with no gaps or other outliers that needed to be addressed through additional interviews.

**Focused Coding**

The next phase of analysis included focused coding, which concentrated on studying and assessing the initial codes and aided in the continuation of the comparative process. One of the goals of focused coding is to determine the conceptual strength of the initial codes (Charmaz, 2014). This process included assessing the initial codes and comparing them with the data to distinguish which codes had the greatest analytic power. Comparing the codes assisted in identifying the direction of analysis and in clarifying ideas (Charmaz, 2014).

Starting this process, I reviewed all initial 127 codes, examining the original data to identify the strength of the codes and their contribution to the analysis. I pinpointed any codes that could be collapsed or removed from the code list. In several instances, I recognized that several codes had been created to cover the same topic and should be combined. Other codes only included one or two pieces of data from interviews and were best suited in another more applicable code, or the code was removed from the codebook. As this process continued, notes
were kept of the decisions made, noting codes to revisit during the next iteration of focused coding, and any questions or topics to discuss during the peer review process were noted.

The large number of initial codes resulted in four rounds of focused coding. A strength of the grounded theory approach to analysis is that codes are derived from the concentrated, active involvement of the focused coding process. Interacting with and acting upon the data creates an active approach, rather than passively reading through them (Charmaz, 2014). The focused coding phase was not a linear process, as I often returned to earlier codes or notes to initially explore topics that were not easily discernible. Consistent with the grounded theory approach, initial and focused coding is an emergent process (Charmaz, 2014). Throughout this process, I found that one telling code might illuminate another.

Next, I began to identify key codes, as well as recognize and link parent codes and sub-codes (Birks & Mills, 2015). In NVivo, parent codes are an overarching category that pulls together multiple sub-codes into a more meaningful theme or topic for continued analysis. Through this process, I started to identify what codes were parent codes or if a new code needed to be developed. The ongoing process heightened my sense of direction for the analysis and clarified the centrality of specific categories and theme development.

As I worked through the focused coding phase, code descriptions were clarified in the evolving NVivo codebook, and the audit trail documented the ongoing process. The focused coding began to enter the final phase when I was not able to remove, combine or collapse the existing codes. The codebook at this point consisted of 87 parent and sub-codes. However, the focused coding process was not yet complete, as I grappled with the unique ways I saw different codes influenced different parts of services provision throughout a nonprofit organization. I went
back through to review personal journals, audit notes and comb through the codes. During this process, I started to recognize that there were different categories of codes covered. From there, three broader categories were identified that better captured what was happening with the data. As the themes emerged, they fell under one of these three categories: (1) client, (2) program, and (3) agency (see example in Table 3). I then went back through the codes to confirm that there was evidence of different themes emerging under these different categories.

Table 3. Example of Client Category and Theme

<table>
<thead>
<tr>
<th>Category</th>
<th>Focused coding</th>
<th>Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client</td>
<td>▪ Industry standards</td>
<td>Definition of client success</td>
</tr>
<tr>
<td></td>
<td>▪ Funder defined success/performance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Difficulty defining success</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Struggling to measure success concept or impact</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Process definition of success</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Outcome definition of success</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Individual definitions of success measured collectively</td>
<td></td>
</tr>
<tr>
<td></td>
<td>▪ Redefining success</td>
<td></td>
</tr>
</tbody>
</table>

For the final round of focused coding, I reviewed the data and provided clarification to the codes that fell under specific categories. Table 4 shows an example of a parent code, focusing on the overall influence of time on nonprofit programming.

I continued to complete this process until clearly identifying when codes fit into specific categories and when they did not. These denser categories helped explain larger pieces of data and started to identify the foundation for the theoretical coding process (Birks & Mills, 2015; Charmaz, 2008, 2014). Once I could not add, combine, or collapse any more of the existing codes, and I could not identify anything new in the data, I concluded the focused coding phase.
Table 4. Example of Focused Coding in NVivo

<table>
<thead>
<tr>
<th>Interview excerpt</th>
<th>Focused code</th>
</tr>
</thead>
<tbody>
<tr>
<td>“I mean some of the, um, it’s interesting because it takes away from some of their program time, obviously. If we’re doing [evaluation] twice within one season and then once within the other season, that’s taking away from program time, then some of the questions can seem really intrusive, so we’ve done a lot of work like doing kind of mini-tutorials and trainings with our coach to do messaging.”</td>
<td>Program time influence</td>
</tr>
</tbody>
</table>

The final codebook includes 106 codes and sub-codes organized into three overarching categories (see Figure 1).

Figure 1. Three Categories from Focused Coding

After the focused coding process, the client category had three themes, the program category had four themes and one subtheme, and the agency category had four themes. While each category had clearly identified themes, I still needed to examine the relationships between them and look at how everything connected. The next phase of analysis further discusses the analysis process and each category with its related themes.
Theoretical Coding

Theoretical coding was the final phase of coding. Glaser (1978) introduced theoretical codes to conceptualize how the substantive codes may relate to each other as hypotheses and be integrated into theory. This phase focuses on making sense and meaning of emerging theories and clearly explaining themes arising with the data. Utilizing the categories and themes that emerged during focused coding, I decided to develop a concept map as a tool to examine possible relationships. Concept mapping is a technique that aids in visualizing relationships between various concepts (Wheeldon & Faubert, 2009).

This technique provides a complementary strategy to Charmaz’s (2014) coding process and aligns with the constructivist approach to grounded theory research. Researchers cannot assign value to meaning without acknowledging the role they personally play within the construction (Lincoln & Guba, 1985). Concept mapping is used as an additional way to practice reflexivity, using a visual approach to examine my contributions to the construction of meaning. Concept mapping also offers a creative way to engage with the data. Having the ability to visualize and move the themes around provides a secondary tool to view the participants’ experiences and perceptions (Wolcott, 2008). It represented a way for me to design the following analysis stage, using the participant-generated themes to guide a more in-depth analysis (Birks & Mills, 2015; Wheeldon & Faubert, 2009).

In preparation for concept mapping, I printed out a final codebook and attached each parent and sub-code to an index card. I then taped them all onto a wall in the same order as the codebook (Figure 2). Codes were examined from a different visual perspective than via NVivo providing a larger space to move the codes around physically. I initially struggled with the
theoretical coding phase, attempting to do this work solely on the computer. Creating the tangible and moveable map allowed more flexibility for thought processes and allowed the participants’ voices to guide the process more clearly. This visual representation of the data also allowed me to evaluate thought processes and visually examine any ideas noted via reflexive journaling.

Figure 2. Example of Concept Map in Focused Coding Phase

Next, on a separate wall, I placed the research question at the top of the concept map and added the three categories developed in the focused coding phase which included the themes of: client, program, and agency. Under each category, codes were added that emerged during focused coding. Journal entries were reviewed while examining each code and notes kept of the
emerging connections and concepts under each category. The visual representation of the process, coupled with the journaling, assisted me in identifying my biases or assumptions and aided in organizing the ideas that emerged without being locked into one particular way of organizing or thinking about the data.

From here, I was able to group and regroup codes together under one of the three categories. This process allowed me to engage in a different type of analysis and review and revisit initial themes that had emerged under each category during focused coding. The process of moving the data and examining it all at once assisted with processing and structuring the data while better understanding the connections between the different themes. There were a few instances where, upon reflection and engaging with the concept map, it became apparent a code was better suited under a different theme or raised questions about which category a theme should connect. When questions were raised about themes, I had to examine the content to clarify what issues were to be grappled with, followed by identifying through the data what decision were to be made.

Several iterations of the evolving map and peer consultation assisted with the final coding phase. The concept map was another way to affirm that saturation was reached with the interviews (Charmaz, 2008, 2014). When I was no longer adding new codes or categories to the map after the interviews, it confirmed the assumption that saturation had been reached during the initial coding phase.

During the final theoretical coding phase, I documented the process visually and through journal entries. Identifying the primary themes under each category allowed for a visualization of the directionality of connections and to clearly identify linkages as the data analysis progressed.
The map also aided in showing the linkages between the data (Figure 3). While the concept map did not drive the data analysis process, it did aid as a tool assisting with the thought process and provided a flexible visual representation to work through my ideas and concepts. Lastly, I could refer back to the map iterations throughout the theoretical coding phases and explore the ideas and concepts while writing down an understanding of the theory emergence.

Figure 3. Example of Concept Map in Theoretical Coding Phase

During the process of meaning making, it was important to avoid forcing data into preconceived codes and categories (Birks & Mills, 2015; Lincoln & Guba, 1985). Continuing to engage in reflexive journaling and peer debriefing, I was able to think critically about the preconceptions while engaging in the iterative theoretical coding process.

For example, several themes emerged which are commonly discussed in the literature. During a peer debriefing session, I discussed how the data aligned with the literature, when it
was pointed out that such a conclusion was not made through data analysis. I assumed that the themes were based on alignment with the literature. From here, I was able to go back and examine the data again and draw conclusions.

Once the concept mapping process was completed, a theoretical model focusing on concepts and connections was developed. The research question was revisited once again to guide this process: What are the perceptions of administrators and staff members within nonprofit organizations regarding the influence of funder-mandated performance metrics on service provision? I focused on how the data answered the “what” in the question and then I started to organize the data into information that answered the question.

Once a model was sketched out, the focus moved to writing out an explanation in a long journal entry. The write-up focused on identifying connections and concepts and capturing ideas regarding implications for social work or future research. This extensive write-up was then reviewed and discussed in an intensive two-hour peer debriefing session with committee member Dr. Perkins. I focused on explaining my findings, discussing questions from my notes, and processing through any questions raised. After this, I went back through the document, made edits and clarifying points, and saved the final copy to guide the writing of the theory development. Details of theory development are presented in chapter five.

**Peer Debriefing**

During the interview process, I engaged in peer debriefing with members of my doctoral committee. This process was essential to reflect on the interview process and to use as a sounding board while navigating the changes taking place during the COVID-19 global pandemic. I discussed engagement approaches in a virtual environment and the ongoing changes
to participants’ work environments throughout the interview process. During these Zoom meetings, I was able to discuss my thoughts, concerns, ideas, and talk through anything from an interview via notes made in the reflective journal. This process took place throughout the six-month span of scheduling and rescheduling interviews.

In addition, I engaged in peer debriefing during the data analysis phase. After the first round of initial coding was complete, continued consultation was made with dissertation committee members, Dr. Megan Farwell, and Dr. Johanna Barry for ongoing guidance and feedback. The doctoral committee members aided in the process of coding, assisting in developing what approach was best for the focused and theoretical rounds of coding. After each iteration of the codebook, meetings were held with Dr. Perkins to discuss areas where codes could be clarified, creation of the node and sub-node system, and identifying when to combine codes if they were addressing the same or similar topics.

Dr. Farwell was available to discuss ideas and the process of developing several iterations of my codebook. During the coding process, Dr. Farwell assisted in the evolution of a coding scheme and process at the focused and theoretical coding stages. She also aided in keeping me accountable to the findings and discussing or challenging any ideas that the data may not have supported. She would encourage me to reflect on personal beliefs, experiences, or biases that may place judgment or values on codes, rather than allowing the interview data to guide the process.

I also relied on informal weekly meetings with doctoral program colleague, Dr. Johanna Barry. Dr. Barry was likewise in the process of completing her qualitative dissertation, and we met virtually to talk about our projects, ask questions, share NVivo knowledge, discuss struggles
and offer feedback to one other. Dr. Barry had completed her interviews and analysis process a few months prior and offered insight into her own learning experiences. It was immensely beneficial to discuss ideas for reflexive journaling, documenting an audit trail, and her process in creating an evolving codebook. The peer debriefing process with these individuals increased the credibility of the study, strengthening the data collection and analysis process.

**Reflexive Journaling**

At the end of the data collection process, I had written 26 journal entries, and following a review of the transcribed data and analysis, had 34 entries. These journal entries included general thoughts after each interview, questions raised, notes on the constant comparison process, and any personal feelings related to the interview or the participant. I also utilized the journaling process to reflect on data, examine any biases, and consider my positionality within the interview processes. The entries were often referenced when engaging in constant comparison between interviews and during the data analysis process. The records provided context regarding what I was thinking or feeling about code development during the analysis phase. Given that the interviews ended up being spread out over a much longer time frame than anticipated, due to rescheduling during the stay-at-home orders, they often served as a way to refresh my memory and assist in reentry into the data collection process.

Additionally, I made journal entries to guide peer debriefing sessions or notes on modifications to the interview process. During the analysis process, journal entries consisted of ideas about specific interviews, the coding process, and emerging themes and insights into theory development. The journaling process was essential in working through the emerging ideas about how different themes connected. It was also a place to note ideas I had for future research topics,
potential implications, or discussion content and served as a space to leave short notes or reminders regarding information to place into the audit trail document. This process assisted in increasing transparency within the research process and contributed to increased trustworthiness and objectivity within the study (Denzin & Lincoln, 1994).

**Audit Trail**

The development of an audit trail was an ongoing process that began at the protocol development phase for IRB submission through the data analysis phase. The audit trail served two primary purposes: tracking information about the data generation process and providing information regarding the data analysis. The grounded theory approach can be inherently messy, and the audit trail became an important way for me to organize and document the entire process. Additionally, it was an excellent resource when I needed to review a protocol decision or create dense descriptions to communicate the trustworthiness of the work. I initially created a memo entry into NVivo for each separate instance I needed to make notes for the audit trail. Realizing the number of times I would need to create entries, one memo document was created in NVivo. For each new entry in the document, I added the date along with key information or discussion that had taken place. The following is an excerpt of an audit trail entry regarding a protocol decision to change the inclusion and exclusion criteria.

3.4.20- I have run into an issue that the agencies with budgets under 250k are not meeting the inclusion criteria. The inclusion criteria are: must be a 501c3, that they have at least 4 programs, do not have 50% or more of their budget funded by one source, have funder-mandated metrics and then have a person in direct service and an administrator to interview. I have screened 10 agencies within that budget range, and none of them had funder-mandated metrics and then did not meet at least 1 other criteria. Additionally, the smaller budget sized nonprofits are conveying more significant challenges adjusting and sustaining services during COVID-19. Based on this, I felt like these barriers are likely to continue in trend and might not be a viable budget range option for the study. I spoke to Michael, Nathan, and Maria to consult on the issue. Based on the issues and given the
general ease of recruiting in the other budget ranges discussed, moving in the direction of expanding the sample size in the other 3 budget categories. The final decision was to move from interviewing 2 agencies (4 people) in each of the 4 budget category ranges to interviewing 3 agencies (6 people) in the top 3 budget categories. It increases my N from 16 to 18 and provides more data for each budget category. It would also increase the diversity of the different types of agencies within each budget range.

The following chapter provides information regarding the participants, nonprofit organization information, and the study’s findings.
CHAPTER FOUR

FINDINGS

This chapter presents the findings from the analysis of semi-structured interview data using a grounded theory approach developed by Charmaz (2014) to examine the following research question: What are the perceptions of administrators and staff members within nonprofit organizations regarding the influence of funder-mandated performance metrics on service provision? Findings arise from interviews conducted with 16 participants, in addition to information related to the agencies in which participants were employed. The findings identified three primary categories influenced by funder-mandated performance metrics on service provision that include (1) clients; (2) nonprofit programs; and (3) agencies. Each category includes themes and subthemes that emerged from the data (Table 5). The following provides a detailed explanation of the findings.

Table 5. Primary Categories, Themes, and Subthemes

<table>
<thead>
<tr>
<th>Category</th>
<th>Themes and influence upon:</th>
<th>Subthemes</th>
</tr>
</thead>
</table>
| Client  | 1. Definitions of client success  
|         | 2. Client-provider relationship  
|         | 3. Direct influence on clients  | 2a. Evaluation tools and metrics limitations  
|         | 2a. Program assessment  
|         | 3. Service modifications  | 3a. Funder relationships influence on service provision  |
| Program | 1. Time and staff capacity  
|         | 2. Program assessment  
|         | 3. Service modifications  | |
| Agency  | 1. Data learning  
|         | 2. Agency planning  
|         | 3. Agency financials and mission  
|         | 4. Metrics systems development and sustainability  | |
Participant Demographics

The study included a total of 16 participants who came from 8 human service nonprofit organizations, all located within one major metropolitan city in the Midwestern United States. Table 6 provides information regarding the participants’ roles and titles within the agency and the number of years they have been in their current positions. It is important to note that several participants had been employed with their agency for a more extended time than noted, but in a different role (e.g., former coordinator and now current director). Pseudonyms were assigned to each participant to assist the reader with basic information about each person while removing any identifying information.

Table 6. Participant Pseudonym, Job Title, and Years in Position

<table>
<thead>
<tr>
<th>Pseudonym</th>
<th>Job title</th>
<th>Years in position</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>Development director</td>
<td>4.5</td>
</tr>
<tr>
<td>Brittany</td>
<td>Program director</td>
<td>5</td>
</tr>
<tr>
<td>Clare</td>
<td>Senior director of programs</td>
<td>1</td>
</tr>
<tr>
<td>Dennis</td>
<td>Executive director</td>
<td>17</td>
</tr>
<tr>
<td>Drew</td>
<td>Director of research and evaluation</td>
<td>2</td>
</tr>
<tr>
<td>Emma</td>
<td>Assistant director of development</td>
<td>2</td>
</tr>
<tr>
<td>Heather</td>
<td>Program manager</td>
<td>3</td>
</tr>
<tr>
<td>Josh</td>
<td>Program director</td>
<td>7</td>
</tr>
<tr>
<td>Kayla</td>
<td>Director of programs</td>
<td>7.5</td>
</tr>
<tr>
<td>Kristy</td>
<td>Clinical services director</td>
<td>5</td>
</tr>
<tr>
<td>Melanie</td>
<td>Executive director</td>
<td>2.5</td>
</tr>
<tr>
<td>Molly</td>
<td>Development director</td>
<td>4.5</td>
</tr>
<tr>
<td>Norma</td>
<td>Executive director</td>
<td>32</td>
</tr>
<tr>
<td>Samantha</td>
<td>Program director</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 7 provides general demographic information about all participants, including gender, sexual orientation, age range, ethnicity, and their highest level of education.

Table 7. Participant Demographic Information \((N = 16)\)

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>12</td>
</tr>
<tr>
<td>Male</td>
<td>3</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>1</td>
</tr>
<tr>
<td><strong>Sexual orientation</strong></td>
<td></td>
</tr>
<tr>
<td>Bisexual</td>
<td>2</td>
</tr>
<tr>
<td>Heterosexual</td>
<td>11</td>
</tr>
<tr>
<td>Gay</td>
<td>1</td>
</tr>
<tr>
<td>Queer</td>
<td>2</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>20–30</td>
<td>1</td>
</tr>
<tr>
<td>31–40</td>
<td>9</td>
</tr>
<tr>
<td>41–50</td>
<td>4</td>
</tr>
<tr>
<td>51–60</td>
<td>1</td>
</tr>
<tr>
<td>61–70</td>
<td>1</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>2</td>
</tr>
<tr>
<td>White</td>
<td>13</td>
</tr>
<tr>
<td><strong>Highest education level</strong></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>1</td>
</tr>
<tr>
<td>Master’s</td>
<td>14</td>
</tr>
<tr>
<td>No response</td>
<td>1</td>
</tr>
</tbody>
</table>
Agency Descriptions

Eight nonprofit organizations were identified to interview the 16 participants for this study. All nonprofit agencies had a 501c3 status, an annual budget of at least 1 million dollars, and provided a minimum of four programs or services in the community. The following narrative describes each organization, summarizing their work and the clients they serve. In addition, Table 8 provides agency information, including budget size, number of programs or services, number of people annually served, number of years in operation, and the number of employees.

Table 8. Agency Demographics Information

<table>
<thead>
<tr>
<th>Agency name</th>
<th>Budget range</th>
<th>Number of programs</th>
<th>Number of people served annually</th>
<th>Number of years providing services</th>
<th>Number of employees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agency 1</td>
<td>1–4.99 million</td>
<td>5</td>
<td>10,000</td>
<td>57</td>
<td>22</td>
</tr>
<tr>
<td>Agency 2</td>
<td>1–4.99 million</td>
<td>5</td>
<td>3,600</td>
<td>25</td>
<td>15</td>
</tr>
<tr>
<td>Agency 3</td>
<td>1–4.99 million</td>
<td>6</td>
<td>6,000</td>
<td>45</td>
<td>21</td>
</tr>
<tr>
<td>Agency 4</td>
<td>1–4.99 million</td>
<td>4</td>
<td>2,000</td>
<td>46</td>
<td>17</td>
</tr>
<tr>
<td>Agency 5</td>
<td>5+ million</td>
<td>9</td>
<td>390</td>
<td>35</td>
<td>80</td>
</tr>
<tr>
<td>Agency 6</td>
<td>5+ million</td>
<td>15</td>
<td>18,000</td>
<td>150</td>
<td>180</td>
</tr>
<tr>
<td>Agency 7</td>
<td>5+ million</td>
<td>4</td>
<td>1,330</td>
<td>35</td>
<td>58</td>
</tr>
<tr>
<td>Agency 8</td>
<td>5+ million</td>
<td>5</td>
<td>900</td>
<td>21</td>
<td>40</td>
</tr>
</tbody>
</table>

Agency 1 is an advocacy and community organizing agency that provides services to residents within three diverse neighborhoods. The organization focuses on community-based
organizing for housing and immigration and facilitates programming to support youth organizing and leadership.

Agency 2 provides sports and leadership programs to school-aged girls all around the city. Partnering with schools across the area, the agency provides afterschool programming as well as one-day workshops. The organization focuses on providing an empowering experience to girls in a fun, team environment.

Agency 3 provides services across the city, focusing on supporting women experiencing homelessness. The agency provides a variety of housing services, in addition to comprehensive case management, health, and employment services.

Agency 4 provides health and mental health services, and health education for clients on the city’s northside and surrounding areas. The agency provides multiple services and focuses on creating a respectful environment for people to receive empowering care.

Agency 5 is a multi-issue agency providing services to youth between the ages of 10 to 17. Partnered with schools across the city, the agency provides services within school settings and offers opportunities to participate in afterschool and weekend services.

Agency 6 is a multi-service organization that provides comprehensive services to immigrant and low-income families across the city. The agency works with children, youth, and adults, providing many social services, resources, and referrals.

Agency 7 provides services across the city, focusing on supporting families experiencing homelessness or poverty. The agency provides permanent housing and wrap-around support services.
Agency 8 provides services to support individuals and families in building financial strength and addressing financial and racial inequity. The agency provides programs and workshops for children and parents to interrupt the generational transmission of poverty.

**Client Category and Themes**

Interviews revealed that funder-mandated performance metrics systems within nonprofit organizations directly influence the clients served by the organization. These themes focus on how a client could be influenced by implementing mandated metrics systems within the program or agency. The three themes that emerged from the data illustrated that funder-mandated metrics influence how client success might be defined (definitions of client success), as well as the relationship between the client and their direct service provider (client-provider relationship), and the clients’ experiences when engaging in care (direct influence upon clients). The following provides detailed findings emanating from the three themes.

**Definitions of Client Success**

Analysis of interviews revealed that funder-mandated evaluation metrics can substantially influence how success is defined among client populations, as further outlined under this first theme in the following section.

For nonprofit programs, annual goals and outcomes are determined to reflect the program services and the intended impact of the services upon clients. Various definitions of client success are created that vary across agency, program, and client population to determine if programs have met these goals. Participants discussed two primary ways in which definitions of client success were measured, which are via process evaluation and outcome evaluation.
Process and Outcome Evaluation Challenges. Agency employees discussed the challenges and struggles they faced when trying to measure a predefined process or outcomes through evaluation because the client process was more complex than the metric could capture, or the definition of success did not align with the client’s individual goals. Implementation of funder-mandated metrics shaped the types of data collected. It also informed how client success or progress was defined within the parameters of the evaluation system utilized to measure such progress. Providers felt that mandated metrics often missed the nuance of the services and many of the critical markers and steps before reaching a predetermined goal. When external funding sources provide their own evaluation tools or require specific metrics, the program’s definition of success can change based on the needs of those specific metrics. As Kristy explained, “[Funder name] creates the evaluation. Every year, it’s different as [funder name] priorities change. [Funder name] may say, you need to measure this thing, or this is a priority this year.”

Susan shared a similar example, stating, “I feel like if you look at that particular program, the larger overall [goal is] improving academic success.” However, Susan discussed that, in this program, attendance was identified as a marker of academic success. She questioned if this definition is actually measuring the improvement of their grades, asking, “Did the student increase their attendance? Did that then improve their grades?”

Industry Standards Usefulness. Additionally, for many of the participants at some nonprofit agencies, the mandated metrics supported a program’s efforts to become more aware of industry standards and metric markers to track and analyze client success. These industry standards provided an evidence-based definition of client success and corresponding mandated metrics to utilize in assessment and evaluation. The standards assisted agencies better in
determining definitions of client success within their programming and enabled the programs to compare their client outcomes to other agencies performing similar work. For example, Molly shared, “We tend to base a lot of our metrics on industry standards and looking at national research and our peers’ standards to see what good performance measurements are.”

**Industry Standards Concerns.** Other participants did raise concerns about the use of industry standards. They were more hesitant to apply the industry standards’ definition of client success when their clients’ needs, or circumstances were not reflected in the development of the industry standards. As April described,

So definitely, we do always talk about [industry standards] being in line with the national standards and the [name of school district] standards, and then we talk about [our standards] being built upon with more inclusive, more expansive, more comprehensive [definitions].

Similarly, Samantha talked about the concerns of industry standard definitions being applied to their clients who are unique to the industry defined populations:

I think our [office location 1] and [office location 2] serve a higher level of undocumented family members and undocumented parents. Their program director will say, “not everything on our [financial] scorecard shows the success of the program.” Like if someone’s starting their own business and they're getting paid under the table.

In this situation, the mandated metric definition of client success utilized does not account for cash economy income, nor does it account for barriers that people without citizenship status may face in achieving their financial goals.

**Challenges of Mandated Evaluation and Individual Client Goals.** Finally, analysis revealed that it could be challenging to create individualized client goals or treatment plans when the data collected had to address the predetermined mandated metrics. As a provider, Drew shared,
It might not be [a] feasible goal for some of our members to get to a family sustained wages within two years. Oftentimes, it requires education programs, which takes a lot of folks over the two-year time with learning English. If it’s an immigrant family, who’s just arrived in the U.S., you know, taking time to build a network and get settled. You know those are prerequisites to family sustained wages. If we can't capture that, we're not really telling the right story.

When a definition of client success was already detailed based on the corresponding evaluation, providers often grappled with how to continue to create personalized plans, knowing they had to report progress in a predefined way.

**Client-Provider Relationship**

Analysis of participant interviews also revealed that funder-mandated metrics could positively and negatively influence the client-provider relationship. In fact, the majority of participants discussed several different ways in which mandated metrics could directly and indirectly influence the client-provider relationship. Providers can play a substantial role in a client's progress while providing support in working toward attaining their goals. Mandated metrics can influence the rapport being built, the time the provider can spend with each client, and the overall care they receive.

**Positive Influence on Client-Provider Relationship.** These participants shared that data collected through funder-mandated metrics provided feedback on client satisfaction with programming and their experiences receiving services. This information assisted direct service providers in better understanding the needs of their clients, and in turn, providers then used that information to respond to meeting their clients’ needs directly. The data enabled providers to build stronger relationships with their clients and support their efforts to build trust and rapport. As April shared,
I know that the program directors do take those monthly reports pretty seriously, especially when it has to do with client experience. So, if someone is reporting a negative experience in the waiting room with phone staff, with the provider, even if it’s one, we’re taking that very seriously. And I think that one, it’s because we genuinely care that someone had a bad experience. . . . And if that is being impacted, that has to be addressed.

Similarly, Melanie reflected upon how client feedback is key to ensuring healthy relationships with providers:

You have these really well trained [employees] that are trained to deal with kids, maybe in challenging situations, reacting to [the kids] in trauma informed ways and all that stuff. That can be really powerful. If it’s not done right, it can be damaging.

**Negative Influence on Client-Provider Relationship.** Additionally, participants talked about how important the role of the relationship with the provider and client is when it comes to setting client goals or treatment plans. When mandated metrics shape the definitions of client success or how success will be measured, the provider can find themselves in the position of knowing the client may not be able to meet an outcome or marker of success. Providers talked about deciding not to meet an outcome and focusing on their client’s needs rather than programming or metric expectations. Providers were faced with two pathways for clients facing barriers to program success or among those unable to make progress and strides within an identified time frame. One path would include providing services knowing the client would not meet the goals, ultimately impacting programming metrics and reporting, and the second pathway would be to admit that they would not be able to provide services to meet the metrics criteria. As Kristy explained,

We have a goal of zero evictions here, and we just really strive for that, and if you’re not evicting folks, then you are making a commitment to then have folks that might not meet those outcome thresholds.
Data from the participants suggested that mandated metrics influenced how services could be provided and created instances where services were inflexible. The inflexibility of services could limit the support or resources from a provider and influence the relationship with their clients. Sharon talked about how the mandated metrics limit the providers’ ability to provide additional care or services, stating, “[Mandated metrics] made [the services] a lot more regimented and programmed and maybe not as flexible and creative in some of the types of things that we could do.”

Mandated metrics evaluation processes could also disrupt service provision. Several providers discussed how implementing evaluation processes took time away from providing direct care, was taxing upon clients, and reflected upon how the client viewed the service provider. As one participant noted, “It’s cool, and we understand, but sometimes [evaluation] gets in the way of doing the work that we want to do.”

**Direct Influence Upon Clients**

In addition, data analysis revealed the third theme, focusing on how mandated metrics can directly influence clients. Participants discussed the positive and negative influences of mandated metrics upon their client populations.

**Motivating Client Experiences.** Mandated metrics can be used to incentivize client engagement or set parameters for client participation in services. As Emma stated,

I think that metrics like attendance requirements can have a positive impact. If it’s youth in an out-of-school time program might want to blow it off and not come. But then, they end up having a really enriching experience.

Mandated metrics may require attendance tracking as a condition of service, maintaining eligibility for services, or encouraging clients to continue utilizing services. As Clare explained,
We have those strict kind of attendance requirements. For those participants, I think [it has] a positive impact, because you know if you’re in therapy because of personal experience, like, I know there can be a lot of fear and anxiety around that and even just showing up can be a barrier.

**Punitive Client Experiences.** However, some participants explained that mandated metrics do not always account for barriers that some of their clients may experience and could result in a punitive experience. Clare also provided an example of this via a client experience, explaining,

> [In this program], you must continue [during the COVID-19 pandemic]. You have to do this virtually, e-learning, and figure it out. But [clients] don’t have computers or don’t have internet. They may have just lost their job. They may be facing, you know, maybe they're not going to pay their rent, figuring out how to get food on the table. [The funder is] still, “Well, they need to figure it out.” I feel like sometimes it’s also thinking about what are those other things that people, or folks are working through their challenges or barriers that really prevent them from being successful in the program.

Mandated metrics evaluation systems can be taxing upon clients in several ways. When there are many metrics requirements in a program, participants could be required to complete excessive tasks for activities to meet the expectations tied to the mandated metrics for data collection and reporting. At other times it is taxing upon a client when a large amount of data must be collected from them, such as with continued progress assessments. Additionally, participants discussed that some of the data being collected as a requirement of the metrics could create negative or difficult emotions for clients, as the questions posed, or information gathered might feel intrusive or not relevant to the services being received. As Melanie discussed, “[Data collection] also can be really embarrassing. It just brings up so many more inequities than you would think.”

Lastly, participants shared that mandated metrics could influence how services are provided directly to the client, limiting clients in making progress toward goals or even
maintaining the status quo. Here, Heather discusses how mandated metrics require client participation in specific services as a condition of program enrollment, “Then they are making you do everything, so you’re not learning nothing new, just basic menial tasks year-round, so there is no room for growth or development or where you want to go from there.” Kristy shared a similar perspective, “There isn’t a lot of work around helping—I think a lot of homeless services are really about reactive surviving from a systems perspective.”

**Nonprofit Program Category and Themes**

Findings associated with data analysis revealed that funder-mandated performance metrics implemented within nonprofit organizations influence agency programming, services, and care. These program themes focus on how services could be influenced due to mandated metrics systems being embedded into the delivery of services for clients. The analysis identified four themes, which focus on how direct service providers’ time and capacity to do their job is influenced (*time and staff capacity*), how the process of program assessment influences how services are provided (*program assessment*), how mandated metrics systems create limitations in assessing programs (*evaluation tools and metrics limitations*), and how the implementation of mandated metrics influences how services can be provided or modified (*service modifications*). The following sections and narrative provide a detailed description of each of the four themes.

**Time and Staff Capacity**

Data analysis revealed that participants discussed how program staff time could be influenced by working with mandated metrics and evaluation systems. Participants discussed the additional work evaluation added to their regular jobs, concerns regarding staff capacity, and positive influence dedicated evaluation personnel had on staff support.
Negative Feelings Towards an Increased Workload. More than half of the participant responses were negative, discussing how much additional (often uncompensated) work was added to their primary job descriptions. Some shared that the amount of time allocated to mandated evaluation processes limited the time providers could spend with clients and equally limited the time allocated to provide direct programming. As Heather shared, “We have almost 500 clients, and it’s three of us, so we cannot work with everyone, and we cannot gather all that data.” Concerns were also raised related to how the time allocated for mandated metrics systems could influence the quality of services. As Sharon explained,

So [the staff are] inputting the attendance in the 15-minute increments. They have to collect all the grades. They have to collect discipline referrals. They have to collect attendance. So, they have all of these things that they have to collect, and then they also have to connect to how much time the person is spending in their program and what types of programs they’re spending time in. . . . And there’s so much time that they have to spend on infrastructure on all these different areas.

Concerns Regarding Staff Capacity. There was additional discussion about the influence mandated metrics had upon the capacity of program staff to incorporate mandated metrics evaluation processes into their jobs. For agencies that did not have dedicated personnel overseeing the evaluation system and processes, participants discussed that direct service providers often did not have the necessary skills or were not provided with the appropriate training to track, collect, and process mandated metrics data. The evaluation process was described as being simply tacked onto their existing job description and expectations without support or opportunities to learn more. As Nancy stated, “I think some of it is that [direct service providers] might not have the capacity in terms of the skill sets.” Similarly, Melanie explained that,
It’s a lot to stay on top of, but you also have to because you don’t want to be satisfied that we only have ten [clients] and only half of them ever come. Well, something is wrong, but I don’t know that staff, and this is probably a reflection on us, that we need to better educate them. But I don’t know that they immediately make the connection of well, if our numbers are low, then there is a reason why they’re low and we should fix that reason.

**Positive Staff Support with Dedicated Personnel.** However, agencies with a dedicated staff person or team did appear to better manage the process around mandated metrics in this study sample. For example, several participants expressed that dedicated personnel resulted in more support for direct providers and increased agency capacity to manage metric systems. As Kristy shared, “I just think that agencies can have a dedicated position. I’m not someone that is just pulling data, but [I am] really being proactive and looking into things that can help move the agency’s data quality forward.”

**Program Assessment**

Additionally, a majority of the participants discussed several ways that mandated metrics influenced program level assessment. Funder-mandated metrics provided data to monitor program progress and quality, assess program and mission alignment, convey transparency and accountability, and identify any trends, population changes, or barriers to accessing services.

**Data to Monitor Programming.** First, utilizing mandated metrics through evaluation and assessment processes assisted programs in monitoring program progress, focusing on how the program is working overall, and providing information regarding the progress of individual clients. As Drew shared, “If the [direct service provider] hears that [information] and is able to look at the dashboard that shows the [client’s] previous answer to the question, that can inform the way the [direct service provider] responds.”
Data from the mandated metrics evaluation systems ensured the program accurately captured what was intended to be assessed and measured. As Sydney stated, “I think it’s also just consistently important to hear like, is what we are doing, what we think we are doing? Is [the program] actually working? And is what we think we’re conveying actually getting through?”

Mandated metrics also provided vital information to assess the program’s quality and report the number of services offered. As April shared,

I know that not everyone does [this type of data analysis], but that’s the approach we take. And the primary thing I see is a pressure towards growth orientation. So okay, we say this many young people in our [program] last year, and the expectation is we will see more next year.

**Program and Mission Alignment.** Assessment data provided additional key information to assess the level at which services aligned with the goals of the program and agency mission.

Molly discussed how the program evaluation could support the program in ensuring they are meeting program goals and doing work that aligns with the overall agency mission,

I think [evaluation data] are things that we had to look at in how they connect to our mission and [are] driving us towards our program goal of how many [clients] we serve. And does serving these [clients] necessarily mean that we are housing them? Or does it mean that we are supporting them in securing and maintaining housing?

**Conveying Transparency and Accountability.** The culmination of this information also ensured that programs and services could utilize the data to demonstrate program transparency and accountability and convey the impact of the programming. April continued to explain,

The providers obviously would be using that [program data] consistently, and to be able to represent, okay, some progress means that this many folks are having this type of success. But also, we started breaking it out because we wanted to see, are we actually better at helping people with smoking cessation than we are with nutrition? Are we better at supporting people in some health goals than others?
Identifying Trends or Changes. Additionally, the mandated metrics and the corresponding data were useful to identify trends or themes arising within the programs. Emma explained,

So, all of that data is compiled and averaged, and I think it’s a really effective way to check in about programming and just make sure that everything is working out, you know, and if they’re—it’s an easy way to pinpoint any problems or new things.

Samantha shared similar thoughts, “I think sometimes it's figuring out those- being able to spot those trends and those nuances and how to then adapt.” Additionally, the program assessment data aided in identifying any needs or anticipated changes to services. For example, as Josh explained,

So, part of changing the program is—I mean, that’s part of the ethos of the program itself. It’s for [clients] to have high investments, and part of having high investment is them having some level of decision making. . . . A few times a year, we’ll take an hour or two and talk to them and say, “Hey, over the next three to four months, what are some things, as a group, you all want to accomplish in each of these?”

In some instances, mandated metrics data provided information that assisted programs in identifying shifts within and across their target population(s). Kayla shared, “[The data] made me think a lot, you know, in what we change. One of the big things that we changed is who we serve.”

Identifying Barriers to Access Services. Finally, this assessment information aided programs in identifying any barriers that clients may be facing in accessing services and completing programming. As Heather explained,

So, we do an assessment . . . and the barrier assessment is part of that. We were trying to figure out—we were trying to understand how much time it takes us from the day somebody is referred to us in the [name of] program to the day that we kind of deem the [client]-ready, what are those steps. And then, we also wanted to see in order for us to achieve those numbers, how many people should we be working with, and we need to
understand the population, lots of our folks have barriers. So, kind of not categorizing them, but trying to see what are some of the obstacles that people face.

There were a few participant comments related to the limitations of program assessment as well. Participants expressed that mandated metrics data did not always assist the program in learning more about their programs, yet instead resulted in solely creating data to meet the requests or needs of their funders. Therefore, funder-mandated metric data was required but provided minimal to no useful programming assessment information. As Sharon shared,

I think that our other funding sources and the metrics are not really as burdensome . . . they’re fine with us collecting fairly minimal data, but they just want to see one [key change such as], what are the contacts we’re making? . . . If you want to measure us based on that, we are a smashing success. But we know that’s not really what we need to work on.

Some participants discussed limitations regarding the depth of the information mandated metric program assessments could provide. Often, there was little to no opportunity to provide any contextual background in addition to quantitative program data. One of the participants expressed a potential explanation for this, conveying that program reports generally prioritize quantitative data over qualitative. As Kayla explained,

In general, it’s more numbers that [the program is] collecting for outputs. And it seems like, you know, on the one hand, it’s a way for funders to be able to sort of compare one agency to another. But . . . then there are all these people who are striving and making successes, but they're not counted, and that’s hard too.

Emma also shared that she wished for more flexibility in reporting and wanted an opportunity to showcase more of what the program was doing, as she explained,

I would ask for more flexibility in what we’re reporting. Or, I guess, more flexibility if we’re feeling a deficit in what we’re reporting. [There are] different ways that we could show that we're doing the work, even if we’re not always meeting the metric.
These participants explained that qualitative program data was often considered supplemental to submission of quantitative mandated metrics. Without including some of this key contextual qualitative information, interview participants believed that many aspects of the program were not clearly or fully assessed. They raised concerns that such limited program data did not robustly capture a more complete picture of the service impact upon their clients.

**Evaluation Tools and Metrics Limitations**

In addition to the influence of mandated metrics on program assessment processes, data analysis identified a first subtheme focusing on the concerns and limitations of tools used in program evaluation. Participants shared concerns using funder-identified evaluation tools to collect mandated metrics information, explicitly noting that the tools could be problematic and limiting when considering the programs’ population.

One of the concerns raised in more than half of the interviews discussed the limitations in using standardized or other evidence-based evaluation tools for program assessment. These limitations included a lack of diversity among the populations utilized to develop the actual assessment methods and/or that the assessment tools were not created by including a population similar to the programs’ client demographics. Additionally, these mandated metric tools may not account for the cultural nuances or other unique variables that the programs and services were addressing. The assessment tools could be a poor fit to accurately capture the influence of programming upon the client population, or the metrics could misrepresent the influence of programming based on what data was being collected. As Melanie explained,

I think another thing we’ve come up against with the surveys . . . is they’re these validated scales, they’re not always validated for kids growing up in maybe urban underserved areas, and so they’ll be questions—like there’s one question that I know
we’ve asked about violence and it’s well, sometimes, I only have two choices like hit someone or have them hit me… They’re living in neighborhoods that struggle with violence. You don’t know what messages they’re getting outside of the program. Is it wrong that someone might teach them to defend themselves because they might need that? I think that’s been a struggle with some of the validated scales that you want to be able to say you’re using, but they’re not always appropriate.

Similarly, almost all participants shared the concern that numerous variables could not be accounted for in the mandated metric tools provided. Participants talked about different types of variables that could impact a client’s ability to engage in services or that failed to reflect actual progress that could not be captured by the required measurement tool. This missing information could negatively impact data reports and fail to identify potential barriers or changes necessary to the program’s success. As Brittany explained,

I mean, the other part of [evaluation] is less evaluation metrics and more numbers which sometimes, we don’t have great numbers at a [program], and there could be a million reasons for that. But having to then explain to a funder this is what we thought we were going to serve [in numbers], and this is what we actually did.

Josh also talked about the numerous variables they are not able to measure that may influence a clients’ outcome,

I think there’s dozens of factors that affect the [client’s] ability to succeed or not succeed in the academic year. And we’re just one small piece of that. So, there’s obviously home life and the teacher and the fact that they’ve never been in a middle school before, so all these other factors. So, for us to say, “Ooh, a kid was in our program. And their grade went from a B to an A,” or vice versa, “a kid was in our program, and their grade declined.” There are so many other factors.

In addition, participants shared concerns about the accuracy of the collected mandated metrics data, as Clare stated,

[The parents/guardians] are going to be like, yeah. We got our log. The cute thing is the kids will hold them accountable, actually, because they’re super cute. They’re like, “No,
we didn’t read.” Their mom is like, “Yeah, we did. We get it.” [Then their kid says] “No, we didn’t.” It’s also the integrity of the data.

April likewise shared concerns about accurate responses from clients,

> With young people, giving them a form to fill out may not be the setting or the tool through which they're going to actually express some of their central thoughts, concerns, questions, right, even if it’s anonymous.

These concerns and others were raised about the potential issue of response bias when clients had to engage in frequent assessment or if clients responded based on what they believe they should say concerning their progress within a program.

*Service Modifications*

Additionally, participant interviews identified ways in which mandated metrics could influence how program services are delivered or modified. For example, programs might be modified based on the results and data provided from the mandated metrics. When programs utilized the metric data successfully, resultant information provided insight into any needs, barriers, or changes necessary to improve upon programs and services. As Drew stated,

> We knew a little bit about how [the program] works when we went into [the data], and that is a lot easier to scale—scale it up rapidly. . . . Where we can learn what works and doesn't work and adapt from there, which is a big asset for an organization like ours.

Likewise, Kayla spoke about how the mandated metrics information informs how they provide services and adjust based on what they learn, stating, “You know [what] we are thinking about is our intervention working right now based on the outcomes we're getting or is there something that we need to shift and then evaluating what direction [we] should go with that.”

Programs could also use the information to shift how services are provided and what approaches to care are utilized. As Heather shared, “I think over time, even understanding how
we serve the clients and really being able to integrate the different models like harm reduction and trauma-informed care, I think [the data has] informed a lot.”

Alternatively, programs could be modified as a result of external funders requiring specific external goals or outcomes that were more in line with the funder’s priorities or mission than those of the organization. It often meant that “when you adapt your metrics for funders, it gets messy” (Samantha). In some instances, the program changes happened to ensure that services being provided would produce data aligned with the mandated metrics. As Sharon explained,

[The funder will] say, you know, you need to provide some more time for tutoring or add some little time of homework help in addition to enrichment activities, I believe. Then we figure out how to provide those services and what that looks like.

**Funder Relationship’s Influence on Service Provision**

A second subtheme emerged in the findings as a result of the data analysis. Several participant interviews underscored how important their relationship was with funders when making any changes with programming or services. The more positive a funder relationship, the easier it was to support program growth or changes that supported the agency and funder. However, negative relationships could result in a power imbalance that pushed nonprofit staff to appease the funder over the agency’s needs.

**Positive Funder Relationships.** Participants who felt there was a positive funder relationship expressed a more robust understanding of the services provided, a clearer sense of the reality of how the programs work and impact clients. This positive funder/provider relationship supported program efforts to make changes to services that aligned with both the program and funder needs. As Melanie stated,
We have good relationships, so we can kind of negotiate, and some of our biggest long-
time funders like [funder name], they funded us, and they’re great. You can have this 
conversation with them, and they’re going to get it, and they’re going to come back and 
say, okay, our focus here—these are the things that we want you to look at and then 
you’re like okay, that’s fine.

Additionally, a positive relationship with a funder could influence the funder in shifting their 
perspectives regarding how program services or care are best provided to specific populations.

Finally, a positive relationship could influence how a funder viewed a program’s expanding 
services or new approach to addressing a social issue through service provision. The following 
response from Sydney highlights an experience discussing changes with their funder.

We’re [the funder] trying to get our board of this foundation on board with that. So, in the 
meantime, know that we’re supporting your work and that we want to talk more about the 
ways in which this foundation could be supportive in a more overtly queer inclusive way 
than we currently are.

**Negative Funder Relationships.** However, when participants described an 
underdeveloped or strained relationship with a funder, it was more challenging to engage in 
conversations around the influence of mandated metrics on programs. Participants shared that 
funders with less of an understanding of the program pushed compliance with the funder’s 
expectations to ensure services would align with the mandated metrics required as a condition of 
continued funding support. As Josh stated, “That’s a tough balance, but I mean, there wasn’t 
even an opportunity to have a discussion. I’m not about that at all.” Emma, likewise, elaborated 
on the influence of the relationships,

I think that a lot of those program officers report to very out of touch boards that are like 
why are you still proposing this number of plans when you reach this many clients like 
you surpass your goals, that means you should raise your goal. And then, it’s like, but 
you’re not raising the grant [funds], you know?
The relationship with funders, specifically program officers, had a strong influence on how the agency was able to communicate with the funder and navigate areas of misalignment between the funder and program.

**Agency Category and Themes**

Analysis of the interviews revealed that mandated metrics systems within nonprofit organizations could influence how the agency as a whole functions and makes decisions. The four themes in this section examines how mandated metrics influence how the agency learns how to utilize and interpret mandated metric data (*data learning*), discusses how data can be utilized to guide agency decision making (*agency planning*), influences agency financial planning and growth (*agency financials*), and shows the influence of mandated metrics on an agency approach to evaluation system development and sustainability efforts (*metric systems development & sustainability*). The following provides a detailed description of the four themes in the agency category.

**Data Learning**

Analysis of the interviews revealed that participants discussed how mandated metrics, data, and information influence agency operations and decision making. Josh discussed how it’s a combination of looking at the data and then having [staff] reflection about the benefits of the programs. . . . We’re trying to relatively frequently be doing evaluation and then tweaking, and then also setting aside time to annually sit down and think about bigger picture stuff.

Kayla discussed how the data is used to make decisions about services, while sharing,

We evaluate and change trajectory. I think [data collection analysis] is more ongoing, and so, any of the goals that are involving any of those direct service program[s], I am a part of those [administrative] conversations, and then, I’m going back with my [program] teams in figuring out how to implement any changes.
Agencies typically tracked data from an array of required tools to meet the requirements and needs of multiple funders. This resulted in a large quantity of various sources of information being required for a variety of funders. Often, in addition to any funder required information, the agency was collecting various types of data for their internal use. For many of these agencies, direct service providers and administrators had to learn how to collect, analyze, and interpret the data for funders and internal users alike. Samantha explained an instance where they had to learn what the data was telling them,

A really common aspect we’ll get is for funders to ask for the amount of money that [clients] have been able to save. So, they’ll want [the agency] to say something like, on average, [clients] have increased their savings by $500. And that sounds like a commonsense way to measure, measure impact in savings. But it’s actually not, we’ve learned over the past years. That’s kind of a misleading indicator to most anyone who’s building savings. It’s usually meant to be spent on something. It goes up and down over time. It’s not something that just increases yearly. We have a lot of [clients] who, you know, set a goal to have money . . . so that when they achieve the goal, their saving balance actually goes down. So, when a funder asks us for this [information], the average change in saving was negative for our [clients]. But, when we looked closer, a lot of those [clients] had actually achieved major goals.

**Barriers to Data Learning.** However, many participants talked about a lack of personnel or resources to collect, analyze, or utilize all available data. Many of these participants talked about data excess, which resulted in having more data than the agency could actually utilize. Some agencies had to prioritize the required data for funders, and the rest were often left unused. As Sharon explained, “there are so many different things you could input through, but there’s certain key things that we have to input [for funders].” Dennis explained that “we did this survey for 50 [clients]. That’s great. . . . Then it just sits somewhere and you’re like, ‘oh yeah. We did a survey. We should probably look into that.’”
Agency Planning

In addition, data analysis revealed that participants discussed the influence mandated metrics could have at the agency level, specifically related to agency-wide planning and decision making. Many participants also talked about the influence mandated metrics could have upon the strategic planning process. Mandated metrics contribute to the information necessary to assess progress markers in current strategic plans and set annual goals and objectives for future plans. Molly discussed how the continued data collection has assisted in a more robust planning process, when stating,

I really feel like when we do the next strategic plan, we’re going to come at it with an even stronger theory of change, which we’re working on. And we’re going to be able to set more precise goals. Not that they would be less ambitious, but that I think we have more data resources now than we did the last time and can really be even more specific.

A few participants shared that mandated metrics information created a better understanding of the agency’s long-term impact over several years. Norma spoke about how it assists funders in looking at long-term growth and the impact upon clients when stating,

[The funders] don’t expect you to change the world in a year. I mean, their guidelines make it sound that way. But of course, we know, and they know—that reality sets in. And that’s why you also have goals around the number of people involved.

The information produced from the mandated metrics could support planning at the staff and program level as well as the board level. The mandated metrics data and any internally tracked agency data could be used to assess current outcomes and inform next steps. As Brittany explained,

In our strategic planning process, um, usually, we have a two-level process. One where we set some of our goals with our strategic plan with the board, and then we have goals that are developed through our staff. I help lead that full effort.
This level of data collection and planning required that the agency confirm that all funder-mandated metrics were fully implemented throughout the agency programs. This required that mandated tools were successfully implemented, and tracking, analysis, and reporting were ongoing. Often, multiple funders had different requirements, software platforms, data entry requirements, and due dates, which required continued agency planning and assessment through a fiscal year to ensure grant compliance. April shared what the agency process looked like for her agency,

We have to either revise or build the tools. We still use Excel spreadsheets for putting that data in and testing that. And then at that point, I'm reviewing typically quarterly what those—you know, if everything’s looking okay, at least quarterly, sometimes more. And the development and volunteer manager is really overseeing [if] surveys [are] getting entered, are they getting entered correctly across all programs, including the new clinical services ones. . . . I would say there's also a lot [of work]. At least twice a year, I have to have conversations about are we getting enough surveys? Are people giving out surveys? Surveys are important. That's always also kind of in the back of my mind. And then at mid-year and you’re getting them ready for reports or those weird in-between or just reports to pull numbers for.

Of the participants who discussed strategic planning, a number of them discussed how the data produced from mandated metrics could assist in thinking through more extensive agency changes. The information could assist in identify areas of growth, either within a program or in expanding services. Dennis explained what this process looked like at their agency, when sharing,

Yeah, to start out, [data analysis] was primarily to assess what’s the prevalence and what kinds of issues are [clients] dealing with? The second phase of that is, depending on the results of that in year one, how much do we have to build out our [services], more staff for that to specifically meet those needs? Because our current staff couldn’t possibly take on another whatever case load of [clients]. So that was kind of a phased-in approach. Depending on the results of those assessments, the goal would be, okay, yeah, let’s find
the money to provide more [staff] that are in those [programs], working directly with those [clients].

Additionally, the mandated metrics information could provide information in making difficult decisions, such as closing a program or cutting a service. Emma shared their experience, explaining,

The [programs or services] that are not generating revenue, or, mission-aligned that was in [the data]. So, we’re more aligned to the [agency] mission. And, as [the administrators] were [examining] all the data, we eliminated two of our programs because we’re like, this isn’t really what we’re doing as an agency.

The mandated metric data were helpful in decision making and also utilized to explain to funders why the program was not sustainable or not in alignment with the agency’s mission.

**Agency Financials and Mission**

Participant interviews also identified a theme focusing on the influence of mandated metrics on agency financial planning, decision-making, and growth. Participants talked about how mandated metrics can influence an agency’s fiscal circumstances. The ability to successfully utilize or implement mandated metrics systems influenced an agency’s ability to increase or decrease revenue and often contributed to escalated financial costs to support any agency evaluation, development, and sustainability efforts.

**Tension Between Money and Mission.** Many of the study participants discussed the tension between the need for financial support and the potential to compromise an agency’s mission as a result. Agencies often had to determine what was best for them and their client services. Many talked about discussions with administrators related to the willingness to let go of funding opportunities in order to stay true to their mission. Melanie discussed how it is a “fine line between you don’t want to change who you are and what your mission is, but you need the
money.” Emma shared the tension between adjusting services as a result of metric requirements and the agency mission, when stating,

Does [the funder mandate metrics] push you to provide your services differently, too, right?…what’s the trade-off, you know, if you need money coming in versus that type of thing, it gets—like, that’s always the dance, right?

Other participants talked about the times when the agency sought funding to increase revenue that was not the best mission fit. This resulted in adopting mandated metrics that did not always align with the agency or program’s work. Molly spoke about the situations that agencies might face, “Organizations were just so desperate for money. The mentality—and this is something I work against all the time. It’s something I am really passionate about. It’s getting away from this mentality of scarcity.” For others, the funding versus mission tension was brought about when securing new funds, as Samantha shared,

We used to get money from [funder name] to do digital literacy, and it’s like, we don’t [provide] digital literacy. Why are we doing this? . . . We’re not a good fit. It’s not a good use of our time or someone else’s time. . . . We don’t have the capacity; we are a very small team right now.

Many participants felt that their agencies had made progress to ensure that the agency mission prioritized funding and the corresponding mandated metrics. Kristy shared,

I think your agency also has to be brave enough to say no to opportunities that will take you away from your mission or take you away from—because I think in this competitive environment, we’re all chasing dollars, but we can’t chase dollars just to chase dollars and I think that’s where a lot of agencies got stuck too.

Likewise, Molly spoke about her agency’s effort to prioritize mission over funder priorities, sharing, “We really try not to apply for funding to create new programs that we normally wouldn’t do, to change our existing programs and metrics to meet a specific funder’s requirements.”
In other instances, agencies could lose income or were unable to increase their revenue due to their limited evaluation capabilities. Sharon discussed a situation that their agency faced,

“You got to be careful with that too because there’s one point over the summer, last summer, the [program officer] in charge of the [grant name], she’s like, “Sharon, you guys haven’t input much [data] into the [required data system]. I think maybe, this is just not the right fit for you guys.” I’m like, I don’t want to lose the funding.

Molly shared an experience that limited the agency to apply for funding,

“We don’t have those kinds of [funding] opportunities right now. We’ve really looked into what would [funders] consider, this kind of aspirational change or impactful change for a just society. While we are obviously making an impact on society, we don’t currently have a way to partner with other organizations or to measure systematic change.

**Increased Agency Expenses.** Participants also discussed the increase in agency expenses to support the work and implementation of evaluation systems. There was often a struggle to increase funding to invest in software, databases, or evaluation technology for some agencies. Others struggled to increase funds to hire employees or consultants to support their evaluation needs. Kayla and Heather shared their experiences of funding implications regarding development personnel. Kayla explained that the agency struggled to secure funding to support a staff member to manage their systems with many direct service providers on staff, when noting,

“You know funding is one big thing. We can fund some database or some parts of the database with some of our [grant] contracts, but not all of it. And you know something that [an administrator] talked about is it’s very different to be an administrator of a system like this for ten users, but if we’re talking about having 40 or 50 users, it’s a completely different ballgame.

Heather explained how complex the funding situation was to finance a person’s position to support evaluation work when stating,

“Our [development] salary comes from those contracts, and so, based on what amount of money we have been allocated [from grant funding]. . . . Based on the money that is
available, how much is available and from what contact. Each one of us have—I [am funded] somewhere from 10 plus contracts.

The majority of respondents shared that their funders rarely allocate funding to support investment in evaluation systems, nor do they consider such funding requests to support agency staffing efforts. Norma explained that funders “may allow you to charge a little bit to the grant, but not to really create what you need.” While funders have become more aware of the resources it takes to manage evaluation systems tied to funder-mandated metrics, there is little investment in supporting those agency efforts. As Brittany explained, “I think funders have gotten better about having broader [evaluation options], like you can name your own evaluation [process], but they do require a certain amount of it, but none of them want to fund it.”

**Metric Systems Development and Sustainability**

In addition to agency planning and financial influence, mandated metrics could influence how agencies approach evaluation system development and sustainability efforts. All the participants discussed the ways mandated metrics could influence how agencies develop assessment tools, modify, and improve assessment tools, and implement efforts to standardize evaluation systems within programs or across the agency.

**Evaluation Development.** Mandated metrics often required that agencies create their own evaluation and tracking systems. While the funder required specific information to be collected and reported, participants stated that the funder often did not provide assessment tools. The tool development and evaluation processes then had to be implemented within specific programs or across the agency. April shared what this type of evaluation systems development and improvement looks like; when considering the mandated metrics requirements,
[Staff and administrators] have talked about in every grant for the last few years that we are going to advance our evaluation measures by having these multiple perspectives for funders. I think in the [program name], that has also gotten more nuanced where we got feedback from a funder because we do pre and post testing and everyone does an evaluation. The feedback from the funder was, “hey, you’re seeing people fourth grade through college age, you’re seeing adult allies, you have workshops.” To what extent are these segmented, age-appropriate language? So, I think now we have 12 different evaluations for the [program name] and four or five different pre and posttests because its age range gets information on these topics, and this is language appropriate for them. That was again not a case of you need to do this, or we’re not going to fund you now. But it was like, this is a weakness we’re seeing in your evaluation, and we were able to put it in the grant for the next year and say, we heard you, we made changes, and they were glad, and they gave us money.

Heather also discussed how the agency had to develop metric tools to capture the required funder metrics in a specific way,

I don’t think that [the funder] provides us with tools, but they do have data—certain data that they want us to collect. . . . We have X number of people who should maintain that employment for 30 days, 60 days, 90 days, and a year. So, we have to keep track of that data.

**Difficulties in Evaluation Development.** One of the struggles agencies faced was determining how many systems needed to be developed for internal use and utilizing funder-provided database systems. Some of the participants discussed that, while the agencies developed evaluation systems to meet mandated metrics and agency data needs, they often had to resubmit some of the data into a second external system. Sharon stated that “So yeah, we have about—I think we have about 5, 6, 7, different systems that were inputting information into [from our tracking systems].” The agencies often were unable to standardize use of any one system, rather utilizing multiple systems in complicated efforts to gather information.

Likewise, Molly shared an experience of submitting data into external tracking systems from the agency evaluation systems when noting,
I think a lot of our metrics are things that we can pull from [our] system. I think the challenge that we have now is that our [program data], most of that, does not get entered into [the external database] because those programs are not managed or funded by the same sources. So, our Program Manager is really trying to create another system and has to track data and ask case managers for those programs to input information into a different system and then has to pull from that and synthesize it into the rest of the data. So, there is a little bit of work in compiling all of that.

Once evaluation systems have been developed, all participants discussed the work at the agency level to maintain and sustain the evaluation systems. Analysis of the interviews revealed that agencies had to continually modify evaluation systems based on funder priorities. In some instances, when a funder priority shifted or a new funder was added, agencies had to work to change current systems to meet these requirements. For example, as Sydney shared,

If a specific funder comes up in the middle of our year, we might have to create additional tools. So, there was a time in which we were doing peer education, and a funder was really excited for us to do it. We needed to create pre and post-tests for education and evals for peer education, which we weren't necessarily doing when we first sort of piloted the program.

Additionally, participants noted that their agencies had to modify and sustain internal assessments and the tools developed to capture funder-mandated metrics. For some agencies, the additional metrics systems were set up to learn more about a program or service before providing public information. Other agencies developed additional metrics and assessment tools because the mandated metrics did not capture certain aspects of the services essential to the agency’s mission and work. April explained a situation in which data was collected for internal agency use only,

So that’s not data that we share externally to funders. We’re focused just on the [program] goals portion. And now we’ve revised that tool to try to capture multiple [program goals]. So now that’s done, the next iteration is trying to capture multiple [program] goals that may overlap or may not. So that’s how we developed the tool.
Challenges with Evaluation Sustainability. Most of the agency participants also discussed various struggles they faced associated with metric system sustainability. For some of the agencies, there was a struggle to secure funding to support evaluation and assessment efforts. The funding shortage might be related to an inability to hire people to support the work, or a lack of income to purchase evaluation systems to support the program and agency needs. Clare shared an experience at their agency:

Most recently, we’ve invested a lot of resources into building out our development department. . . . We switched from ETO to Salesforce. And that is really done in partnership with [the development employees], to think through what are we doing? What data are we collecting? What data do we need from [evaluation]? What reports do we need? And then we build that out in our system.

Likewise, Josh spoke about the agency’s internal metrics sustainability efforts, “No. [Some metrics are] not a requirement of a funder. We used to have our own internal database through a Microsoft program, but it was real bare bones. We got the job done, barely.” Norma spoke more to the financial struggles to sustain required evaluation systems:

It was still that mostly government money that was really limited in the amount of admin that you could charge. That was really the problem more than anything. It wasn’t the performance metrics [systems]. It wasn’t the systems we had to use. It was the fact that they would not let us charge more than 5% [to support administrative costs], which was ridiculous, not sustainable.

For other agencies, there had been efforts from funders to support mandated metric tool development that ended unsuccessfully. For some agencies, the systems utilized were not a good fit for agency needs, while for others, it was fiscally unsustainable. Sharon noted, “When the funding ended, it had not become—like, it hadn’t embedded itself deeply in the organization. People weren’t going to use it anyway, if we didn't have funding for a part-time dedicated person to do the data entry.” Melanie shared a similar experience of the flawed system fit for their
agency:

[Funder name] is a big partner of ours. [Funder name] can sometimes come to us with really specific totally unrealistic [systems]. But they’re great about it when we go back, and it’s like, yeah, that is not going to happen.

In addition, participants talked about the efforts to sustain mandated metrics systems when the funder required tools or data requirements did not align with an agency mission.

Sydney provided an example of this situation:

Wow, this [funder-mandated] measurement feels really challenging to us, almost like it is reinforcing some harmful assumptions about identity and/or limited perspectives on young people, or it really is focused in a way that we don’t actually model. That’s not how we approach young people or our work... That’s a little bit more complicated, I think, but to say what about this is uncomfortable for us and articulate why, and then think about a way that we could work around this and/or how we would want to communicate [to the funder], why this doesn’t align with our model or our mission.

While funder-mandated metrics influenced the logistics of various agency evaluation processes, it also impacted the agency culture related to impressions, feelings, and concerns associated with assessment and evaluation. Thirteen of the participants who discussed agency metrics system development shared that mandated metrics influenced the overall agency evaluation culture. In these interviews, evaluation culture included impressions associated with how staff feels and thinks about evaluation; leadership philosophy about evaluation; and how well the nonprofit integrated evaluation into key functions and processes. Some participants felt their agency had a “minimal” evaluation culture, while others spoke about how developing systems around mandated metrics and cultivated a fully integrated assessment culture. Agencies with little evaluation culture sometimes had administrators focusing their time and energy on ways to shift the agency evaluation culture more positively. Melanie described the limited evaluation culture:
It’s going to be a culture shift a little bit and getting people—it’s always hard to get people to look at the bigger picture when they’re caught up in, this [co-worker] called in sick and I have to make phone calls for these [clients] for Saturday and all this other stuff. . . . I think if it becomes a routine, hopefully, it’ll get some of our program managers to more immediately think that way.

In agencies where administrators had or identified resources to improve systems, train staff, or hire staff with essential skill sets, participants discussed more substantial program staff “buy-in” associated with a positive evaluation culture. The agency participants who felt that they had a strong and fully integrated agency culture shared that evaluation systems were considered and included as part of the work, supported with resources, time, capacity, and planning efforts at all levels. Samantha talked about the culture at their agency, stating, “It is a pretty—it is a data-driven culture for sure, and I think it also has kind of folded into with our [clients] because they’re just used to [evaluation].” Likewise, Molly shared what an integrated agency evaluation culture looks like:

Every year, [staff and administrators] look at the various programs to determine what metrics are going to drive us towards those goals and also what we feel that we’re able to measure that accurately represent the work we’re doing and—it shows the level of work that staff are putting into it.

Agencies that were more successful at integrating an evaluation culture into their organizations and programming reflected that a more significant portion of the agency staff at all levels (from direct services to administrators) was included in the evaluation process and engaged in a more collaborative approach to evaluation systems and program development.

The following chapter will present a theory addressing the research question by assessing and discussing the various themes previously described and how they are connected. The next
chapter also includes implications for social work education, practice, and the promotion of social justice. Limitations and directions for future research are also examined.
CHAPTER FIVE

DISCUSSION

This chapter furthers the discussion of this research study examining the impact of funder-mandated performance metrics on organizations, services, and client populations. First, this chapter will begin with a review of the study’s key findings, including an in-depth discussion of the relevance of such findings, while also examining the emergent theoretical connections identified throughout the data. Then, the chapter will conclude by exploring significant implications for the fields of social work practice and education, as well as the promotion of social justice. Finally, limitations of the study will also be provided associated with the study design, methods, and analysis.

Review of Key Findings

Findings from this research study examined the influence of funder-mandated performance metrics on services within nonprofit organizations. 16 participants from eight nonprofit organizations were recruited using a purposive sampling method. Participants were interviewed in person, via phone, or Zoom over six months during one in-depth interview. Following the constructivist grounded theory approach developed by Charmaz (2014), data analysis followed a three-phase process. These phases focused on initial, intermediate, and advanced coding procedures correlated with low, medium, and high-level conceptual development levels (Birks & Mills, 2015). The findings identified three primary categories
where funder-mandated metrics had an influence within nonprofit organizations at the client, program, and agency level.

Client Category

Within the client category, three primary themes emerged as a result of analysis. The first theme underscored the key influence that funder-mandated metrics can have on how client success is defined. When mandated metrics are utilized in program evaluation and assessment systems, it can shape the parameters by which progress is defined and measured. Such a formalized and often rigid method or measure does not always fully account for individual client circumstances, needs, opportunities, or challenges. The second client theme highlighted the ways funder-mandated metrics could influence the relationship between a client and the agency employees they work with while participating in services. Analysis revealed that providers could play a significant role in supporting a client’s progress. Study participants raised concerns related to how required data collected as part of funder-mandated metrics might influence rapport building and potentially hinder the client-staff relationship. Participants also discussed the challenge that may arise when clients require services but do not meet the mandated required markers or benchmarks utilized to measure client success and/or program outcomes. Last, the third client theme focused on how funder-mandated metrics can directly influence client’s experiences when engaging in services. In some instances, metrics could be used to have a positive impact on clients, helping clients set goals or track their progress by using the assessment tools. However, for other clients, the use of mandated metrics in program evaluation could be experienced in a punitive manner, hindering their experiences, or creating additional barriers to engaging in services.
*Program Category*

Findings associated with the program category included three primary themes and two subthemes. The first theme illustrated the influence funder-mandated metrics could have on program staff time and capacity. The work required to facilitate performance measurement systems was often described as additional, time-consuming, and uncompensated work. Participants also raised concerns about the lack of trainings associated with conducting performance measurement, data collection methods, resources, and use of data entry systems to engage in this detailed work, contributing to an overall negative experience when implementing and collecting data.

The second program theme focused on the way that funder-mandated metrics influenced program-level assessment. The evaluation process often provides a way to monitor and evaluate the impact and quality of programs. Data collected from the evaluation systems also supported programs in identifying trends, barriers, or needs that could be addressed through service modifications. The culmination of utilizing required metrics also provided programs with informative data to demonstrate program successes and challenges in a transparent and accountable way to the agency, funders, and community alike. In addition to the context of program assessment, a subtheme was identified as participants shared additional concerns regarding the utilization of funder-mandated performance metrics. Questions were raised about the limitations of standardized or evidence-based evaluation tools, specifically concerning the historical exclusion of diverse populations in tool development. These issues could result in an inadequate representation of the program’s impact on clients. The third program theme identified ways in which mandated metrics could influence the delivery of services. For example, program
data could be used to guide changes to services or clinical approaches to care. Alternatively, programs often had to alter services to be more in line with a funder’s or stakeholder’s priorities or align services to meet the required measurement systems. A second subtheme emerged under the program theme, underscoring the importance of the agency relationship with the funder. Agency employees with a positive working relationship with funding entities felt it was easier to make changes to services that fit both the program and funder needs. However, in the absence of a positive relationship, participants described difficulties engaging in conversations around the influence of mandated metrics on clients, programs, or services. With less flexibility, program staff often felt compelled to comply with funder expectations to ensure continued financial support.

**Agency Category**

The third category focused on the influence of funder-mandated metrics on nonprofit agencies, which resulted in four key themes. The first program theme described how mandated metrics influence agency staff learning to utilize and interpret metric data. Participants reported having to use an array of tools to keep track of the requirements of multiple funders and collect additional information for internal agency use. They reported a lack of resources (e.g., time, financial, and/or technology) to input the data, resulting in a multitude of evaluation systems and a large amount of excess data that the agency did not have the capacity to evaluate or utilize. The second program theme illustrated the influence that mandated metrics have on agency-wide planning and decision making. Data collected from evaluation contributed to the information utilized to assess markers in strategic planning or goal setting. This also required significant
planning at the agency level to ensure that all required metrics were tracked, collected, analyzed, and reported to funders throughout different time points during the fiscal year.

Mandated metrics could also influence decisions regarding the expansion or closure of programs and assess how well services aligned with the agency's mission. The third program theme examined the influence of mandated metrics on agency financial planning, decision-making, and growth. The successful implementation of evaluation systems could influence the agency’s ability to increase revenue or result in revenue loss. Participants discussed a continuous tension between the need for external financial support and accepting funding that compromised the agency’s mission. Others discussed grappling with accepting funds from new revenue sources when the funder’s priorities were not aligned with the agency’s mission. There was also a critical need identified by participants calling for funders to allocate resources towards technology, software, and ongoing training to utilize the required evaluation systems effectively.

The last program theme illustrated the influence that mandated metrics might have upon how agencies approach evaluation system development and sustainability efforts. Funders often required agencies to develop their own tools or systems to track required metrics. These metrics influenced how agencies approached system development and the efforts required to sustain and maintain evaluation processes. Agencies often have to balance required forms of data collection for existing funders along with compiling additional data the agency deemed necessary for internal use.

The Purpose and Influence of Funder-Mandated Metrics

Overall, the findings from this study do suggest that funder-mandated metrics align with the intended purposes of evaluation systems at the program and agency levels, which is well-
supported by the literature (Behn, 2003; Cairns et al., 2005; Carman, 2011; Dawson, 2010; Despard, 2017; Lee & Clerkin, 2017b; LeRoux & Wright, 2010; Moxham, 2009). The findings further suggest that the implementation of performance evaluation systems can influence program and agency functioning, creating additional challenges to be addressed as a result of utilizing the funder-mandated metrics. Notably, these findings support the minimal literature available on the topic of the influence of funder-mandated metrics (Bryan et al., 2021; Carman & Fredericks, 2010; Carnochan et al., 2014; Kim et al., 2019; Lee & Clerkin, 2017a; Williams-Gray, 2016). Finally, findings at the client level suggest several unintended influences that funder-mandated metrics can have on service provision, an area in which the literature is relatively sparse (Beer & Micheli, 2018; Benjamin & Campbell, 2015; Carnochan et al., 2014; Mosley et al., 2019). The following discussion provides a more in-depth examination of the findings supported by the literature while also exploring the findings that are notably absent in scholarly literature, associated with the intended and unintended purposes of evaluation. First, the following discussion examines the intended purpose of evaluation and utilization of funder-mandated metrics, followed by discourse centering upon the additional influence of mandated metrics, concluding with an examination of the unintended influence of mandated metrics.

**Intended Purpose of Evaluation and Utilization of Funder-Mandated Metrics**

At the program and agency levels, evaluation is intended to assist nonprofits in several ways. The data collected from performance evaluation systems can be used to evaluate, learn, inform decision making, as well as to make a positive impact upon changes to services, and aid in the budgeting process. Much of the research to date reflects that funder-mandated metrics
assist agencies in taking these actions (Behn, 2003; Cairns et al., 2005; Dawson, 2010; Moxham, 2009); key findings from this study support this literature.

For example, at the program level, findings from this study indicated that data collected from funder-mandated metrics were primarily used to assess programs. Utilizing the data provided staff with an opportunity to examine the overall program achievements towards the pre-defined goals and identify the service impact upon clients, as extensively highlighted across the literature (Behn, 2003; Cairns et al., 2005; Dawson, 2010; Moxham, 2009). Utilizing outcome data also enabled staff to monitor ongoing program progress (or lack thereof) and see the differences before and after an intervention or service had taken place (Behn, 2003; Dawson, 2010; Moxham, 2009).

Additionally, the data can be utilized to report to funders and other stakeholders about program outcomes and the effectiveness of services while conveying transparency and accountability (Cordery & Sinclair, 2013; Despard, 2017; Moura et al., 2019; Treinta et al., 2020). As nonprofit agencies continue to compete for funding, institutional theory posits that the use of evaluation data to convey this transparency and accountability creates organizational legitimacy and increases opportunities for an organization to obtain resources for its ongoing survival and growth (Carman, 2011; Lee & Clerkin, 2017a). Funder-mandated metrics and assessment systems can continue to contribute to these efforts toward legitimacy and can help agencies more readily position themselves to compete for funding. However, it should be noted that, while the literature supports the use of evaluation systems as a form of program assessment, to date, there is limited scholarly attention given to examining whether outcome measures are
relevant to the program and if they are actually linked to program success and agency mission attainment (Lee & Clerkin, 2017a; LeRoux & Wright, 2010).

Furthermore, findings from this study supported by the literature suggest that funder-mandated metrics can influence program service modifications. The data derived from program assessments aided in identifying program trends or themes. For many programs, the data provided direct client feedback on the quality of care and supported efforts to understand how clients best respond to different models of care, interventions, or services (Carman, 2013b; Carnochan et al., 2014). The program assessment process can provide a baseline analysis and help staff to continue gathering additional information for comparison, resulting in key data and information necessary to improve service outcomes (Despard, 2017). These data might then inform the decision-making process related to changes in service delivery, intervention methods, or modifications to client care approaches (Carman, 2007, 2011).

Similarly, existing research indicates that performance measurement data can be utilized to learn about the agency as a whole and inform agency-wide planning and budgeting processes (Lee, 2020; Lee & Clerkin, 2017a; LeRoux & Wright, 2010). Findings from the current study support these results, with analysis revealing that the culmination of all program data can serve as a learning tool to reflect on individual programming and provide insight into agency strengths, weaknesses, and opportunities for growth (Lee, 2020; LeRoux & Wright, 2010). This information can then guide infrastructure decisions associated with strategic planning, reorganization, goal changes, capacity building, or the need for program closures. In addition, agency leadership can obtain a better overall understanding of the inner workings of the agency and identify programs demonstrating strong markers of progress or hone in on any program or
staffing issues that need direction, support, or intervention (Behn, 2003; Bryan et al., 2021; Carman, 2013a; Festen & Philbin, 2010).

All of the data used to support agency decisions can then be utilized to present a compelling justification to funders and other stakeholders for any changes and potentially make a stronger case to increase or expand funding (Carman, 2007, 2013a; Williams-Gray, 2014). In addition, findings from this study suggest that data can identify where financial support needs to be allocated and what additional resources the agency needs to obtain in order to accomplish its goals and mission (Behn, 2003; Prentice, 2016).

Additional Influence of Funder-Mandated Metrics

The previous section discussed the study findings supported by extensive literature associated with the intended influence of mandated metrics. These metrics aid agencies in program assessment, agency planning, budgeting decisions, and service delivery changes. The following discussion centers on the additional influence mandated metrics can have on program and agency employees when utilizing the metrics as intended. Findings from this study suggest that the agency’s commitment to using mandated metrics and evaluation systems inadvertently creates additional employee responsibilities and can create power dynamic challenges among funders. Some of these challenges include increased staff time for data entry and reporting, capacity concerns, resource strain, and navigating the power dynamics of funder relationships with regard to service modifications and agency planning. Several of these findings have been acknowledged throughout the literature (Carman & Fredericks, 2010; Kim et al., 2019), but there remains a dearth of research and scholarship devoted to examining these topics.
The requirement of more complex and/or expansive performance measurement systems has resulted in nonprofit organizations increasing the allocation of staff time toward managing these efforts (Carman & Fredericks, 2010; Carnochan et al., 2014; Kim et al., 2019; Lee & Clerkin, 2017b). Study findings suggest that program staff time was negatively impacted due to the increased responsibilities and often uncompensated work associated with expectations for evaluation efforts in addition to primary and existing job requirements. Such additional job responsibilities often reduced the amount of time program staff spent on direct care and delivering services, while also increasing the overall time employees had to work to complete their job requirements successfully. Analysis of general participant responses associated with this finding was negative and pessimistic, often viewing these responsibilities as a burden, time consuming, and cumbersome. These findings also raise questions associated with how the lack of human capital may influence data collection quality and to what extent data analysis actually provides adequate information to assess or improve services accurately.

The findings associated with staff time appear to be related to staff capacity as well. The data suggests that direct service providers and administrators often did not have the necessary performance measurement skill set, background and training, or resources. Increased demands on staff time and a lack of resources necessary to develop these skills could be reducing the quality and integrity of the data collected, let alone other job responsibilities and client care provision. Previous research has found that organizational capacity could significantly influence the successful use of performance metrics for planning and budgeting (Behn, 2003; Carman, 2013b; LeRoux & Wright, 2010; Willems et al., 2014). It could be possible that agencies facing capacity barriers may be spending more time and energy collecting and analyzing mediocre or
misrepresented data, reporting it to funders, and executing their overall mission based on an inaccurate assessment of potentially flawed findings. Therefore, if an agency is struggling to allocate resources towards performance measurement systems, it should not be surprising if it cannot successfully convert the data into meaningful or useful information during the analysis or reporting phase.

In addition to concerns regarding staff capacity, previous research has indicated that one of the most significant obstacles to successful evaluation implementation is a lack of financial resources to support the investment necessary to develop and maintain quality systems (Bryan, 2019; Bryan et al., 2021; Carman & Fredericks, 2010; Williams-Gray, 2016). A lack of funding for such systems often results in nonprofit organizations lagging in technology purchases and resources that enable the implementation of efficient metrics systems. In addition, the lack of support for necessary human and technology infrastructure can lead to agencies struggling with how to best collect data in a way that meets the needs of their funders constricted by limited financial means to do so (Bryan et al., 2021; Carman & Fredericks, 2010; Moxham, 2009; Williams-Gray, 2016). The current study supports these findings, with analysis revealing that mandated metrics and evaluation systems influenced how agencies developed their evaluation tools. Participants shared that their agencies struggled to determine what systems to use given their limited financial capacity and lack of capacity-building resources and training for staff. As funders continue to evaluate programs and agencies as part of the accountability movement, nonprofit leaders will likely face considerable ongoing pressure to demonstrate program effectiveness and efficiency, as well as client performance and outcomes, as defined by external funding entities. However, as previously discussed, studies suggest that despite agency efforts to
increase the number of measurement tools and reporting mechanisms, limited evidence suggests that nonprofits are measurably improving upon their performance (Bryan et al., 2021; Ebrahim, 2019).

Moreover, there is limited evidence to indicate that program evaluation produces adequate and effective data that best serve the organization or clients (Lynch-Cerullo & Cooney, 2011; Moxham, 2014). Rather, program evaluation may only be demonstrating an ability to complete evaluation for the sake of compliance but is less helpful to organization functionality (Bryan et al., 2021; Carman, 2010; Hoefer, 2000). With the lack of capacity necessary to develop and sustain time consuming and often multiple evaluation systems, agencies might be functioning much less efficiently, negatively impacting the quality and quantity of time allocated to providing direct care and services. The lack of financial investment in employees and agency resources has historically placed the nonprofit sector in a position far behind in technology and technological capacity when compared to the for-profit sector (Akintobi et al., 2012; Bryan et al., 2021; Glisson, 2015; McBeath et al., 2019; Parsons et al., 2017).

This highlights the issue of nonprofit overhead ratios, which have long been used as a proxy for nonprofit efficiency. The assumption is often made that higher levels of budgetary support for non-program expenses are evidence of wasteful or inefficient spending (Park & Matkin, 2021; Qu & Daniel, 2021; Tian et al., 2020). This mindset significantly minimizes the investments that agency staff can make towards developing and sustaining quality evaluation systems. Although there is strong evidence to suggest that limiting overhead can greatly constrain the overall sustainability of agencies and even debilitate their effectiveness over time, there continues to be a powerful mindset from funders that nonprofit organizations should not be
“wasting” money on overhead and should prioritize funds directly for programming (Charles et al., 2020; Qu & Daniel, 2021). Funders may need to reconsider financially investing in and supporting organizational efforts to enhance performance measurement that best serves the client populations as well as the agencies and funders. For many funders, this could require a significant shift in thinking about nonprofit overhead expenses. Investment in evaluation systems, personnel, training, up-to-date technology, and outside-the-box evaluation concepts would not necessarily take away from providing direct services. Rather, this could be an important investment to cultivate a greater understanding of agency and programmatic impact, addressing critical social issues, and providing key information that may result in necessary innovations and/or large-scale social changes.

Lastly, funder-mandated metric system requirements and implementation influenced the relationships between nonprofits and their financial funders. As previously discussed, mandated metrics can provide useful information to help nonprofit organizations make important decisions, plan, and effectively budget. However, findings from this study suggest that these agencies may have varying levels of autonomy over this process based on their relationship with various funders. Metrics often create the parameters for services or inform the agency’s development of their theory of change, based on funder priorities conveyed through data collection, reporting, and analysis requirements.

Viewed through the lens of resource dependency theory and institutional theory, the demands of funders, and nonprofits working to satisfy those demands, can create a power dynamic that affects the overall relationship (Carman, 2011; Thomson, 2010, 2011). Agencies reporting stronger and more positive relationships with funders conveyed a sense of a more
equitable relationship in which power was collectively shared. These relationships provided more autonomy over program decision making, approaches to service provision, and decisions related to what was to be measured. For agencies with underdeveloped or strained relationships, this study found more feelings of coercive pressure to make programmatic changes based on funder expectations and create evaluation systems that met external needs but not necessarily internal priorities or a focus on client care and success. The financial decision-making power of the funder resulted in fewer opportunities to negotiate, collaborate, or discuss agency needs. Agencies were more likely to submit to those external pressures rather than risk a reduction or loss in funding or strained relationship with a funder. Such power dynamics have the potential to create program goal misalignment or cause agencies to experience mission drift, as well as move away from an emphasis on client-centered care provision.

**Unintended Influence of Funder-Mandated Metrics**

There is significantly less research and literature examining the potential unintended influences of funder-mandated metrics. However, the findings of this study suggest there are several ways that funder-mandated metrics, both directly and indirectly, can influence client experiences with services and outcomes based upon receipt of care. Examples of the unintended influence of mandated metrics may relate to varying definitions of client success, and the nature of the client-provider relationship, among other areas.

When metrics are required for organizations, they can impose subjectivity and may also place social values and meanings upon the definition and measurement of client success (Beer & Micheli, 2018). A funder defining desired behaviors or outcomes raises questions about whether those outcomes are based on a genuine understanding of a client’s individual background and
needs; environmental, historical, and structural factors; and the importance of self-determination, encouraging individuals to participate actively in defining their treatment or approaches to care (Carman & Fredericks, 2010; Carnochan et al., 2014). If agencies are already serving people who are marginalized, the use of externally constructed interventions and assessments can disproportionately subject a client to behavior change interventions not based on principles of empowerment or resilience (Mosley et al., 2019). Such metric systems often focus on client behaviors while minimizing systemic structures that potentially contribute to or perpetuate a client’s circumstances and lived experiences. It raises questions related to failure in considering inequalities at the mezzo (community) and macro level of policies when determining appropriate metrics to gauge success (Mosley et al., 2019). These challenges are further examined in the implications for social work practice and social justice section explored later in this chapter.

Social workers are trained to understand the needs of their clients within the context of the environment, and therefore quite often, nuanced practice approaches with clients can make it challenging to assume a predetermined definition of success. The findings from this study suggest an individualist mentality of progress may fail to acknowledge the often nonlinear pathway or any systemic challenges that might pose significant barriers to a client’s eventual success (Benjamin & Campbell, 2015; Carman & Fredericks, 2010). Focusing solely on a “bootstraps” approach to assume or define client success is problematic. This narrow approach raises critical questions about the lack of emphasis given to the impact culture, privilege, and ideology can have perpetuating forms of oppression, while also investing in interventions that place the onus of responsibility for change solely upon the individual.
Similarly, the findings from this study suggest that applying industry standards can be a valuable way to compare client outcomes with national research and benchmarks while providing helpful insight into the client service process. However, this also certainly raises questions related to the development of such industry standards. Is one model of evaluation used to create an industry standard, thus privileging certain types of interventions? If so, continued research should examine the influence that such interventions have upon unique client populations that do not fit within the model upon which the standard was built. For example, most research studies still privilege White, cisgender, and heterosexual individuals (Mosley et al., 2019), raising questions about the generalizability of findings among populations that have been historically excluded from such studies based on race, gender, sexual orientation, etc.

Additionally, the use of funder-mandated metrics that correspond with industry standards and definitions places the potential power with funders to dramatically influence such industry standards and norms (Thomson, 2010). Implementing evaluation standards that do not include an agency’s client population might perpetuate the application of socially constructed norms that are not reflective of a client’s reality of “success” or their ability to achieve “success” as defined. However, the challenge remains that failure to comply with industry standards can make agencies appear “non-compliant” or perceived as “less legitimate,” placing them at risk of losing necessary funding.

Additional findings from this study suggest that funder-mandated metrics can influence definitions of client success due to the lack of inclusion or consideration of cultural nuances as well as structural and historical factors in the metrics. For example, one participant shared that success in a housing program was defined as an individual moving into their own apartment or
home. This dominant cultural expectation views living alone as a marker of independence but does not account for different cultural considerations, such as the commonality of multigenerational households cohabitating among certain family groups. Applying such a narrow definition of success can result in specific populations being reported with higher rates of unsuccessful outcomes, solely based on the failure to be more culturally inclusive.

Further, there may not be consideration given or ability to weigh influential external variables (e.g., experiences of discrimination or factors stemming from forms of oppression) that cannot be controlled or predicted, but which can lead to vastly divergent outcomes for each client (Bromley & Meyer, 2017; Mosley et al., 2019; Thomson, 2010). The participants in this study raised such concerns and questions about the numerous ways that quantifying definitions of client success and corresponding outcomes can influence when client success or progress is rewarded or becomes punitive.

The present analysis suggests that mandated metrics’ definitions of success can be used to incentivize client engagement and set parameters for client participation in services. Though this may be perceived as a positive influence, such an approach solely underscores the concept of personal responsibility as the pathway to success, diminishing the role of systemic issues that may be creating barriers to access, care, or progress toward meeting the ultimate outcome. While ignoring the social constructs at hand, blame for any failure to meet desired outcomes or make expected progress is often passed on to the client (Bach-Mortensen & Montgomery, 2018; Benjamin & Campbell, 2015; Mosley et al., 2019). It may be argued that solely placing this responsibility for success on an individual removes any agency or funder responsibility to
actively focus on creating social change and equity, thus perpetuating a system of oppression that promotes the status quo.

Another theme from this study highlights the unintended influence of funder-mandated metrics associated with the client and provider relationship. Providers can play a pivotal and supportive role throughout a client’s journey and progress while they are engaging in care, services, or treatment. While the relationships are individual and unique, each provider’s preferences, experiences, judgments, and biases often shape service delivery (Mosley et al., 2019). Some research even suggests that the racial and gender identities of direct service staff and clients interact in complex ways, potentially shaping service experiences differently (Watkins-Hayes, 2009). In addition, direct service providers build rapport and unique relationships with clients, variables that are not considered within the context of predetermined metrics or definitions of success.

Findings from this study suggest that mandated metrics can influence the flexibility or inflexibility of service delivery. When mandated metrics come with a limited definition of client success or narrow scope of clinical interventions, services become more regimented. These restrictions can limit the input a client has with regard to their care plan, reduce outside-the-box thinking, discount practice wisdom, and decrease professional autonomy. The prioritized mandated evaluation outcomes also waylay valuable community-based knowledge, again increasing the likelihood that community culture, history, and traditions are ignored. As the standard expectation of evidence-based models or externally constructed client outcomes and success continue to equate with agency legitimacy in the eyes of financial funders, providers may experience pressure to conform to a narrower approach to care. When faced with the
pressure to meet reporting requirements, direct service providers may have less autonomy to meet clients “where they are” and explore alternative treatment or intervention options, therefore limiting self-efficacy, empowerment, and promotion of social justice (Benjamin, 2008; Benjamin & Campbell, 2015; Carman & Fredericks, 2010; Thomson, 2011).

Findings also suggest that as a result of service inflexibility, providers can feel conflicted, perpetually aware of the predetermined external expectations set by funders and metrics, while also attentive to an individual client’s needs and circumstances, which can make determining care for clients complex and challenging. Collaborating with clients and working from a strengths perspective, providers often work together with their clients in developing individualized program goals and treatment plans. Clients are not always ready or able to fully complete programming or services, solely based on the definition of successful outcomes. However, they may be able to make progress on their own terms or over an extended period of time. Participants conveyed that providers were often faced with the reality that they would have to work with clients who would more likely meet the predetermined outcomes. This resulted in limited or no access to clients with more significant support needs, thus hindering their ability to make any progress or access any services. A continued focus on individualist, predefined approaches to and definitions of client success might be potentially harming the most marginalized groups (Mosley et al., 2019).

This discussion highlights that there is value in utilizing funder-mandated performance metrics and other evaluation systems while equally illustrating potential issues that may hinder client outcomes, specifically with the lack of emphasis on measuring success individually and via assessing one’s social environment. In the profession of social work, we must ensure that
“the legitimacy of science and data does not become the end in itself, with a concurrent decline in the actual ability of the field to produce significant social change” (Mosley et al., 2019, p. 2). Thus, the profession’s continued research and scholarship in this area must include a comprehensive examination of the programs that successfully aid and support clients, as well as the effective design of services, systems, and policies at the macro level that address structural causes of inequity and inequality. These ideas are further explored in the implications section. The discussion section will examine theory development, which explores the connections and relationships between key findings from this study as well as relevant themes.

**Emergent Theory and Connections**

In keeping with the constructivist grounded theory approach, the aims of this research study were to interpret the findings and examine the connections between the themes, as described by the participants’ perceptions. A theory is presented that provides an explanatory scheme comprising all the concepts related to each other through patterns of connectivity that emerged (Birks & Mills, 2015; Charmaz, 2014). While moving through the various levels of coding and reviewing the codebook, creating an audit trail and memos, and analyzing participant interview transcripts, I realized that funder-mandated metrics influenced agency practice at the agency, program, and client levels. In addition, although these themes were distinctly identifiable within their own categories, there was also data to suggest that many of the themes influenced one another. Through continued analysis and concept mapping, I examined a more comprehensive picture of the participants’ experiences and their perceptions of the reach that funder-mandated metrics can have within a nonprofit organization, particularly exploring the intersectionality between and across the categories. When funder-mandated metrics influenced
one area of service provision, there was a continuing influence over one or more areas of agency functioning (Figure 4). Therefore, the reach of funder-mandated metrics spans across client, program, and agency levels, while all interconnecting to influence many aspects of organizational functioning. The following sections examine the connections and relationships between the themes within each category.

Figure 4. Connections Between Category Themes

**Agency-Program Connections**

The agency category interviews revealed that funder-mandated metrics influenced the data learning process, organization-level planning, financial planning, evaluation system
development, and sustainability. Further analysis revealed that there were connections between these agency-level influences and the findings at the program level. For example, when participants spoke of the ways that evaluation data were utilized for agency-level learning, they also shared that this information was gathered from the program-level assessments and pooled to form a larger picture of agency functionality. The data sets collected could also inform modifications at the program level and assist in any changes across programming or agency processes as a result of the information learned.

Participants shared stories to highlight ways that organizational planning and budgeting directly influenced functioning at the program level, including issues related to staff time and capacity, assessment of program effectiveness, and modifications to program-level service delivery. Through data learning, agency leadership can set long-term goals, make changes to the organizational or programmatic budget, conduct personnel recruitment, seek professional development strategies, and more. These decisions influenced how services were provided, with programs adjusting to address budgetary changes or agency-wide planning efforts. Finally, agency-level planning and budgeting influenced the time program leadership staff and frontline workers were able to allocate to services, modified job descriptions based on metrics requirements, and even shifted how or to whom services could be provided. Funder-mandated metrics requirements that influence decisions made at the agency level and result in program capacity issues for direct service providers. For example, an agency level decision to accept funding with new evaluation requirements, may require an investment (e.g., time, financial, technology) in training program staff, hiring people with the appropriate skill sets or the use of new program technology to track evaluation data. If a program does not have the financial or
other resources to meet these needs, staff may face significant capacity issues and challenges, such as trying to adapt inadequate technology to meet the funder needs.

**Agency-Client Connections**

In addition to connections between agency and program themes, participant interviews conveyed the connections between agency and client themes. When discussing the influence of funder-mandated metrics on agency planning and financials, there was also discussion about the implications these decisions or changes could have on the client-provider relationship and their direct influence on the client. For example, one participant explained that metrics requirements shifted the way agencies instructed direct service workers to provide care, and in some cases, this also limited services. She explained that this shift in care resulted in program staff members having less time to spend with their existing clients, fewer referrals, and a narrower scope of goals that could be set due to limited time and capacity. Clients reported feeling that they had less of a say in setting their goals, had fewer opportunities to access referral services, often felt upset with their provider, and faced more barriers when trying to meet their needs. Thus, the decisions made at the agency level had a ripple effect that influenced services and, as a result, direct client experiences with their providers.

In addition, definitions of client success were greatly influenced by decisions made at the agency planning level and through evaluation system implementation, rather than being led by the client populations themselves. Leadership utilized the combined program data to inform agency level planning while also incorporating the needs of funders. This led to changes in definitions of client success to solely align with funder priorities, thus, reducing client self-efficacy, empowerment and limiting the participatory role clients could have when setting goals.
with their provider. Additionally, agency decisions to implement evaluation methods also influenced how client outcomes were defined through the framework of mandated metrics.

**Client-Program Connections**

Last, there were additional connections between program and client themes. The definitions of client success, client-provider relationship, and influences on the client all had connections with program assessment and service modification.

Participants discussed that funder-mandated metrics often influenced how client success or outcomes were defined because they had to align with the mandated metric or fit within a predetermined set of requirements. Nearly all participants shared a story or example of a time when a client outcome was viewed as unsuccessful by a metric, but the provider believed the client to be making excellent progress given their individual circumstances and barriers to overcome, as well as in relation to their personal goals. For example, one participant shared that her clients often “failed” to obtain employment as defined by the funder. However, these clients primarily worked in a cash economy and did not have bank accounts, which made it very difficult to show proof of employment in the way that was expected. This example underscores how definitions of client success can be altered at the program level and how these service modifications can influence the client’s overall experience with services. When the parameters for success are not attainable for the client given their personal circumstances, the program can feel punitive because clients are not able to achieve this standard of success.

The program assessment process tied to funder-mandated metrics also influenced the relationship with providers. Program providers could utilize the assessment data to better understand or meet their clients’ needs, build stronger relationships, or limit the scope of
services, as discussed earlier in the agency and client sections above. When program assessments restricted the scope of care, providers were often faced with a difficult decision to either prioritize clients they believe could meet the assessment benchmarks or work with clients knowing they likely would not meet metrics criteria, negatively impacting the reporting data provided to funders.

When the data connections between the different category themes are examined, it is apparent that funder-mandated metrics can influence multiple areas of practice. Agency decisions often influenced both program functioning and client experiences. This intersectionality reflects how vast a reach funder-mandated metrics can have within a nonprofit organization—influencing long-term, agency-wide planning as well as the day-to-day experiences of the clients.

**Implications**

Performance evaluation systems have become an accepted (and often required) aspect of nonprofit funding and functionality since the 1990s in the United States. Such systems are also required for many nonprofits to competitively compete for funding at the local, state, and national levels. As nonprofit organizations continue to utilize performance measurement systems, an understanding of the influence of funder-mandated metrics helps to provide insight into the experiences of administrators, program staff, and client populations, and the challenges and opportunities they face when working under such requirements. Thus, there are a multitude of implications for funders, social work practice, and the field of social work education that emerge. The following examines the implications for these areas while also highlighting the social justice implications.
Funders

There are several implications for funders that support nonprofit organizations. First, results of this study along with previous literature identify a clear need for funders to take direct action towards building capacity among nonprofits related to the ability to effectively conduct performance evaluation (Cairns et al., 2005; Carman, 2011; Despard, 2017; Lee & Clerkin, 2017a; LeRoux & Wright, 2010; Moxham, 2009). Second, there is a unique opportunity that exists for funders to directly support organizations by sorting out specifically what individual agencies and their client populations truly need rather than assuming a “one size fits all” model of performance measurement, standards, and expectations. For example, building evaluation capacity might result in financial investment in technology for performance measurement at one organization or personnel dedicated solely to conduct evaluation at another. Resource needs could also be supported through staff training or creating learning networks that include peer learning models, collaborating with community leaders, and/or mentoring of nonprofit leadership across organizations. In addition, building capacity around performance measurement systems can help nonprofit organizations address some of the documented barriers that agency employees are experiencing, while also collaboratively exploring how outcome measures are relevant to service provision and link to client success and the accomplishment of an agency’s mission alike.

Funder support for building capacity based on individual organizational needs or through a collaborative approach also creates an opportunity to examine current practices around the influence of funder-mandated metrics on the asymmetrical power dynamics among funders and recipients. Requiring the use of metrics defined by funders places many nonprofit organizations in a position of upward accountability to the funder. However, care must be taken from a social
justice perspective when considering the adequacy of the indicators used (Campbell & Lambright, 2016). An organization’s ability to assess and provide evaluation results does not mean that the agency is inherently effective, especially if the mandated metrics are rooted in false narratives about the causes of social issues or if the defined outcomes are not those desired by the clients and communities being served. Examining and potentially addressing these power dynamics could create opportunities to learn more about how resource dependency shapes performance measurement practices. For example, funders could request feedback about current benchmark practices, reporting requirements, barriers to program implementation, and areas to support capacity building efforts around performance evaluation. Additionally, funders could create spaces for nonprofits addressing the same or similar social issues to network, express their shared understanding of client and community issues, and look for ways to support those agency-identified needs.

There are also implications for funders with regard to the current system of performance measurement. Findings from this study related to the unintended influence of definitions of client success underscore concerns associated with social justice that warrant reconsideration of current practices in several ways. First, the use of mandated metrics raises questions about what practices are privileged, what assumptions are normalized, culture-bound, and/or deemed “appropriate” or “desirable.” Imposing mandated metrics has the potential to reaffirm the view that marginalized populations need to be “fixed,” “saved,” or conform to standards of more privileged cultures while situating the problem with the individual (Mosley et al., 2019).

Secondly, with the majority of philanthropic organizations operated primarily by White individuals and board members (Dorsey et al., 2020; Sullivan, 2020), there are concerns that
mandating metrics or imposing parameters for outcomes could fail to challenge privileged culture-bound norms. A lack of meaningful, diverse board representation can diminish higher levels of critical analysis when members who effectively represent the organization’s constituency are excluded from membership. Homogeneous boards are less likely to consider alternative perspectives or varied approaches to evaluation. Instead, they continue to conform to the norms of the past or to those that are reflective of their experiences (Bromley & Meyer, 2017; Dorsey et al., 2020; McConville & Cordery, 2018).

Additionally, these findings suggest that funders may need to explore how data are being effectively utilized (or not) and examine any potential alternatives to the current data collection and analysis models. This is not to say that existing measures are wholly uninformative or flawed but suggests reconsideration in striving for measures that have a more universal application to the nonprofit sector as a whole (Pennerstorfer & Rutherford, 2019). For example, funders could reward organizations that use evaluation information in important and meaningful ways that best support clients and service provision or support agencies in designing their own evaluation systems that are mutually beneficial (Carman, 2011; Carman & Fredericks, 2010). This fosters organizational curiosity and learning while also creating the potential to use community definitions of social issues and benchmarks for progress in a more participatory fashion.

Lastly, the findings suggest a need to value client-defined outcomes, which may differ from those articulated through funder-mandated metrics. Placing greater value on client-defined short-term goals provides flexibility to determine best how to meet the client’s needs with their provider and within the parameters of the client’s environment, culture, and experiences. This client-centered approach based in empowerment may require funders to shift their approach to
philanthropy, following the lead of the communities being served, and collaboratively reshaping the expectations of performance evaluation systems.

**Social Work Practice**

There are also several implications associated with social work practice. First, the need to build capacity around performance measurement applies to efforts at the nonprofit level, as well as with funders. Several capacity and training issues and barriers have been identified in the literature (Bryan et al., 2021; Carman & Fredericks, 2010; Glisson, 2015), but each agency has the best understanding of its needs, limitations, and strengths. Second, as nonprofit employees make efforts to impact social change, agencies must articulate their own vision for promoting social justice. Those efforts may include difficult conversations with funders regarding an investment in overhead support or other resources, in addition to any issues or concerns related to the implementation of funder-mandated metrics. Organizations may also need to shift their fundraising efforts towards private donors or funders with evaluation practices that provide greater agency autonomy in leading the performance evaluation process.

The agency may also need to cultivate more ownership and responsibility to build capacity and staff skill sets around program evaluation, regardless of funder requirements, in order to be true to their mission and genuinely serve the community. For example, job descriptions, evaluation requirements, and compensation may need to be reassessed to clearly identify staff roles and ensure the necessary support to engage in evaluation processes, such as additional training or peer mentorship. Hiring staff with the necessary skill sets or creating new positions could also increase capacity efforts.
The unintended metrics influence on definitions of client success and on the client-provider relationship also has implications for practice. Organizations may want to reexamine their current evaluation methods and approaches, looking critically at the assumptions imposed by privileged cultural norms. If services are designed and measured to examine behavior change or personal progress without consideration for the context of their clients’ lives, the services may assist them to some degree, but will not (nor are they designed to) transform societal structures or lead to political change (Mosley et al., 2019). Organizations might need to examine the influence of evaluation systems on a client’s right to self-determination, and the inclusion or exclusion of community-based knowledge, definitions of social issues, and markers of progress. Identifying these issues creates opportunities to address unintentional barriers or problematic evaluation standards, ensuring that the agency is not perpetuating any systems of oppression. Supporting a client’s right to self-determination and respecting their dignity and worth is required at the individual, mezzo, and macro levels of practice.

In addition, social workers and other nonprofit employees should recognize the relational nature of the work. Client and provider relationships should respect client autonomy and agency, be collaborative with the communities served, and tailor services to meet community needs best. Social workers and other frontline staff members are uniquely positioned in their day-to-day work to examine and respond to the implications of mandated metrics. Their unique perspectives can inform evaluation systems and integration with greater attention towards organizational priorities and increased focus on organizational learning. These providers may also be able to identify creative solutions or changes for current practices and provide context for funder narratives, explaining the necessity for changes or modifications to evaluation systems.
Finally, at the practice level, data suggest that funder-mandated metrics greatly influence client experiences, their autonomy over service options, and ultimately their outcomes. As social workers, one of the most important tools we have to serve communities and support clients is through nonprofit organizations. The approaches to programming are typically assumed to be catalysts of social justice, supporting client growth and empowerment, particularly when utilizing evidence-based practice and evaluation systems. However, the findings of this study suggest a need for caution as intervention, implementation, and evaluation science become increasingly central to our work (Mosley et al., 2019). As professionals guided by the ethical principles to challenge social injustice and respect the inherent dignity and worth of a person, greater attention must be paid to the role funders have in determining the outcomes of others. For example, data collected from metrics can be used to inform best practices or shape policy. However, if these data continue to be unexamined, it has the potential to perpetuate outcome expectations based on privileged culture and reinforce the idea that marginalized populations need to conform to these expectations if they are to be viewed as successful. Moreover, as nonprofit organizations continue to abide by current evaluation processes that focus the responsibility of change on the client, an unintended consequence might be that the nonprofit system is unintentionally absolving itself from societal responsibility to address social change.

Social Work Education

Examining the social justice implications of how performance evaluation is shaped by privileged culture is a critical part of cultivating strong social work advocates and agents of social change. Preparing students to do so has several implications for social work education, undergraduate and graduate curricula, and teaching methods. First, interprofessional
opportunities have the potential to facilitate dialogue regarding some of the topics raised in this study, such as barriers to successful evaluation implementation or issues identified around definitions of client success. Social workers employed at nonprofit organizations work alongside individuals with various degrees, and board members have a variety of educational backgrounds. Students from these various backgrounds might benefit from interprofessional education and training opportunities (de Saxe Zerden et al., 2018; Rubin et al., 2018) to more effectively prepare them for performance evaluation in multidisciplinary environments. For example, building interdisciplinary courses, assignments, or projects connecting social work students with students in fields such as business, philanthropy, public health, or public administration would create space for co-learning. Curricula developed for interprofessional courses create an opportunity to learn about other professional approaches to practice, evaluation, and assessment. They could provide insight into how each profession approaches evaluation, what values and perspectives are placed on evaluation systems, and what influences definitions of client success. Social workers have a unique educational perspective that focuses on viewing the client as part of their socially constructed environments and through a social justice lens. This perspective could bring about important conservations and provide a unique opportunity for students to co-construct knowledge, such as the implications of metrics developed within culturally bound norms or examining what assumptions are made about client circumstances or what outcomes are deemed appropriate for them.

Second, future development of social work curricula should consider the implications of nonprofit evaluation and performance measurement metrics on direct services and client populations. Micro, mezzo, and macro practice classes could benefit from educational learning
opportunities around evaluation systems and the influence of performance metrics on service provision. Many social work courses still divide courses into a micro versus mezzo/macro practice framework, failing to reflect the integrated role of social workers practicing in nonprofit organizations across all levels of practice (Ezell et al., 2004; Hill et al., 2010; Iverson et al., 2021). Educational opportunities associated with performance evaluation in classroom discussions, assignments, and/or via field opportunities could create unique moments for students to understand the intersection among direct practice; nonprofit administration; the role of community, policy, and advocacy work (Iverson et al., 2021; Rothman & Mizrahi, 2014).

For example, research classes that typically focus on design, methods, evaluation, and assessment might examine the practice challenges associated with staff capacity related to navigating evaluation systems and mandated performance metrics. Social work courses could provide students with educational experiences that can translate into applicable skills and increase social workers’ confidence entering the profession. Advancing the scientific methodologies vital to nonprofit organization needs could address the evaluation and skill set capacity issues identified. Providing students with the opportunity to learn about the role performance evaluation has within nonprofits can better prepare them to engage in assessment as a provider, have a more holistic understanding regarding the purpose of funder-mandated evaluation, and contribute to any evaluation systems development necessary.

Lastly, policy and practice courses could also be impacted by learning about nonprofit evaluation standards and challenges. As evaluation data can shape direct practice approaches and policy, it is imperative to incorporate conversations associated with the practices that shape and inform the performance evaluation process. An examination of the individualistic, “bootstraps”
mentality and approach to evaluation and universal applicability metrics raises critical questions regarding what external influences shape approaches to care.

**Future Directions for Research**

Findings from this study present several key findings that should be examined in future research. First, while this study solicited the perceptions of participants from a diverse set of nonprofit organizations, future research should seek to explore the influence of funder-mandated performance metrics on service provision across a broader demographic of agencies and providers. For example, this study included two agency budget categories, including 1 to 4.99 million and 5 million dollars or more. While there were no unique differences between these two budget categories, ongoing research examining agencies with smaller budgets may result in differing perspectives. There was also a notable difference in the influence of metrics on services associated with government sources of funding rather than via private funding sources that future research should explore. Participants from this study came from a large urban setting compared to those working within rural areas. There are significant differences between urban and rural service provision, the impact of funder-mandated metrics, and impact upon client populations that should be examined in future research as well.

Additionally, with regard to the sample of participants from this study, the perspectives from frontline workers (not in an administrative role) must be further examined due to their unique perspectives and proximity to direct client experiences. Organizations often measure what is required rather than what is perceived by frontline workers to be useful (Perez Jolles et al., 2017). If frontline workers do not believe that the performance indicators are useful or valid, the overall quality and robustness of the required data may be limited. Findings from research in this
area would likely provide critical information about the quality of current data collection practices. They should also provide additional information regarding the influence of funder-mandated metrics on definitions of client success and a better understanding of the important relationship between a client and their provider.

Future research would also benefit from the continued examination of studying the adequacy of current metrics as indicators of success, the return on social investments, and the impact on an organization’s theory of change (Pennerstorfer & Rutherford, 2019). Continued research in these areas can provide additional insight into their influence upon service delivery, agency functioning, and client experiences. In addition, this research area may identify ways to reconsider and revise data collection processes to address historic challenges that agencies often face. Finally, there is also limited research exploring perceptions of the purpose of evaluation. For example, additional research is needed to determine if agencies are primarily collecting information because they feel obligated to do so (Campbell & Lambright, 2017), if the evaluation process assists with agency growth and success, and how to successfully address some of the barriers identified in this study, such as resources, technology, time, and capacity issues.

In addition to researching ways to address barriers with regard to resources and capacity, future research should also examine how agency staff and administrators navigate and address the overlapping evaluation needs of multiple funders. When nonprofits receive money from multiple funders, they must develop a complex set of organizational goals and subjective assessments of nonprofit effectiveness based on a variety of external priorities. Combining all of these evaluation requirements may limit an agency’s ability to shape programming and to define
what constitutes success. It may also deter nonprofit leaders from gathering more pertinent information that could be used to influence decision making by policymakers or other community leaders (Benjamin, 2008; Campbell & Lambright, 2016; Cumberland et al., 2015; Ebrahim, 2019; Willems et al., 2014).

There is also a need for ongoing research to examine how much performance measurement metrics impact organizational effectiveness and efficiency. Obtaining a better understanding of how evaluation processes and requirements influence agency staff at different levels (e.g., volunteers, direct service roles, supervisors, administrators, board members) could aid in supporting organizations examining the relationships between organizational capacity and the adoption of outcome measurements.

Moreover, there is limited research that examines various aspects of evaluation requirements from the perspective of funders. Future research should seek to explore the meaning of transparency and accountability in the context of funder evaluation requirements. Identifying what constitutes transparency or accountability in relation to program effectiveness is complex and under-researched (Kim et al., 2019; Lynch-Cerullo & Cooney, 2011; Moura et al., 2019; Treinta et al., 2020). Understanding this relationship can provide information associated with power dynamics formed between nonprofit organizations and funders. In addition, future research should examine funders’ perspectives on metrics and identify if current practices and expectations truly provide the necessary information funders seek.

Future research should consider exploring the implications of funder-mandated metrics and standards on client definitions of success. In addition, research in this area should examine how socially constructed norms framed by White, cisgender, heteronormative perspectives are
reflected in performance evaluation and how this may impact client outcomes and specific minority populations with unique culture-bound norms.

There is also a need to examine the relationship between funder-mandated metrics, standardized evaluation systems, and descriptions of positive client change. Understanding how individualized and tailored client treatment plans can result in positive client outcomes is an under-researched area requiring further attention.

This study identified a need for research exploring measurements that examine the broader social impact of service provision in addition to the current focus on individual client change. In addition to current forms of data collection standards, research should focus on what metrics and data could improve the understanding of the enduring effects of racism, educational inequality, disinvestment in poor communities, and other systems of discrimination and oppression (Mosley et al., 2019). Therefore, ongoing research, scholarship, and advocacy should be directed toward improving the quality of a client’s overall circumstances within a nonprofit setting, their own community, and society at large.

Finally, research is needed to examine alternative approaches to current assessment practices, identifying their overall usefulness and exploring the different types of outcomes they may provide. For example, little is known about trust-based philanthropy in relation to performance metrics, an approach that focuses on redistributing power between foundations and nonprofits, providing multiyear unrestricted funding, and developing metrics in a mutual learning environment (Colinvaux, 2018; Easterling et al., 2019; Easterling & Main, 2016). Similarly, an organizational learning approach focuses on the organizational context in which the interventions or services are implemented, increasing agency learning and taking systematic
approaches to best understand how to build the most robust programming (Campbell & Lambright, 2017; Carman, 2011; Lynch-Cerullo & Cooney, 2011; Mosley et al., 2019). These approaches may also be more in line with the mission of the social work profession and the promotion of social justice. Research in these areas or other alternative approaches to metrics could provide important insight into how they inform service provision within the scope of diverse client experiences and organizational realities while void of the constraints of required external metrics.

Limitations

There are several limitations to the current study. First, the sample was limited by utilizing purposive sampling, even while this was the most effective method in working with a smaller sample size and selecting the most extensive variety of respondents. With a limited number of participants, the purposive approach allowed the selection of agencies and individuals that provided unique and rich information (Suen et al., 2014). The sampling method permitted an extensive representation of the data within the inclusion criteria identified for the study (Charmaz, 2014; Schurch, 2015). The approach was the most feasible regarding accessibility to agencies, affordability, and effective use of resources. It also allowed for intentional sampling of nonprofit organizational leadership. However, since the sampling method selected participants within the population required by the study, it is possible the sample did not fully represent a comprehensive understanding of staff and administrator perspectives on the influence of mandated metrics on service provision.

In addition, I was only able to interview participants once, which removed the opportunity for prolonged engagement and observation of participants (Lincoln & Guba, 1985;
Padgett, 2008). This has the potential to impact the credibility of the study. However, I did have participants engage in the member checking process to review my findings, share their feedback, and ensure their voices and perspectives were accurately conveyed.

Additionally, during the course of this research study in early 2020, the COVID-19 global health pandemic propelled the world into an unprecedented health crisis. This unanticipated event resulted in a mandatory stay-at-home order, except for essential personnel, and altered the normality of everyday life for the researcher and study participants alike. Thus, the global pandemic altered the current research study in several ways.

The mandated stay-at-home order resulted in the rescheduling of 15 participant interviews. Initially, 18 in-person interviews were scheduled in a six-week period in March and April 2020. The rescheduling process occurred over six months, with two participants withdrawing from the study. Ideally, all the data collection from participants would have occurred in the same six-week time span, but this was not feasible. The time between data collection allowed some participants to reflect on the research topic within the context of COVID-19, while the initial participants were interviewed prior to the pandemic. This may have altered participant perspectives or experiences when discussing the research topic. To enhance the credibility of the interview process, I employed reframing and expansion of questions during each interview. These actions supported the interviews’ internal consistency (Krefting, 1991).

The member checking technique was also used after the completion of data analysis in an effort to support internal consistency. This strategy included providing all participants with the opportunity to review the findings, ensuring that I accurately translated the viewpoints, and decreased chances of misrepresentation. Emails were sent to all 16 participants asking them to
review the findings and engage in the member checking process. Four participants responded and reviewed the content, sharing that they had found no discrepancies, felt their contributions to be represented, did not convey any concerns, and stated the findings aligned with their experiences. Although these responses aligned with the presented findings, it is a limitation that almost half of the participants were no longer employed with the organization and that only a quarter of participants reviewed the content.

Next, the sampling inclusion criteria were modified. Initially, nonprofit organizations with a budget range between $100,000 to $1 million were included in the study. These smaller nonprofit agencies had limited capacity to participate in the study due to increased demands to pivot to online service provision during the pandemic or due to budget cuts and the temporary closure of programming or the agency. Therefore, it was determined that this budget category should be eliminated, therefore resulting in two budget categories rather than three. Including these additional budget categories could yield different areas of influence, and additional research is warranted in this area. After these changes were made, the study included participants from eight nonprofit organizations. Four agencies had budgets between $1 and $4.99 million, and the other four had budgets greater than $5 million. The resulting 16 participants could be considered a comparatively small sample size and potentially impact the transferability of the findings. However, due to the maximum variation sampling approach, the resulting participants and nonprofit organizations still provided a range and breadth of experiences. Additionally, sampling saturation did occur during data analysis when no new codes or categories emerged from the data. There was also a detailed audit trail and dense descriptions, which provided increased credibility of the study by presenting accurate information of the participant’s
experience that other individuals who also shared the experience could recognize in the descriptions (Krefting, 1991; Lincoln & Guba, 1985).

During the transition to a stay-at-home order and shifting to working from home, the interview process was altered from an in-person format to phone or online via Zoom. The primary advantage of in-person interviews is the rapport developed between the participant and interviewer (Creswell, 2014; Rubin & Babbie, 2017). This relationship has the potential to increase the credibility of the data by reducing response bias, the ability to tailor the order of the questions and ask follow-up questions to clarify the participants’ responses. Overall, interviews conducted via Zoom provided an enhanced interface to engage with participants in a similar way to being in person, along with the added benefit of convenience and ease of use. In addition, both Zoom and phone interviews provided the necessary and safest approach to conducting interviews during the global pandemic. Zoom was preferred to phone interviews and served as a ‘stand-in’ for face-to-face interviews for all but one of the rescheduled participants. Both approaches ultimately provided more flexibility for the participants, scheduling times that best suited their work and family obligations and schedules while working from home. However, this change in format had the potential to impact how participants may have engaged in the interview and responded to the interview questions.

Last, these participants found themselves in new work environments, struggling to balance family responsibilities and address a series of changing safety guidelines and mandates within their nonprofit. Many of these organizations were operating in crisis response mode, and the interview participants were facing new and unprecedented challenges in the organization’s day-to-day functioning. As the COVID-19 global pandemic continued to impact the participants’
lives in such significant ways, it should be noted that it is difficult to discern how this experience influenced their involvement in the study. Moreover, the pandemic continued throughout the entire dissertation process, making it difficult to determine the potential impact of this historical event on the overall results and the transferability of the study.

As the researcher, I was faced with similar challenges during the global pandemic, moving to a remote work format for school and through my employer and managing family obligations. These experiences likely influenced the ways in which I navigated the research process. The data analysis and the emergent theory were created through the lens of my experiences, interpretation of participant interviews, and previous knowledge of the content. Others may have interpreted the data differently, making it imperative to create detailed information regarding the entire process. In keeping with a constructivist approach to grounded theory, I engaged in reflexivity practice during the entire study. Reflexive journaling, peer debriefing, and the construction of an audit trail were used to assess the influence of my own perceptions with regard to the study and became an additional way to reflect upon my perceptions of the impact of the COVID-19 global health pandemic on the research process. Ultimately, these tools guided the data meaning making process and resulted in dense descriptions of the process, procedures, participants, and findings (Charmaz, 2008, 2014; Hussein et al., 2014). In addition, the dense background information about the participants, research settings, and content are available to allow others to assess the transferability of the findings (Krefting, 1991).
Conclusion

Nonprofit organizations provide essential services that address complex, multifaceted social issues to diverse client populations. As part of the nonprofit funding process, organizations receive financial support from various external sources, such as foundations, government contracts, or corporate sponsorship. These financial contributions require nonprofit agencies to engage in performance evaluation in an effort to measure program progress and convey accountability and transparency to each funding entity. Often, the evaluation process is driven by funder priorities and the metrics they require agencies to collect data on. Therefore, as agencies continue to navigate these requirements, understanding nonprofit staff perceptions on the influence of funder-mandated metrics upon service provisions becomes more relevant. These perceptions provide a better understanding of how these metrics impact different areas of agency functioning and the clients. This research project focused on increasing the knowledge regarding staff perceptions of funder-mandated metrics on service provision and contributes to a larger body of literature that identifies the strengths and limitations of the mainstream performance evaluation approach.

Findings from this study suggest that funder-mandated metrics influenced nonprofit organizations at the client, program, and agency levels. Overall, mandated metrics do fulfill the intended purpose of evaluation systems. The evaluation data can support learning, inform decision making, guide service modifications, and inform strategic planning or budgeting processes. However, the findings also suggest that the influence of funder-mandated metrics can create additional challenges, such as a lack of consideration for client content and culture, increased staff time, limited resources to cultivate employee evaluation skill sets, and limited
resources, such as technology needs. As a result, these challenges contribute to additional work that staff must navigate. Notably, findings suggest that several outcomes unintentionally influence definitions of client success, the client-provider relationship, and client experiences. Funder-mandated metrics requirements can impose subjective and restrictive definitions of success on clients. This raises questions about how accurate metrics are regarding actual client progress. These metrics can also influence the relationship between direct service providers and clients, influencing the scope of services and time allocated to working with clients. Finally, the influence of metrics can be directly experienced by clients, who often felt that they have less autonomy over their personal goals and decision-making. Continued research in these areas can examine the relevance of performance evaluation to agencies and funders, as well as explore if the metrics actually link to program success and client goal attainment.

This research also raises several important questions regarding the social justice implications of funder-mandated metrics that warrant further exploration. There is a clear need for funders to take direct action towards building more evaluation capacity among nonprofits, as well as reevaluate metrics requirements of agencies, particularly with concern around what metrics are normalized, cultural-bound, or privileged. Within social work practice, agency leadership needs to examine their internal ability to cultivate more ownership and responsibility around program evaluation, exploring the role the client-provider relationship holds, as well as the impact unexamined metrics have on their clients and towards greater social change. Finally, within social work education, there is a potential to expand the dialogue around metrics through interprofessional education opportunities, as well as an evaluation of curricula to be more inclusive of micro, mezzo, and macro perspectives within course work.
APPENDIX A

EMAIL COMMUNICATION TO POTENTIAL PARTICIPANTS
Dear Colleagues,

I hope this email finds you well. I am Ph.D. Candidate in the School of Social Work at Loyola University Chicago and am seeking individuals to offer their valuable opinions in an important study. The primary purpose of this study is to examine the perceptions of nonprofit administrators and managers within nonprofit organizations regarding the influence of funder mandated performance metrics on service provision. My hope is that this study will provide insight on how funder mandated performance metric systems have on agency programming and client services. You are being asked to participate because you are an employee of a nonprofit organization that receives private or public funding to support programming and agency services.

To participate in this study, you must currently work at a Chicago-based nonprofit (501c3) organization that has an annual budget of $1 million or greater, receive funding from more than one primary funding source, provide human services programming, run at least 4 programs, and be either an administrator or manager that works directly with program measurement and evaluation.

As a participant in this study, you will be asked questions pertaining to your agency’s programming, how you measure program goals or outcomes, and the performance measurement tools used to do so. The in-person interview will take approximately 1-2 hours to complete. Participating in the study include contributing to a better understanding of funder mandated metrics on nonprofit services provision. It also may provide insight into the relationship dynamics between funders and nonprofits, identifying ways nonprofits have been successful or unsuccessful in these relationships.
This research has been reviewed and approved by the Research Ethics Board of Loyola University Chicago (Project #2921). If you have any questions, concerns, or comments about your rights as a participant, please contact the Office of Research Services at: 773-508-2689. I would welcome the opportunity to meet with you and interview you. If you are interested please contact me, Melissa Iverson, at miverso@luc.edu to schedule a time and location that is convenient for you. Thank you for your time and consideration.

Sincerely,
Melissa Iverson, MA, MSW
PhD Candidate
School of Social Work
Loyola University Chicago
APPENDIX B

CONSENT TO PARTICIPATE IN RESEARCH
Project Name: An Examination of Perceptions of Funder Mandated Performance Measurement in the Nonprofit Sector

Researcher: Melissa Iverson, MSW, MA

Faculty Sponsor: Michael Dentato Ph.D., MSW

Introduction: You are being asked to take part in a research study being conducted by, Melissa Iverson, Ph.D. Candidate in the School of Social Work at Loyola University Chicago.

Purpose: The purpose of this study is to gather information about nonprofit employees’ perceptions regarding funder mandated performance metrics influence on service provision. Participants for this study will consist of employees in administrative and program management positions that work on measuring program outcomes. The study examines what perceptions participants have regarding the influence of funder mandated performance measurement on service provision.

Procedures: If you agree to be in the study, you will be invited to participate in:

▪ An in-person interview that will be conducted by the researcher in a private space at your place of work or location you determine to be confidential.
▪ An interview that will be audio recorded.
▪ An interview that is expected to last between 60 to 120 minutes in total.
▪ Reviewing the study findings to ensure that your experiences are represented in the interpretation of the data.

Risks/Benefits:
Participation in the interview has no perceived risks involved in participating beyond those experienced in everyday life.

There are no direct benefits to participants, other than participating in the study which could contribute to a better understanding of funder mandated metrics influence on nonprofit services provision. However, the information provided by participants may provide insight into the relationships between funders and nonprofits. Additionally, the research may provide insight into the relationship dynamics between funders and nonprofits, identifying ways nonprofits have been successful or unsuccessful in these relationships.

Confidentiality:
Following the completion of the interviews, the audio recordings will be used to transcribe the conversations by the researcher. In order to protect participant confidentiality, the researcher will use participant pseudonyms during the transcription process. Audio files, transcripts and related computer data files will be stored on password protected computer and will not contain names or other identifying information and will be assigned an identification number. Audio files will be
destroyed immediately following transcription. Transcripts and related computer data files will be kept for seven years following the date of the final interview as is the requirement of the LUC IRB.

**Voluntary Participation:**
Participation in this study is voluntary. If you do not want to participate in this study, you do not have to do so. Even if you decide to participate, you are free not to answer any question or to withdraw from participation at any time without penalty.

**Compensation:**
There is no compensation for participating in this study, however your responses will help provide insight regarding how funder mandated performance metric systems have on agency programming and client services.

**Contact and Questions:**
If you have any questions about this research study, please feel free to contact Melissa Iverson, MSW, MA, doctoral student in the School of Social Work at Loyola University Chicago at miverso@luc.edu. The dissertation chair is Dr. Michael Dentato, who can be contacted at mdentato@luc.edu.

If you have questions about your rights as a research participant, you may contact the Loyola University Chicago, Office of Research Services at 773-508-2689.

**Statement of Consent:**
Do you agree to participate in this study and to have it audio recorded?

___ No, I decline to participate in this study.

___ Yes, I agree to participate in this study and to have the interview recorded.

Your signature below indicates that you have read the information provided above, have had an opportunity to ask questions, and agree to participate in this research study.

Participant’s Signature
Date

Researcher’s Signature
Date

Faculty Sponsor Signature
Date
APPENDIX C

INTERVIEW PROTOCOL AND QUESTIONS
Introduction to the study:
My name is Melissa Iverson, and I am a doctoral student in the School of Social Work at Loyola University Chicago. The purpose of this study is to examine the perceptions of nonprofit administrators and managers within nonprofit organizations regarding funder mandated performance metrics influence on service provision. The following interview will be comprised of questions that focus on 4 areas including your professional background, information about the agency’s programming and perceptions of funder mandated performance measurements the influence on service provision. The interview will take 1-2 hours to complete. You can stop the interview at any time. Do you have any questions before we begin?

Opening questions:
How did you come into your current role?
What do you do in your role that involves program evaluation?
   Probe: Do you work on goals setting? In what capacity?
   Probe: Do you work on measurement? In what capacity?
   Probe: Do you work on program evaluation? In what capacity?
   Probe: Do you write part or all of reports to funders?

The next set of questions will pertain to information about the agency’s programming and funder mandated performance measurements:
Can you tell me about your agency’s services/programs?
Can you tell me about the funding your agency receives that requires you to use measurement tools provided by the funder?
Can you talk about the measurement tools your agency uses and how these are related to services that are provided?
   Probe: Can you talk about why your agency uses internally developed and funder required measurement tools?
   Probe: Can you talk about how your agency collects data to meet the needs of multiple funders?
How do you think these measurement tools impact how services are provided?

Probe: Can you talk about how the required metrics impact efforts to increase program quality? Please elaborate or provide specific examples.

Probe: Can you talk about how required metrics impact efforts to increase program effectiveness? Please elaborate or provide specific examples.

Probe: Can you talk about how required metrics impact the mission of your program? Please elaborate or provide specific examples.

Probe: Can you talk about how required metrics impact clients? Please elaborate or provide specific examples.

Probe: Can you talk about how metrics impact definitions of client or program success? Please elaborate or provide specific examples.

What areas in your programs do you think are overlooked when current performance metrics are used?

Probe: If you could create additional metrics to help your organization to evaluate and improve current services, what would they be?

Probe: If you could eliminate certain metrics because they are not helpful to evaluating and improving quality or effectiveness of service, what would those be? From where do the metrics you identified originate?

This now completes the interview process for this study. Do you have any questions about the topics we discussed? Are there other things that you would like to discuss at this time that you think would help me understand what we have been discussing?

Thank you for participating in this study. If you have additional questions I can be reached at miverso@luc.edu.
APPENDIX D

DEMOGRAPHICS QUESTIONS
1. What is the annual operating budget for the agency?
   - $100,000 to $249,999
   - $250,000 to $999,999
   - $1 million to $4.99 million
   - $5 million or more
2. How many programs does the nonprofit have? ____________
3. How many people does the nonprofit serve annually? ________
4. How many years has the agency been in business? ____________
5. How many board members does the agency have? ___________
6. How many staff does the agency have? __________

Participant information:
1. What is your current position at your agency? ________________________________
2. How long have you been in your position? _________________________________
3. What is your age? _________________
4. Please specify your gender. ________________
   - Male
   - Female
   - Other (Please fill in the blank): ________________
5. Please specify your sexual orientation.
   - Heterosexual
   - Gay
   - Lesbian
   - Bisexual
   - Other (Please fill in the blank): ________________
6. Please specify your ethnicity. (Check all that apply)

☐ White
☐ Hispanic or Latino
☐ Black or African American
☐ Native American or American Indian
☐ Asian
☐ Native Hawaiian/Other Pacific Islander
☐ Other (Please fill in the blank): ________________

7. Please specify your highest level of completed education:

☐ High School
☐ Bachelor’s
☐ Master’s
☐ Doctoral
☐ Other (please specify): ________________
APPENDIX E

EMAIL COMMUNICATION TO PARTICIPANTS FOR MEMBER CHECK PROCESS
Dear NAME OF PARTICIPANT,

I hope this email finds you well. Thank you for your time and input into my research study examining the perceptions of nonprofit administrators and managers within nonprofit organizations regarding the influence of funder mandated performance metrics on service provision. As part of the research study, each participant has the opportunity to review the study findings to ensure that your experiences are represented in the interpretation of the data. It is an opportunity to provide feedback regarding the study and its findings.

I am writing today to let you know that the data has been collected and analyzed and to offer the opportunity to review the findings. If you would like to participate, please respond to this email, and I will provide the findings for your review. You can provide any feedback via email, over the phone or we can arrange a time to meet.

Thank you again for your time and input regarding this study.

Sincerely,

Melissa Iverson, MA, MSW
PhD Candidate
School of Social Work
Loyola University Chicago
REFERENCE LIST


Publications.


Frayne, D. (2014). *Nonprofit leader perceptions of effective organizational performance measurement: A Q methodology study* [ProQuest Dissertations and Theses database]. UMI No. 3648297


McKeever, B.S., (2015, October) The Nonprofit sector in brief 2015: Public charities, giving and
volunteering. (Brief) Retrieved from
https://www.urban.org/sites/default/files/publication/72536/2000497-The-Nonprofit-


Mitchell, G. E. (2013). The construct of organizational effectiveness: Perspectives from leaders of


performance measurement and management in non-profit organizations* (No. 18). https://doi.org/10.2139/ssrn.402880

http://ej.issuelab.org/resources/1458/1458.pdf

human service organizations, and what social work can do to fix it. *Human Service

Moura, L. F., Pinheiro de Lima, E., Deschamps, F., Van Aken, E., Gouvea da Costa, S. E.,
nonprofit and public administration organizations. In *International Journal of
Group Holdings Ltd. https://doi.org/10.1108/IJPPM-06-2018-0236

Moxham, C. (2009). Performance measurement: Examining the applicability of the existing body of
knowledge to nonprofit organisations. *International Journal of Operations and
Production Management, 29*(7), 740–763. https://doi.org/10.1108/01443570910971405

literature review. *International Journal of Productivity and Performance Management,*


VITA

Dr. Melissa Mae Iverson was born on February 21, 1983 in La Crosse, Wisconsin. She graduated from Westby High School, Westby, Wisconsin, in 2001. She earned her Bachelor of Arts in Psychology from Winona State University, Winona, Minnesota, in 2001. In 2007, Dr. Iverson received a Master of Arts in Gender Studies and Master of Social Work from Loyola University Chicago, Chicago, Illinois. After graduating, she worked as a social work administrator for several nonprofit organizations in Chicago. Currently, Dr. Iverson is an Instructor at Northeastern Illinois University in the Social Work Program. She lives in Chicago, Illinois.