Photo by Johnny Killroy on Unsplash

JPNA
JOURNAL OF PUBLIC AND NONPROFIT AFFAIRS

Research Articles

JPNA
JOURNAL OF PUBLIC AND NONPROFIT AFFAIRS
a journal of the Midwest Public Affairs Conference

Statement of Purpose
The *Journal of Public and Nonprofit Affairs* (*JPNA*) focuses on providing a connection between the practice and research of public affairs. This is accomplished with scholarly research, practical applications of the research, and no fees for publishing or journal access. *JPNA* publishes research from diverse theoretical, methodological, and disciplinary backgrounds that address topics related to the affairs and management of public and nonprofit organizations.

JPNA
JOURNAL OF PUBLIC AND NONPROFIT AFFAIRS

a journal of the Midwest Public Affairs Conference

Spring 2025   Volume 11, Issue 1

Research Articles

Editor's Introduction

Christopher R. Prentice – University of North Carolina Wilmington
Richard M. Clerkin – University of North Carolina Wilmington

This issue of the *Journal of Public and Nonprofit Affairs* highlights recent research that advances understanding in nonprofit and public sector studies. Each article addresses pressing issues for nonprofit leaders, practitioners, and policymakers, with attention to themes such as efficiency, resilience, evaluation, social innovation, and workforce challenges. Together, these works reflect the diversity of approaches and insights shaping the public service today.

The discussion begins with an examination of efficiency, a central yet often misunderstood concept in nonprofit management. Berrett (2025) investigates how leaders from Habitat for Humanity affiliates across the United States assess organizational efficiency. Through interviews with 36 practitioners, the study finds that while affiliates employ a wide range of metrics, many fall short of accurately capturing efficiency. The results point to the need for stronger education on measurement practices and a multidimensional approach that better aligns scholarly frameworks with practitioner realities.

Building on this theme of organizational capacity, De Oro (2025) explores resilience within the nonprofit sector. Using a systematic review guided by Cochrane-Campbell protocols, the study identifies three recurring themes in the literature: disturbances to organizational systems, leadership and management strategies, and financial resilience. The review underscores the importance of developing holistic frameworks that integrate people, structures, and relationships, particularly for smaller nonprofits that may struggle to fully recover yet demonstrate adaptability and survival.

Complementing these organizational perspectives, Berghmans and Vandenabeele (2025) examine how evaluation practices influence the transformative potential of social innovation. They argue that dominant monitoring and evaluation methods, based on simple or complicated intervention logics, often limit transformative outcomes. Instead, approaches grounded in complex intervention logics are more compatible with systemic change. The authors call for cultivating new evaluation habits and propose an agenda for further action and research.

Turning to a specific national context, Alaimo, Pejcal, and Smrčková (2025) investigate the evaluation capacity of Czech nonprofit social service organizations. While these organizations are often required to evaluate their programs by national and EU entities, many struggle with evaluation capacity building (ECB). Using surveys and interviews with directors, the study finds that challenges resemble those seen in other countries but are also shaped by directors' interpretations of government guidelines and their limited understanding of ECB. The authors conclude with practical recommendations for improving evaluation practices.

Finally, questions of workforce performance and sectoral differences are addressed by Bednarczuk (2025), who analyzes work absence patterns across public and private sectors in the United States. Drawing on data from the National Health Interview Study, the findings reveal that public sector employees generally report worse physical health, but better mental health compared to private sector workers. While health alone did not explain absence gaps,

decomposition analysis showed that demographic and organizational differences largely account for the variation. These results highlight the role of sector-level characteristics in shaping workforce outcomes.

Together, these articles expand understanding of critical issues facing nonprofit and public organizations. From measurement and evaluation to resilience and workforce dynamics, the research emphasizes the need for frameworks and practices that reflect the complexity of organizational life and support more effective management and policy.

References

Alaimo, S. P., Pejcal, J., & Smrčková, S. (2025). The Evaluation Capacity of Czech Nonprofit Social Services Organizations. *Journal of Public and Nonprofit Affairs, 11*(1), 67 – 95. https://doi.org/10.20899/jpna.jeabf764

Bednarczuk, M. (2025). Health and the Work Absence Gap Across Employment Sectors in the United States. *Journal of Public and Nonprofit Affairs, 11*(1), 96 – 120. https://doi.org/10.20899/jpna.8xaca558

Berrett, J. L.. (2025). Exploring Efficiency Metrics in Nonprofits: A Case Study of Habitat for Humanity. *Journal of Public and Nonprofit Affairs, 11*(1), 5 – 27. https://doi.org/10.20899/jpna.g0199p91

Berghmans, M. & Vandenabeele, J. (2025). Evaluation and Transformative Social Innovation: A Call for New Habits. *Journal of Public and Nonprofit Affairs, 11*(1), 48 – 66. https://doi.org/10.20899/jpna.w8sydp51

De Oro, G. C.. (2025). The Role and Relevance of Resilience in the Nonprofit Sector: A systematic review of the literature. *Journal of Public and Nonprofit Affairs, 11*(1), 28 – 47. https://doi.org/10.20899/jpna.ysdjay56

Acknowledgement

Exploring Efficiency Metrics in Nonprofits: A Case Study of Habitat for Humanity

Jessica L. Berrett - University of Colorado Colorado Springs

This exploratory case study examines nonprofit efficiency by interviewing 36 leaders from Habitat for Humanity affiliates across diverse geographic locations in the United States, focusing on understanding efficiency from a practitioner's perspective. Habitat for Humanity, composed of numerous nonprofits dedicated to housing, provided a rich context for this investigation. The study finds that while nonprofit practitioners use a diverse array of metrics to assess efficiency, many of these metrics do not accurately measure it, indicating a need for greater education around efficiency measurement. Additionally, the metrics cover diverse organizational areas, highlighting the importance of a multidimensional approach to efficiency assessment. The study suggests that aligning scholarly methods with practitioner needs and promoting a broader understanding of efficiency across multiple organizational dimensions could improve both theoretical and practical applications in nonprofit management.

Keywords: nonprofit efficiency, normative measures, instrumental measures, Habitat for Humanity

Introduction

Nonprofit organizations are increasingly confronted with the challenge of optimizing their managerial efficiency—the ability to turn inputs into outputs (Coupet & Berrett, 2018). The pursuit of doing more with less has become a central theme in nonprofit scholarship and discourse (AbouAssi et al., 2016; Alexander, 2000; Bishop, 2007; Coupet & Berrett, 2018; Hackler & Saxton, 2007; Ridder et al., 2012; Ritchie et al., 2004). These organizations must navigate a complex landscape of diminishing government support, intensified competition for scarce resources, and heightened expectations from funders, who increasingly demand concrete, demonstrable outcomes while insisting on cost containment (Alexander, 2000). As a result, nonprofits are subject to heightened scrutiny and accountability (Bishop, 2007; Ritchie et al., 2004).

Efficiency metrics, which evaluate how effectively an organization converts resources into results, play a crucial role in this environment. These metrics are used both internally, for performance management and strategic planning, and externally, for public communication and benchmarking against other organizations (Poister, 2008). External benchmarking is often emphasized by watchdog organizations like Charity Navigator, Charity Watch, and BBB Wise Giving Alliance. For instance, Charity Navigator (2024) assesses metrics such as the program

expense ratio and fundraising efficiency, while Charity Watch (2024) uses similar measures. BBB Wise Giving Alliance (2024) also evaluates program and fundraising expense ratios, whereas GuideStar by Candid (2024) is moving away from traditional metrics, offering new alternatives to reduce the emphasis on overhead ratios.

Nonprofits face significant challenges in assessing efficiency, due to a lack of clarity on what metrics to measure, insufficient data collection practices, and limited capacity to analyze and interpret data. Often, organizations struggle to define appropriate indicators of efficiency, unsure of whether to focus on financial metrics, program outcomes, or operational processes. This uncertainty is compounded by the fact that many nonprofits do not consistently collect the necessary data to evaluate their performance, either because they lack the resources or because data collection is not integrated into their regular operations. Additionally, even when data is available, nonprofits may not have the expertise or tools to analyze it effectively, making it difficult to draw meaningful insights or make informed decisions. These challenges create a significant barrier to understanding and improving efficiency, leaving many nonprofits unable to demonstrate their impact or optimize their use of resources.

Despite the importance of these metrics, there has been limited discussion between scholars and nonprofit practitioners on this topic. This paper aims to bridge that gap by providing clarification and insight into how efficiency is understood and measured by those working in the field. Guided by the research question "What efficiency metrics are employed by nonprofit practitioners?" this study conducts an exploratory case study on nonprofit efficiency, featuring interviews with executive directors and board members from various Habitat for Humanity affiliates across the United States. The perspective of practitioners is crucial, as highlighted by Mosley et al. (2019) in their critique of the "What Works" movement. They argue that evidence-based practice, while aimed at ensuring service quality through standardized models, often overlooks the contextual expertise of frontline workers and the unique needs of diverse communities. They advocate for an organizational learning approach that values adaptability and integrates multiple forms of knowledge, which can better support effective practice in complex social environments (Mosley et al., 2019). The goal of this study was to understand efficiency from the perspective of practitioners. Habitat for Humanity, as a prominent organization within the human service subsector—the largest nonprofit subsector in the U.S. (McKeever et al., 2016)—offers a useful sample.

The relevance of this study is highlighted by the current landscape in which nonprofits operate. The COVID-19 pandemic and subsequent economic disruptions have further strained resources and increased the demand for services (Kim & Mason, 2020; Newby & Branyon, 2021; Shi, 2022), making efficient use of resources more critical than ever. Nonprofits are under immense pressure to adapt and find innovative ways to achieve their missions with limited resources. Therefore, understanding and improving efficiency is not just an academic interest but a practical requirement for organizational survival and impact.

Moreover, this study offers practical insights for nonprofit leaders and managers. By highlighting the limitations of current efficiency measures and proposing alternative approaches, it aims to equip practitioners with the tools needed to navigate the complex landscape of nonprofit management. The findings suggest that a multidimensional approach can provide a more accurate understanding of efficiency, ultimately helping organizations to better achieve their missions in a resource-constrained environment. In sum, this study is timely and essential as it addresses the critical need for a better understanding of nonprofit efficiency in today's challenging environment. It connects practical application with academic theory, offering valuable contributions to both

fields and providing nonprofit organizations with the insights needed to enhance their operational effectiveness.

Literature Review

Upon examining the prior research that assesses efficiency measures within the nonprofit domain, the literature review has unveiled a multitude of efficiency metrics employed (see Table 1). These metrics can be effectively categorized into two overarching categories that align with Mitchell's (2018) dual perspectives on efficiency. One view, termed normative, accentuates the importance of overhead reduction, while the other perspective, deemed instrumental, underscores the imperative of cost minimization per unit of impact. However, scholarly literature has begun to expose the complexities and limitations inherent in some of these metrics.

Integrating institutional theory further enriches this discussion by exploring how nonprofit organizations navigate and reconcile these normative and instrumental logics within their institutional environments. Institutional theory proposes that organizations are shaped by external pressures from their institutional environment, including regulatory requirements, professional standards, and cultural expectations. These pressures can lead to isomorphism, where organizations in the same field become increasingly similar in their structures and practices (DiMaggio & Powell, 1983).

In practice, nonprofit managers often operate within a complex interplay of normative and instrumental logics. For example, a nonprofit may adhere to ethical standards (normative) while also employing performance metrics (instrumental) to demonstrate accountability and effectiveness to funders. Balancing these logics requires managerial acumen to ensure that the organization maintains its legitimacy and stakeholder trust while achieving its mission efficiently.

Table 1. Selective Review on the Efficiency Measures from the Academic Literature

Instrumental Efficiency Measures	Studies
Administrative expenses + fundraising expenses + special event expenses / total revenue	Ecer et al. (2017); Hager et al. (2001)
Cost per dollar of receipts x technical efficiency index	Callen & Falk (1993)
Data envelopment analysis (DEA)	Ayayi & Wijesiri (2018); Ba et al. (2022); Berrett & Hung (2023); Burgess & Wilson (1995); Callen & Falk (1993); Coupet (2018); Coupet & Berrett (2018); Coupet et al. (2021); González-Torre et al. (2017); Luksetich & Hughes (1997); Min & Ahn (2017); Miragaia et al. (2016); Roh et al. (2010)
Fundraising + administrative costs / total revenue	Bowman (2006)
Fundraising expenses / contributions	Ashley & Faulk (2010); Frumkin & Keating (2011)

Fundraising expenses + special event expenses / total contributions + gross special event income	Ecer et al. (2017); Hager et al., (2001)
Program expenses / number of participants	Hung & Berrett (2023)
Project expenses / total revenue	Rocha Valencia et al. (2015)
Stochastic frontier analysis (SFA)	Bishop & Brand (2003); Coupet & Berrett (2018); Hung & Berrett (2022)

Normative Efficiency Measures	
Administrative expenses / total expenses	Ashley & Faulk (2010); Callen et al. (2003); Chikoto & Neely (2014); Coupet & Berrett (2018); Ecer et al. (2017); Frumkin & Keating (2011); Frumkin & Kim (2001); Tinkelman & Mankaney (2007); Trussel & Parsons (2007); Lu & Zhao (2019)
Administrative + fundraising expenses / total expenses	Lecy & Searing (2015); Kim (2017); Coupet & Berrett (2018)
Fundraising expenses / total expenses	Callen et al. (2003); Chikoto & Neely (2014); Frumkin & Keating (2011)
Program expense / total expense	Trussel & Parsons (2007)
Total expenses / program expenses	Callen et al. (2003)

Note. The list is meant to be representative but not exhaustive. Also, see Appendix A for a list of the input and output measures used in each study's DEA and SFA analysis.

Instrumental Measures

The instrumental logic, drawing inspiration from March and Olsen's (1998) theory of consequences, centers on the notion that individuals make choices by carefully evaluating the likely outcomes concerning their personal or collective objectives. Within the instrumental perspective, managerial decisions are not influenced by adherence to a predetermined set of normative rules; instead, they are driven by a deliberate assessment of the potential consequences associated with each decision.

The managerial viewpoint is marked by its inward focus, a clear orientation toward specific objectives, and a strong emphasis on achieving impact (Berrett & Sudweeks, 2023). Under the instrumental framework, efficiency is construed as a ratio juxtaposing costs against outputs or impact. In nonprofit organizations, this instrumental orientation toward efficiency entails pursuing organizational goals by amplifying social impact while concurrently curbing or maintaining costs (Berrett & Sudweeks, 2023). Within the nonprofit literature, instrumental measures manifest through three distinct avenues.

Data Envelopment Analysis: Data envelopment analysis (DEA) is a method used to evaluate how efficiently different organizations use their resources (like money, time, or staff) to produce results (like how effectively shelters use their budgets to provide housing for the homeless or how

effectively counseling centers use their staff to improve client well-being). It is a way to compare the performance of different organizations doing similar work to see who is getting the best results with the resources they have. It helps identify who is most efficient and where others might improve. Essentially, the application of DEA is instrumental in determining an organization's efficiency relative to its peers, yielding a scoring metric ranging from 0 to 1. This score is derived by maximizing the summation of input-to-output ratios for each organization, thereby generating an efficiency score (Charnes et al., 1978, 1981). An organization achieving a score of 1 is deemed the most efficient, effectively establishing an efficiency frontier. Subsequently, peer organizations are assessed and assigned efficiency scores, reflecting their proximity to the most efficient entity or entities.

For instance, in a study conducted by Coupet (2018) investigating the impact of government funding on efficiency in nonprofit and public colleges, DEA is employed as the initial step. Each college's efficiency is calculated, considering inputs such as instruction, academic support, and student services, compared with the output metric of graduation rates. Subsequently, a regression analysis employs the efficiency score as the dependent variable of interest, with public funding as the independent variable. The findings of this study reveal that public funding experts have no discernable influence on the efficiency of nonprofit colleges, but conversely have a negative effect on the efficiency of public colleges (Coupet, 2018).

In another study, DEA is harnessed to gauge the efficiency of food banks across 13 European countries, as undertaken by González-Torre et al. (2017). This analytical framework leverages inputs including the foundation year, number of volunteer staff, and number of permanent staff. Outputs encompass metrics such as amount of food distributed and number of recipients of food. The outcomes of this analysis facilitate identifying the most and least efficient food banks, thereby affording the means to establish benchmarks for enhancement.

Stochastic Frontier Analysis: Akin to the application of DEA, certain scholars within the nonprofit domain employ stochastic frontier analysis (SFA) as a method to measure efficiency, as exemplified by the works of Bishop and Brand (2003), Coupet and Berrett (2018), and Hung and Berrett (2022). As explained by Kumbhakar et al. (2015), SFA entails utilizing econometric models to estimate production, cost, or profit frontiers, then assessing efficiency relative to these established frontiers. For instance, in a study by Bishop and Brand (2003), the researchers probe into the nexus of public funding, volunteer engagement, and efficiency. Efficiency is achieved by evaluating total running and maintenance costs incurred by museums during the study year, with the number of full-time equivalent workers within each museum constituting the inputs. Outputs are captured by the number of physical visits and subscribers garnered by each museum. Through a regression analysis incorporating public funding and volunteer activity as independent variables, the study discerns a negative impact on efficiency.

In another study, Hung and Berrett (2022) examine the relationship between commercialization and nonprofit efficiency with the moderating roles of government funding and organizational size in nonprofit arts organizations. They measure efficiency using stochastic frontier analysis with the number of participants as the output, and total program expenses, total administrative expenses, and the number of full-time equivalent employees as the inputs. Regressing commercialization on nonprofit efficiency, they find a positive correlation. Moreover, they find the relationship is more positive when less government funding is received.

Simple Input-to-Output Ratios: Efficiency measures can also include simple input-to-out ratios. For instance, some scholars focus on fundraising efficiency by calculating the ratio of fundraising expenses to contributions (Ashley & Faulk, 2010; Frumkin & Keating, 2011; Hager et al., 2001).

Ashley and Faulk (2010) found that organizations with higher fundraising expense ratios tend to receive lower grant amounts. On the other hand, in a large sample of nonprofits across the U.S., Frumkin and Keating (2011) discovered that revenue concentration enhances efficiency. Other researchers have examined revenue efficiency by analyzing different expense-to-revenue ratios (Bowman, 2006; Ecer et al., 2017; Hager et al., 2001; Rocha Valencia et al., 2015). For example, Hager et al. (2001) used the ratio of overhead expenses to total organizational revenues. They found that the efficiency levels vary based on factors like organizational size, age, and subsector. Ecer et al. (2017) also used this ratio and observed that organizations relying primarily on commercial revenues more efficiently manage overhead costs. Additionally, some scholars have focused on program efficiency. For example, Hung and Berrett (2023) used a ratio of program expenses to the number of participants in a sample of arts nonprofits, finding that efficiency does not mediate the relationship between commercialization and free access in nonprofits.

Normative Measures

The normative managerial framework draws upon March and Olsen's (1996, 1998) concept of appropriateness logic. In this paradigm, managerial actions are guided by established norms and rules that stem from the organization's institutionalized identity or role. This institutionalization process gives rise to a set of norms and regulations that dictate what constitutes appropriate behavior within the organization, or in other words, socially acceptable standards. These decisions and actions are largely influenced by the need to uphold these institutionally defined standards of conduct rather than being solely grounded in rational expectations (Berrett & Sudweeks, 2023). This perspective posits that a socially or normatively accepted mode of behavior exists for nonprofit organizations, and managers are tasked with making choices that align with and perpetuate this accepted conduct.

Expense Ratios: The normative measures of efficiency are epitomized by the array of expense ratios, a category encompassing the overhead, administrative, fundraising, and program ratios, as substantiated by a corpus of scholarly works (Ashley & Faulk, 2010; Bowman, 2006; Callen et al., 2003; Chikoto & Neely, 2014; Coupet & Berrett, 2018; Ecer et al., 2017; Frumkin & Keating, 2011; Frumkin & Kim, 2001; Kim, 2017; Lecy & Searing, 2015; Lu & Zhao, 2019; Tinkelman & Mankaney, 2007; Trussel & Parsons, 2007). Expense ratios are considered normative because they are grounded in the concept of appropriateness logic, which dictates that managerial actions should align with established norms and rules within an organization (March & Olsen, 1996, 1998). These norms, derived from the organization's institutional identity, define what is considered acceptable behavior, such as reducing overhead costs. Expense ratios, like overhead, administrative, and fundraising ratios, are commonly used as proxies for efficiency because they are easily accessible, simple to compute, and allow for comparisons across organizations. However, they focus on input-to-input ratios rather than measuring outputs, which limits their ability to fully capture organizational efficiency (Coupet & Berrett, 2018). Despite these limitations, expense ratios are widely used because they reflect socially accepted standards within the nonprofit sector.

For those who employ expense ratios as their metric of choice, the reduction of overhead is construed as a critical marker of efficiency. For example, an investigation conducted by Ashley & Faulk (2010) hinged upon the administrative and fundraising cost ratios, seeking to unravel the impact of efficiency on grant acquisition. Their findings unveiled an inverse relationship, wherein organizations exhibiting higher fundraising ratios tended to secure less grant revenue. Furthermore, a study by Frumkin and Keating (2011) delved into revenue concentration, establishing a linkage with heightened efficiency. This investigation employed the administrative and fundraising expense ratios as proxies for efficiency, effectively demonstrating that revenue

concentration was positively correlated with enhanced efficiency. However, caution must be applied to these findings for the reasons explained below.

Challenges With the Normative Measures: The evaluation of nonprofit efficiency is fraught with challenges, but the increasing pressure on these organizations to demonstrate efficiency emphasizes its importance for scholars and practitioners alike. A key concern in the discourse is the concept of construct validity, which refers to the extent to which a metric accurately represents the phenomenon it is intended to measure while also aligning with the established theoretical frameworks (Garson, 2016).

Several scholars have cautioned against overreliance on expense ratios as a means of comparing nonprofit organizations. Steinberg and Morris (2010) argue that an excessive focus on these ratios can lead to unintended and harmful consequences, such as increased compliance costs, the spread of misleading solicitations, donor misguidance, inefficient fundraising efforts, and the suboptimal delivery of charitable outputs. They point out that high fundraising expenses do not necessarily indicate fraud or misallocation of resources (Steinberg & Morris, 2010).

Bowman (2006) further critiques the utility of overhead ratios, noting their limitations in the comparative analysis of organizations. Research by Coupet and Berrett (2018), using the same Habitat for Humanity data as this study, supports the conclusion that the expense ratios do not validly measure efficiency. The crux of the issue lies in the distinction between input-to-output ratios, which genuinely assess efficiency, and input-to-input ratios, like expense ratios, which do not account for outputs. For example, the administrative expense ratio measures administrative costs as a percentage of total expenses, without considering the outcomes produced by these expenditures. This lack of output consideration undermines the construct validity of expense ratios as a measure of efficiency.

To substantiate their argument, Coupet and Berrett (2018) conducted an analysis comparing expense ratios with direct efficiency measures, using stochastic frontier analysis (SFA) and data envelopment analysis (DEA), both of which are well-regarded methods in the management sciences for evaluating efficiency (Lampe & Hilgers, 2015; Ruggiero, 1996). Their findings showed a statistically significant relationship between SFA and DEA efficiency scores, while showing a negative correlation between these scores and expense ratios (Coupet & Berrett, 2018). Based on these results, Coupet and Berrett (2018) advocate for the use of metrics more closely aligned with true efficiency, such as DEA and SFA, highlighting the construct validity issues inherent to using overhead ratios as proxies for efficiency.

Prentice (2016) also emphasizes the importance of construct validity in nonprofit financial assessments, particularly when complex constructs are oversimplified into single metrics or when different metrics are used interchangeably. Mitchell and Calabrese (2019) further critique the use of overhead as a proxy for efficiency, arguing that it fails to serve as a credible measure of organizational effectiveness and efficiency due to its lack of construct validity.

Takeaways From the Academic Literature

The review of academic literature on nonprofit efficiency reveals several key trends and insights. Scholars have predominately focused on financial efficiency, with a particular emphasis on overhead ratios as a proxy for efficiency. However, it is noteworthy that the most recent study using overhead ratios dates back to 2019, indicating a potential shift away from this metric. This shift is likely driven by increasing caution among scholars regarding the limitations and potential misuse of overhead ratios.

In place of these traditional metrics, there is a growing trend toward the application of more sophisticated analytics methods such as data envelopment analysis and stochastic frontier analysis. Although these methods are not entirely new, their increased usage reflects a broader move toward more accurate assessments of nonprofit efficiency. These approaches allow for a more comprehensive understanding by evaluating the relationship between inputs and outputs, rather than relying solely on financial metrics.

Despite these advancements, the literature consistently highlights that there is no single, definitive way to measure nonprofit efficiency. The complexity and diversity of the sector necessitates a multidimensional approach that takes into account various factors beyond financial performance alone. This academic perspective sets the stage for exploring how nonprofit practitioners themselves approach the measurement of efficiency. The following methods section dives into the research design, data collection, and analysis processes.

Method

Research Design Overview

This research employs an exploratory case study methodology, a method developed by Stake (1995), to gain deeper insight into the measures used by nonprofit practitioners in assessing efficiency. While applying this research question beyond a single case would be valuable, a case study approach was chosen to ensure the study's manageability, particularly given its exploratory nature and the use of interviews. The exploratory design facilitates the identification of emerging patterns, themes, and unanticipated insights, making it well suited to address the research question. Focusing on Habitat for Humanity affiliates allows for an exploration of the practical applications and theoretical constructs of nonprofit efficiency, ensuring that the findings are both relevant and applicable to academic and practitioner audiences. In the context of this research, the study focuses on 36 in-depth, semi-structured interviews conducted with executive directors and board members from various Habitat for Humanity affiliates across the U.S.

Habitat for Humanity International is dedicated to bringing people together "to build homes, communities, and hope" (Habitat for Humanity, 2023). Its mission focuses on building homes and providing affordable homeownership. The organization is hierarchical and decentralized, encompassing several levels of governance and operational units, including global headquarters, regional offices, national organizations, and local affiliates. This study focuses on local affiliates based in the U.S. The selection of this particular sample is supported by the shared operational characteristics and output parameters exhibited by these Habitat for Humanity affiliates, thereby facilitating an evaluation of efficiency within each affiliate.

Study Participants

Thirty-one executive directors and five board chairs from Habitat for Humanity affiliates, representing 19 states,[i] participated in the interviews. Regarding racial background, two individuals (5.56%) identified as Black, while the majority, comprising 34 individuals (94.44%), identified as white. Gender distribution among the interviewees consisted of 16 females (44.44%) and 20 males (55.56%). The nonprofit leaders' backgrounds, prior to assuming leadership roles within Habitat for Humanity, were diverse. Specifically, five leaders (13.89%) had prior experience in the public sector, nine (25%) hailed from the nonprofit sector, eight (22.22%)

possessed a for-profit sector background, and 14 (38.89%) brought a blend of experience spanning multiple sectors.

Regarding the organizations represented by these nonprofit leaders, 15 organizations (41.67%) were classified as highly efficient, nine (25%) as moderately efficient, and 12 (33.33%) as having low efficiency. Additionally, when examining organizational size, six entities (16.67%) were categorized as small, 11 (30.56%) as medium-sized, and 19 (52.78%) as large organizations. Please see Table 2 for a summary of the descriptive statistics.

Table 2. Descriptive Statistics

Category	Subcategory	Details	Percentage
Participants	Role	31 Executive Directors	86.11%
		5 Board Members	13.89%
	Racial Background	2 Black	5.56%
		34 White	94.44%
	Gender Distribution	16 Females	44.44%
		20 Males	55.56%
	Prior Experience	5 Public Sector	13.89%
		9 Nonprofit Sector	25.00%
		9 For-Profit Sector	22.22%
		14 Multiple Sectors	38.89%
Organizations Represented	Efficiency Classification	15 Highly Efficient	41.67%
		9 Moderately Efficient	25.00%
		12 Low Efficiency	33.33%
	Organizational Size	6 Small	16.67%
		11 Medium	30.56%
		19 Large	52.78%

Data Collection

Semi-structured interview protocols were crafted and tested among Habitat executive directors and board members, and the necessary approval was received from the Institutional Review Board. These interviews were conducted from November 2019 to January 2020 via Zoom. While the interviewees responded to a comprehensive set of 15 questions as part of a larger project, the interview questions specific to this study included the following: Do you measure efficiency in your organization? If so, what metrics or measures do you use? Although the interviews ranged from 30 to 60 minutes on average, the questions pertinent to this study typically occupied approximately 2 to 4 minutes of the interviewees' time.

Analysis

The qualitative data analysis was conducted utilizing NVivo12, employing a qualitative content analysis approach. The analytical process for the qualitative data involved the following steps: (1) initial transcript preparation for coding, followed by a thorough review of each transcript to gain familiarity with the data; (2) a second reading of the transcripts, during which text segments were systematically coded; (3) aggregation of similar codes to eliminate redundancy and refine the coding structure; (4) utilization of codes to identify underlying themes and the formulation of theme passage; (5) development of a conceptual map to organize and connect these themes; and (6) crafting a coherent narrative that weaves together all of the identified themes to address the primary research question, in line with the approach outlined by Creswell and Creswell Báez (2021).

The initial coding process involved collaboration between the primary interviewer and a graduate student, jointly creating the codebook and delineating the preliminary themes. To enhance interrater reliability, a third round of coding was conducted with the involvement of an additional graduate student, achieving an interrater reliability score of 0.81. In cases of coding discrepancies, discussion was held until a consensus was reached. Subsequently, the emergent themes underwent extensive analysis and discussion among the research team.

In qualitative research, the assessment of the findings' validity or accuracy is conducted through a multifaceted lens, considering the perspectives of the researcher, participants, readers, and reviewers (Creswell & Creswell Báez, 2021). The researcher, in particular, undertakes a process of reflexivity, contemplating their experiences and backgrounds, and what potential impact these elements could exert on interpreting the data. A search for disconfirming evidence is also employed, which requires establishing themes followed by searching for evidence for exceptions to these themes. Subsequently, the integration of participants' viewpoints is essential. This is realized through collaborative efforts, exemplified by the engagement of seven Habitat leaders in pilot interviews. This participatory aspect enriches the research process by incorporating the unique perspectives of those Habitat leaders directly involved. Furthermore, a commitment to presenting thick, rich descriptions is important in catering to the discernment of readers and reviewers. This commitment involves crafting and articulating detailed contextual information, facilitating a nuanced and comprehensive understanding of the research outcomes.

Results

The results are organized into two sections. The first focuses on thematic analysis based on types of measures, and the second section focuses on thematic analysis based on organizational areas.

Thematic Analysis Based on Types of Measures

In response to the inquiry regarding the evaluation of efficiency, a majority of participants, comprising 69.44% (25 out of 36 respondents), affirmed that they engaged in efficiency measurement practices. Conversely, 30.56% (11 out of 36 respondents) indicated that they do not incorporate efficiency metrics into their assessments.

Among those respondents who acknowledged measuring efficiency, an examination of the metrics they employ was conducted, as detailed in Table 3. The findings revealed various metrics, totaling 34 distinct measures collectively. Noteworthy is the observation that only seven of these metrics constituted direct efficiency measures, explicitly denoting ratios of inputs to outputs. The remaining metrics encompassed expense ratios and input or output measures, reflective of the data requisites and collection practices intrinsic to Habitat International.

Input-to-Output Ratios: The input-to-output ratios provided are instrumental measures of efficiency and are key indicators of efficiency in construction and project management. Several of these ratios are particularly important in the pursuit of efficiency and productivity. First, there is a strong emphasis on completing tasks as quickly as possible. This focus on speed is closely linked to the time required to build a house, with minimizing the construction duration being a top priority. As one respondent noted, "In construction, my director of construction, in particular, has a goal around the speed at which a house gets built. We aim to complete all our houses within 16 weeks or less from the time the roof is raised" (Interview #18). However, the need for speed must

be balanced with cost considerations. Another respondent highlighted the importance of cost efficiency:

> We tend to focus heavily on the cost of the unit produced. There's a question of how many units you can produce in a certain amount of time using mostly volunteer labor. It is going to specifically limit exactly how much construction you can get done in any given timeframe. But it also tends to help lower the price of the unit, and what you get is any time you make a decision that increases the price of the unit, you have to increase the amount of income that the end user is going to need in order to afford the unit. (Interview #31)

Table 3. Thematic analysis based on types of measures.

Themes	Codes
Input-to-output ratios	# of staff per house built - Cost of goods sold - Cost per house - Getting things done as fast as possible - Staffing ratios - Time per project - Time to build a house
Input-to-input ratios	Expense ratios
Inputs	# of applications approved # of applications received # of families applying # of families inquiring # of people who attend orientation meetings # of volunteer hours # of volunteers - Amount of land - Amount of cash accumulated - Bills paid - Board giving - Board member engagement - Cost of operations - Cost savings - Donations - Employees paid - Time - Value of a dollar donated
Outputs	# of families served

- # of houses built
- # of houses sold
- # of mortgages
- # of mortgages overdue vs. current
- # of people connected with
- Accessibility
- How much is saved from the landfill

This approach also involves a careful analysis of the costs of goods sold, ensuring that each unit produced is financially viable. As one respondent explained:

> We measure efficiency by closely monitoring our budgets, but cash flow is even more critical. When evaluating the budget, we compare the cost of goods sold to the home costs and the sales price. We calculate the cost of goods sold divided by the number of homes built. (Interview #27)

Understanding staffing ratios is also essential for achieving efficiency in production. Determining the optimal number of staff members needed to build a single house is crucial for effective resource management. Staffing ratios, which consider the total time spent on projects and the corresponding financial allocation, provide a comprehensive view of workforce optimization. As one respondent shared, "We measure efficiency primarily through financial metrics. We assess staffing ratios and the cost of building a home, including how specific positions are allocated across divisions. These financial metrics are key to defining and evaluating efficiency" (Interview #22).

Input-to-Input Ratios: The input-to-input ratios provided reflect the normative measures of efficiency. Effectively managing financial resources is paramount in any organization, and for some organizations, input-to-input ratios play a crucial role in understanding and optimizing expenditures. A key consideration is how much is spent on homes versus operations, as striking a balance between these two categories is essential for financial sustainability. One respondent highlighted this balance:

> We measure efficiency in two different ways. We have the Department of Agriculture, which has a charity checker. They look at the 990s and financials, and they also assess how much we spend on homes versus operations. We also use Charity Navigator, which does the same thing. People rely on both of these, and it helps increase donations. (Interview #5)

Functional expenses, which encompass various operational costs, contribute to the overall financial landscape. For many organizations, analyzing and optimizing these functional expenses is key to efficient resource allocation. As one respondent emphasized,

> We measure efficiency in terms of financial management and administration. We look at functional expenses mainly because they're easily identifiable and come out in the audit, such as overhead ratios. We look at where our funder's donations are spent, and in every decision, we ask if it's a good use of funds—whether it increases our capacity to do more and if we can serve more families as a result. (Interview #6)

Expense ratios provide valuable insights into the proportion of resources dedicated to overhead costs. Overhead, including indirect costs such as administrative expenses, must be carefully managed to prevent unnecessary financial strain. Balancing the ratios ensures that a reasonable portion of the budget is allocated to essential operational elements without excessively burdening the organization with overhead costs. As one respondent pointed out, "We like to look at how

much of our budget is allocated to administrative overhead versus program expenditures" (Interview #23).

Inputs: Many nonprofits rely on various input measures as indicators of their efficiency. These inputs represent the resources required to execute their programs. For example, some organizations focus on their volunteers. As one respondent shared,

> Well, when I started here, we had nothing. We had no way of measuring anything. So, the reports that we have to turn in to Habitat International, the statistical reports, and that sort of thing, have really forced us and encouraged us to start tracking things like volunteer hours. And we've just recently signed on with a company called Network for Good to track our donors and our online donations. So, we're just starting to learn how to use that. (Interview #29)

Other organizations emphasized board participation. As one respondent noted, "We have metrics that measure board participation; for instance, we strive for 100% board giving and require board members to attend at least three home dedications each year" (Interview #7). This active engagement demonstrates board members' personal investment in the organization's mission.

Some organizations take a more comprehensive approach, tracking not only the number of volunteers but also donations and homeownership inquiries. One respondent explained, "We're constantly evaluating our number of volunteers, our donations, and how many families are applying" (Interview #17). Another added,

> We track our family services by monitoring the number of inquiries for homeownership, the number of people who attend our orientation meetings, how many applications are received, how many are approved, and we also track our land inventory throughout this process. (Interview #30)

Outputs: Nonprofits also utilize a variety of output measures as indicators of their efficiency. Central to these measures is the number of houses built, which lies at the heart of the organization's mission. This metric represents the tangible result of the organization's efforts to provide housing solutions to those in need. As one respondent noted, "An obvious one is the number of houses—we primarily build new houses and do some full rehabs" (Interview #11). This metric, combined with the number of families served, expands the narrative beyond just structures. As another respondent stated, "We have a number of measures—how many houses we're building, how many families we serve" (Interview #13). Closely related is the number of homes sold. For example, one respondent mentioned, "This year, our metric happens to be to sell 70 homes in [city] and 10 in [another city]" (Interview #7).

Another important measure is reflected in the number of mortgages facilitated. As one respondent explained,

> There's the official and unofficial method. Officially, what is required of Habitat International in the quarterly production reports, such as how many houses, how many mortgages, and how many mortgages are overdue vs. current? Unofficially, are we meeting the needs of the families and community? (Interview #2)

Beyond housing and financial metrics, nonprofits also measure their environmental impact. One respondent highlighted the organization's commitment to sustainability, saying, "How much do we save from the landfill?" (Interview #19).

Thematic Analysis Based on Organizational Areas

In addition to categorizing the measures based on their types, an alternative approach involves organizing them according to specific organizational areas, as illustrated in Table 4. Notably, it was observed that the efficiency measures provided did not encompass an overarching evaluation of overall efficiency; instead, they were more granular and specific to distinct organizational areas. The three areas that emerge as themes include finances and resources, management and human resources, and programs.

Finances and Resources: The financial health and resource management of a nonprofit organization are reflected in several key metrics. The amount of cash accumulated and cash flow provide insight into the organization's liquidity and its ability to sustain operations over time. Board giving demonstrates the commitment of leadership to financially support the mission, which can inspire confidence in other donors. The timely payments of bills and employees indicate effective financial management and operational stability. Cost of operations and cost savings are critical in evaluating how efficiently the organization utilizes its resources, while expense ratios help assess the balance between spending on programs versus overhead. Additionally, tracking donations and understanding the value of each dollar donated highlight the organization's fundraising effectiveness and its ability to maximize donor contributions toward achieving its mission.

Management and Human Resources: A nonprofit organization's management and human resources are reflected in several key metrics. The number of staff per house built and volunteer hours highlight the organization's ability to effectively utilize both paid staff and volunteers to achieve its goals. The total number of volunteers and the engagement of board members indicate the level of community involvement and leadership commitment, which are vital for sustaining the nonprofit's mission. Prioritizing efficiency, as seen in the focus on getting things done as quickly as possible, and monitoring staffing ratios, are essential for optimizing workforce productivity. Metrics like time per project and overall project duration provide insights into the organization's ability to manage its time effectively, ensuring that projects are completed on schedule and resources are used efficiently.

Programs: A nonprofit's programs can be reflected in a variety of metrics. The number of applications received and approved, and the number of families applying, inquiring, and served are key indicators of the program's reach and ability to meet community needs. The number of houses built and sold, along with the number of mortgages facilitated and their status (overdue vs. current), illustrate the tangible outcomes of the organization's housing efforts, while the number of people attending orientation meetings speaks to the program's ability to engage and educate potential beneficiaries. Accessibility is important in ensuring that these programs reach those in need, and the amount of land available directly impacts the organization's capacity to expand its services. Financial metrics, such as the cost of goods sold and cost per house, highlight the program's efficiency in resource utilization. Additionally, tracking how much is saved from the landfill highlights the environmental impact of the program, while the time it takes to build a house reflects operational efficiency. Finally, tracking the number of people connected with through these programs ensures ongoing engagement and support, reinforcing the organization's commitment to its mission and community.

Table 4. Thematic analysis based on organizational areas

Themes	Codes
Finances and resources	• Amount of cash accumulated • Bills paid • Board giving • Cash flow

	• Cost of operations • Cost savings • Employees paid • Expense ratios • Donations • Value of a dollar donated
Management and human resources	• # of staff/house built • # of volunteer hours • # of volunteers • Board member engagement • Getting things done as fast as possible • Staffing ratios • Time per project • Time
Programs	• # of applications approved • # of applications received • # of families applying • # of families served • # of families inquiring • # of houses built • # of houses sold • # of mortgages • # of mortgages overdue vs. current • # of people connected with • # of people who attend orientation meetings • Accessibility • Amount of land • Cost of goods sold • Cost per house • How much is saved from the landfill • Time to build a house

Discussion

This study found that nonprofit practitioners utilize a wide variety of metrics to measure efficiency. While some practitioners focus on instrumental measures, such as input-to-output ratios, others continue to rely on normative measures like expense ratios or even individual input or output measures. The efficiency measures were also categorized according to organizational areas: finances and resources, management and human resources, and programs. This categorization highlights that practitioners consider efficiency across multiple dimensions of their operations, demonstrating the need for a multidimensional approach to fully assess an organization's efficiency.

The findings show both similarities and differences when compared to scholarly approaches. In terms of similarities, both nonprofit practitioners and scholars recognize that no single metric can fully capture nonprofit efficiency. Both groups use a diverse array of measures, with scholars and practitioners both employing normative and instrumental measures.

There are also notable differences, however. Scholars have increasingly moved away from traditional metrics like expense ratios, moving toward sophisticated methods such as DEA and SFA (for example, see Ayayi & Wijesiri, 2018; Ba et al., 2022; Coupet & Berrett, 2018; Coupet et al., 2021). By contrast, the study results indicate that nonprofit practitioners are less likely to use these advanced methods, instead relying on simpler metrics and ratios. This gap highlights a divergence in the tools and methodologies used by academics and practitioners to assess efficiency.

Furthermore, while academic research has traditionally focused heavily on financial metrics—likely due to the availability of financial data—the study results suggest that nonprofit practitioners place significant emphasis on other areas, such as human resources and programs. This indicates that practitioners may adopt a broader perspective on efficiency, considering various aspects of their operations beyond just financial performance.

Implications

First, nonprofit practitioners need to learn what efficiency is and how it is best measured. Scholarly work has highlighted the problems with using the expense ratios as a measure of efficiency (Bowman, 2006; Coupet & Berrett, 2018; Steinberg & Morris, 2010) and this is something that needs to continue to be communicated to nonprofit practitioners. Furthermore, the study highlights the importance of evaluating efficiency across multiple dimensions. Practitioners should consider developing and using a comprehensive set of efficiency metrics that capture the full scope of their organization's operations rather than relying on a single measure or focusing solely on financial metrics. Additionally, while more advanced methods such as DEA and SFA may not be realistic for a nonprofit practitioner to use, nonprofits are capturing individual input and output measures and can learn to combine these into ratios to measure efficiency (refer to Appendix A). Lastly, the divergence between scholarly methods and practitioner approaches highlights the need for better alignment and knowledge sharing. As the field evolves, nonprofit practitioners should remain open to refining their efficiency metrics and approaches. By periodically reviewing and updating their methods in line with both practical experiences and academic advancements, they can ensure that their organizations are effectively measuring and improving efficiency across all areas of operation.

Scholars should consider broadening their research to include metrics that reflect human resources and program outcomes, as practitioners emphasize these areas alongside financial efficiency. This shift could lead to a more comprehensive understanding of nonprofit efficiency. Additionally, the divergence in the tools and methodologies used by scholars and practitioners suggests a need for greater alignment. Scholars should explore ways to make advanced methods like DEA and SFA more accessible and applicable to practitioners, potentially developing practical guidelines or tools that translate complex methods into usable strategies for nonprofits. Scholars should promote the use of multidimensional efficiency assessments in their research, recognizing that no single metric can fully capture nonprofit efficiency. By advocating for a more holistic approach, scholars can help shape a better understanding of efficiency in the nonprofit sector. Lastly, scholars might benefit from increased collaboration with nonprofit practitioners to ensure that academic research addresses real-world challenges and that theoretical advancements are

informed by practice experience. This engagement could foster a two-way exchange of knowledge, enhancing both research and practice in the field.

Furthermore, a balanced, integrated multidimensional metric and related approach may also be useful to multi- or cross-sector collaborations involving foundations, other nonprofit organizations, and sectors. Such an approach can facilitate a more comprehensive evaluation and improve coordination and resource allocation among different entities working toward common goals. Exploring how shared beliefs or values relate to preferred outcomes can provide deeper insights into the alignment of organizational goals and the effectiveness of these multidimensional metrics.

Lastly, a multidimensional approach to measuring nonprofit efficiency should incorporate diverse perspectives, including those of people with lived experience and cross-disciplinary experts. Engaging people with lived experience can provide valuable insights into the practical implications of efficiency measures. Cross-disciplinary expertise can help develop, implement, and monitor more holistic and contextually relevant metrics. The paper suggests that while this inclusive approach is broadly beneficial across various nonprofit domains, its importance may vary based on the organization's context and mission. For example, housing nonprofits like Habitat for Humanity can significantly benefit from the insights of community members who directly experience housing challenges. In domains where direct service delivery and community impact are central, such as human services and healthcare, involving people with lived experience can lead to more relevant and impactful efficiency measures. Conversely, for more administrative or advocacy-focused nonprofits, cross-disciplinary expertise might play a more prominent role.

Limitations and Future Research

This study has several limitations that should be considered when interpreting the findings. First, the focus on Habitat for Humanity, an organization with specific characteristics—such as a significant reliance on volunteers and a mission centered on housing as a fundamental human right—limits the generalizability of the results to the broader nonprofit sector. Second, the study employed a case study methodology and relied solely on interviews for data collection, which may also limit the generalizability. Third, the potential for response bias exists, as participants may have provided socially desirable answers, affecting the integrity of the data. Fourth, selection bias is a concern, as the sample of 36 affiliates was drawn from a larger population, which may not fully represent the diversity within Habitat for Humanity. Lastly, the exploratory nature of the study offers preliminary insights without engaging in confirmatory analysis, necessitating further research to substantiate and refine the findings.

To address the limitations identified in this study, future research should consider several avenues to enhance the generalizability and robustness of findings. Expanding the scope beyond Habitat for Humanity to include a more diverse range of nonprofit organizations will help determine whether the identified efficiency metrics apply broadly across the sector. Additionally, integrating quantitative data and data from surveys with qualitative interviews could provide a more comprehensive understanding of nonprofit efficiency. Addressing potential response and selection biases will require larger, more representative samples and the use of mixed-method approaches. Finally, future studies should move beyond exploratory analysis to include confirmatory research that tests and validates the proposed efficiency measures, thereby solidifying their applicability and reliability across different nonprofit contexts.

Conclusion

This study offers insights into the complex landscape of nonprofit efficiency metrics, highlighting the divergence between academic approaches and practitioner practices. While academics are increasingly adopting sophisticated methods like DEA and SFA, practitioners often rely on more straightforward metrics. This research highlights the need for a multidimensional approach to efficiency that incorporates financial, human resource, and programmatic dimensions. Future efforts should aim to bridge the gap between theory and practice, ensuring that nonprofit organizations have access to both robust metrics and practical tools for assessing and enhancing efficiency.

Notes

i Those states are Alabama, Arkansas, California, Colorado, Florida, Georgia, Idaho, Indiana, Kentucky, Montana, Michigan, Nebraska, New York, North Carolina, Ohio, Oregon, Pennsylvania, Tennessee, and Texas.

Disclosure Statement

The author(s) declare that there are no conflicts of interest that relate to the research, authorship, or publication of this article.

References

AbouAssi, K., Makhlouf, N., & Whalen, P. (2016). NGOs' Resource Capacity Antecedents for Partnerships: Capacity as Antecedent for Partnerships. *Nonprofit Management and Leadership*, *26*(4), 435–451. https://doi.org/10.1002/nml.21200

Alexander, J. (2000). Adaptive Strategies of Nonprofit Human Service Organizations in an Era of Devolution and New Public Management. *Nonprofit Management and Leadership*, *10*(3), 287–303. https://doi.org/10.1002/nml.10305

Ashley, S., & Faulk, L. (2010). Nonprofit competition in the grants marketplace: Exploring the relationship between nonprofit financial ratios and grant amount. *Nonprofit Management and Leadership*, *21*(1), 43–57. https://doi.org/10.1002/nml.20011

Ayayi, A. G., & Wijesiri, M. (2018). Better with age? The relationship between longevity and efficiency dynamics of nonprofit microfinance institutions. *Quality & Quantity*, *52*(5), 2331–2343. https://doi.org/10.1007/s11135-017-0668-3

Ba, Y., Berrett, J. B., & Coupet, J. (2022). *Government attention and nonprofit efficiency: A link across boundaries*. Public Administration Review Virtual Symposium.

BBB Giving Wise Alliance. (2024). *BBB Standards for Charity Accountability*. Retrieved from: https://give.org/charity-landing-page/bbb-standards-for-charity-accountability

Berrett, J. B., & Hung, C. (2023). Does Revenue Concentration Really Bring Organizational Efficiency? Evidence From Habitat for Humanity. *Nonprofit and Voluntary Sector Quarterly*.

Berrett, J. L., & Sudweeks, J. (2023). Understanding nonprofit leaders' perceptions of organizational efficiency. *Nonprofit Management & Leadership*, *33*(3), 491–510. https://doi.org/10.1002/nml.21538

Bishop, P., & Brand, S. (2003). The efficiency of museums: A stochastic frontier production function approach. *Applied Economics*, *35*(17), 1853–1858. https://doi.org/10.1080/0003684032000158064

Bishop, S. W. (2007). Linking Nonprofit Capacity to Effectiveness in the New Public Management Era: The Case of Community Action Agencies. *State & Local Government Review*, *39*(3), 144–152. https://doi.org/10.1177/0160323X0703900303

Bowman, W. (2006). Should Donors Care About Overhead Costs? Do They Care? *Nonprofit and Voluntary Sector Quarterly*, *35*(2), 288–310. https://doi.org/10.1177/0899764006287219

Burgess, J. F., & Wilson, P. W. (1995). Decomposing Hospital Productivity Changes, 1985-1988: A Nonparametric Malmquist Approach. *Journal of Productivity Analysis*, *6*(4), 343–363. https://doi.org/10.1007/BF01073525

Callen, J. L., & Falk, H. (1993). Agency and Efficiency in Nonprofit Organizations: The Case of "Specific Health Focus" Charities. *The Accounting Review*, *68*(1), 48–65.

Callen, J. L., Klein, A., & Tinkelman, D. (2003). Board Composition, Committees, and Organizational Efficiency: The Case of Nonprofits. *Nonprofit and Voluntary Sector Quarterly*, *32*(4), 493–520. https://doi.org/10.1177/0899764003257462

Candid. (2024). *GuideStar Launches Tool for Nonprofits to Emphasize Their Progress and Results*. Retrieved from: https://candid.org/about/press-room/releases/guidestar-launches-tool-for-nonprofits-to-emphasize-their-progress-and-results

Charity Navigator. (2024). *How We Assess Charities' Transparency and Financial Responsibility*. Retrieved from: https://www.charitynavigator.org/about-us/our-methodology/ratings/accountability-finance/#accordion-33b1f4ab90-item-efea199a1c

Charity Watch. (2024). *Our Charity Rating Process*. Retrieved from: https://www.charitywatch.org/our-charity-rating-process#:~:text=NR%20%3D%20Not%20Ratable*-,CharityWatch%20considers%20a%20charity%20to%20be%20highly%20efficient%20when%20our,benchmarks%20for%20governance%20and%20transparency.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Charnes, A., Cooper, W. W., & Rhodes, E. (1981). Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through. *Management Science*, *27*(6), 668–697. https://doi.org/10.1287/mnsc.27.6.668

Chikoto, G. L., & Neely, D. G. (2014). Building Nonprofit Financial Capacity: The Impact of Revenue Concentration and Overhead Costs. *Nonprofit and Voluntary Sector Quarterly*, *43*(3), 570–588. https://doi.org/10.1177/0899764012474120

Coupet, J. (2018). Exploring the link between government funding and efficiency in nonprofit colleges. *Nonprofit Management and Leadership*, *29*(1), 65–81. https://doi.org/10.1002/nml.21309

Coupet, J., Berrett, J., Broussard, P., & Johnson, B. (2021). Nonprofit Benchmarking With Data Envelopment Analysis. *Nonprofit and Voluntary Sector Quarterly*, *50*(3), 647–661. https://doi.org/10.1177/0899764020977670

Coupet, J., & Berrett, J. L. (2018). Toward a valid approach to nonprofit efficiency measurement. *Nonprofit Management & Leadership*, *29*(3), 299–320. https://doi.org/10.1002/nml.21336

Creswell, J. W., & Creswell Báez, J. (2021). *30 essential skills for the qualitative researcher* (Second Edition). SAGE Publications, Inc. https://go.exlibris.link/vzwvY3gN

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review, 45*(2), 147-160. https://doi.org/10.2307/2095101

Ecer, S., Magro, M., & Sarpça, S. (2017). The Relationship Between Nonprofits' Revenue Composition and Their Economic-Financial Efficiency. *Nonprofit and Voluntary Sector Quarterly*, *46*(1), 141–155. https://doi.org/10.1177/0899764016649693

Frumkin, P., & Keating, E. K. (2011). Diversification Reconsidered: The Risks and Rewards of Revenue Concentration. *Journal of Social Entrepreneurship*, *2*(2), 151–164. https://doi.org/10.1080/19420676.2011.614630

Frumkin, P., & Kim, M. T. (2001). Strategic Positioning and the Financing of Nonprofit Organizations: Is Efficiency Rewarded in the Contributions Marketplace? *Public Administration Review*, *61*(3), 266–275. https://doi.org/10.1111/0033-3352.00029

Garson, G. D. (2016). *Validity and Reliability*. Statistical Associates Publishing.

González-Torre, P., Lozano, S., & Adenso-Díaz, B. (2017). Efficiency Analysis of the European Food Banks: Some Managerial Results. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *28*(2), 822–838. https://doi.org/10.1007/s11266-016-9814-0

Habitat for Humanity. (2023). https://www.habitat.org/about/mission-and-vision

Hackler, D., & Saxton, G. D. (2007). The Strategic Use of Information Technology by Nonprofit Organizations: Increasing Capacity and Untapped Potential. *Public Administration Review*, *67*(3), 474–487. https://doi.org/10.1111/j.1540-6210.2007.00730.x

Hager, M. A., Pollak, T., & Rooney, P. (2001). *Variations in Overhead and Fundraising Efficiency Measures: The Influence of Size, Age, and Subsector*. Annual Conference for Research on Nonprofit Organizations and Voluntary Action, Miami.

Hung, C., & Berrett, J. (2022). The effects of commercialization on nonprofit efficiency and the moderating roles of government funding and organizational size. *Nonprofit Management & Leadership*, *33*(4), 735–753. https://doi.org/10.1002/nml.21546

Hung, C., & Berrett, J. (2023). When Are Commercialized Nonprofits Less Likely to Offer Free Access? Evidence from the Performing Arts Subsector. *Public Performance & Management Review*, *46*(3), 644–665. https://doi.org/10.1080/15309576.2023.2168283

Kim, M. (2017). The Relationship of Nonprofits' Financial Health to Program Outcomes: Empirical Evidence From Nonprofit Arts Organizations. *Nonprofit and Voluntary Sector Quarterly*, *46*(3), 525–548. https://doi.org/10.1177/0899764016662914

Kim, M., & Mason, D. P. (2020). Are you ready: Financial management, operating reserves, and the immediate impact of COVID-19 on nonprofits. *Nonprofit and Voluntary Sector Quarterly*, *49*(6), 1191-1209.

Kumbhakar, S. C., Wang, H.-J., & Horncastle, A. P. (2015). *A Practitioner's Guide to Stochastic Frontier Analysis Using Stata*. Cambridge University Press.

Lampe, H. W., & Hilgers, D. (2015). Trajectories of efficiency measurement: A bibliometric analysis of DEA and SFA. *European Journal of Operational Research*, *240*(1), 1–21. https://doi.org/10.1016/j.ejor.2014.04.041

Lecy, J. D., & Searing, E. A. M. (2015). Anatomy of the Nonprofit Starvation Cycle: An Analysis of Falling Overhead Ratios in the Nonprofit Sector. *Nonprofit and Voluntary Sector Quarterly*, *44*(3), 539–563. https://doi.org/10.1177/0899764014527175

Lu, J., & Zhao, J. (2019). Does Government Funding Make Nonprofits Administratively Inefficient? Revisiting the Link. *Nonprofit and Voluntary Sector Quarterly*, *48*(6), 1143–1161. https://doi.org/10.1177/0899764019859435

Luksetich, W., & Hughes, P. N. (1997). Efficiency of Fund-Raising Activities: An Application of Data Envelopment Analysis. *Nonprofit and Voluntary Sector Quarterly*, *26*(1), 73–84. https://doi.org/10.1177/0899764097261007

March, J. G., & Olsen, J. P. (1996). Institutional Perspectives on Political Institutions. *Governance*, *9*(3), 247–264. https://doi.org/10.1111/j.1468-0491.1996.tb00242.x

March, J. G., & Olsen, J. P. (1998). The Institutional Dynamics of International Political Orders. *International Organization*, *52*(4), 943–969. https://doi.org/10.1162/002081898550699

McKeever, B. S., Dietz, N. E., & Fyffe, S. D. (2016). *The nonprofit almanac: The essential facts and figures for managers, researchers, and volunteers*. Rowman & Littlefield.

Min, H., & Ahn, Y. (2017). Dynamic Benchmarking of Mass Transit Systems in the United States Using Data Envelopment Analysis and the Malmquist Productivity Index. *Journal of Business Logistics*, *38*(1), 55–73. https://doi.org/10.1111/jbl.12148

Miragaia, D., Brito, M., & Ferreira, J. (2016). The Role of Stakeholders in the Efficiency of Nonprofit Sports Clubs. *Nonprofit Management & Leadership*, *27*(1), 113–134. https://doi.org/10.1002/nml.21210

Mitchell, G. E. (2018). Modalities of Managerialism: The "Double Bind" of Normative and Instrumental Nonprofit Management Imperatives. *Administration & Society*, *50*(7), 1037–1068. https://doi.org/10.1177/0095399716664832

Mitchell, G. E., & Calabrese, T. D. (2019). Proverbs of Nonprofit Financial Management. *The American Review of Public Administration*, *49*(6), 649–661. https://doi.org/10.1177/0275074018770458

Mosley, J. E., Marwell, N. P., & Ybarra, M. (2019). How the "What Works" Movement is Failing Human Service Organizations, What Social Work Can Do to Fit It, 43(4), 326-335.

Newby, K., & Branyon, B.(2021). Pivoting services: Resilience in the face of disruptions in nonprofit organizations caused by COVID-19. *Journal of Public and Nonprofit Affairs, 7*(3), 443–460. https://doi.org/10.20899/jpna.7.3.443–460

Poister, T. H. (2008). *Measuring performance in public and nonprofit organizations*. John Wiley & Sons.

Prentice, C. R. (2016). Understanding Nonprofit Financial Health: Exploring the Effects of Organizational and Environmental Variables. *Nonprofit and Voluntary Sector Quarterly*, *45*(5), 888–909. https://doi.org/10.1177/0899764015601243

Ridder, H.-G., Piening, E. P., & Baluch, A. M. (2012). The Third Way Reconfigured: How and Why Nonprofit Organizations are Shifting Their Human Resource Management. *Voluntas: International Journal of Voluntary and Nonprofit Organizations*, *23*(3), 605–635. https://doi.org/10.1007/s11266-011-9219-z

Ritchie, W. J., Anthony, W. P., & Rubens, A. J. (2004). Individual Executive Characteristics: Explaining the Divergence between Perceptual and Financial Measures in Nonprofit Organizations. *Journal of Business Ethics*, *53*(3), 267–281. https://doi.org/10.1023/B:BUSI.0000039417.15488.17

Rocha Valencia, L. A., Queiruga, D., & González-Benito, J. (2015). Relationship Between Transparency and Efficiency in the Allocation of Funds in Nongovernmental Development Organizations. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *26*(6), 2517–2535. https://doi.org/10.1007/s11266-014-9527-1

Roh, C. Y., Jae Moon, M., & Jung, C. (2010). Measuring Performance of U.S. Nonprofit Hospitals. *Public Performance & Management Review*, *34*(1), 22–37. https://doi.org/10.2753/PMR1530-9576340102

Ruggiero, J. (1996). On the measurement of technical efficiency in the public sector. *European Journal of Operational Research*, *90*(3), 553–565. https://doi.org/10.1016/0377-2217(94)00346-7

Shi, W. (2022). COVID-19 impacts on border community organizations. *Journal of Public and Nonprofit Affairs, 8*(3), 399–422.https://doi.org/10.20899/jpna.8.3.399–422

Stake, R. E. (1995). *The art of case study research*. Sage Publications. https://go.exlibris.link/fRTcLCB1

Steinberg, R., & Morris, D. (2010). Ratio discrimination in charity fundraising: The inappropriate use of cost ratios has harmful side-effects. *Voluntary Sector Review*, *1*(1), 77–95. https://doi.org/10.1332/204080510X497028

Tinkelman, D., & Mankaney, K. (2007). When is Administrative Efficiency Associated With Charitable Donations? *Nonprofit and Voluntary Sector Quarterly, 36*(1), 41–64. https://doi.org/10.1177/0899764006293176

Trussel, J. M., & Parsons, L. M. (2007). Financial Reporting Factors Affecting Donations to Charitable Organizations. *Advances in Accounting*, *23*, 263–285.

Author Biography

Jessica L. Berrett is an assistant professor in the College of Public Service at the University of Colorado Colorado Springs. Her specialty is nonprofit management, with a research focus on financial and performance managmenet.

Appendix A. DEA and SFA Input and Output Selection

Context	Input	Output	Example studies
Colleges (nonprofit and public)	Instruction, academic support, and student services	Graduation rate	Coupet (2018)
Food banks	Foundation year, number of volunteer staff, number of permanent staff	Amount of food distributed and number of recipients of food	González-Torre et al. (2017)
Hospitals (nonprofit)	# of full-time physicians and other health professionals, current assets of each hospital, # of hospital beds in each hospital	Total # of patient days, # of ER visits, # of outpatient visits, # of outpatient surgery visits, total amount of charity care	Roh et al. (2010)
Hospitals (nonprofit, for-profit, and public)	# of long-term hospital beds, registered nurses, licensed practical nurses, other clinical labor, nonclinical labor, and long-term care labor	Acute care inpatient days, case-mix weighted acute care inpatient discharges, long-term care inpatient days, number of outpatient visits, ambulatory surgical procedures, inpatient surgical procedures	Burgess & Wilson (1995)
Housing nonprofits	Program expenditures; management and general expenditures	Number of new houses, rehabilitated houses, recycled houses, and repaired houses	Ba et al. (2022); Berrett & Hung (2023); Coupet & Berrett (2018)
Mass transit systems (private, public, and nonprofit)	Service area population density, total operating	Fare revenue earned, total miles	Min & Ahn (2017)

	expenses, total labor hours	accumulated on active vehicles	
Microfinance institutions	Total assets, operating expenses, # of employees	Gross loan portfolio, inverse of average loan balance per borrower, # of active borrowers	Ayayi & Wijesiri (2018)
Museums	Total running and maintenance costs of the museum in the year; number of full-time equivalent workers in each museum	Number of physical visits and subscribers achieved by each museum	Bishop & Brand (2003)
Performing arts nonprofits	Total program expenses, total administrative expenses, full-time equivalent (FTE) employees	Number of participants	Hung & Berrett (2022)
Rural transit nonprofits	Vehicles in annual maximum service, number of volunteer drivers, operational expenses	Car miles; unique passenger trips	Coupet et al. (2021)
Sports clubs	Human resource costs, general expenditure resulting from the aggregation of operating costs, transports, facilities, and expense with activities	Revenues from activities, subsidies, membership fees, and sponsorship	Miragaia et al. (2016)
Symphony orchestras	Spending on staff, including printing and postage, phone, and other development expenses	Government funding, individual funding, business funding, and foundation funding	Luksetich & Hughes (1997)

The Role and Relevance of Resilience in the Nonprofit Sector: A systematic review of the literature

Grace Catherine De Oro – University of Maryland, Baltimore County

This review examines previous literature on resilience that focuses on nonprofit organizations. Growing literature examines organizational resilience in the context of nonprofits, however most scholars discuss resilience from a community perspective. This review employs Cochrane-Campbell protocols to establish a research question and keyword search protocols in advance. The main findings include themes in the literature around disturbances to the system, leadership and management trends, and financial resilience. Implications include insights for nonprofit managers when considering short- and long-term recovery plans and how to build a surplus beyond financial means. Ultimately there is a need for a holistic framework to bring together structures, people, and relationships in the discussion of fostering and being resilient. While bounce back is nearly impossible for smaller organizations, which are often in vulnerable financial situations to begin with, they still manage to evolve and survive.

Resilience, organizational resilience, non-profit organizations

Introduction

The coronavirus (COVID-19) pandemic led to a global crisis and disturbance that affected individuals and organizations. Nonprofit organizations are not immune to the pandemic's effects, resulting in declines and disruptions to how we live and the communities in which we function. More importantly, it threatened nonprofit organizations' ability to carry out their mission (Finchum-Mason et al., 2020; Maher et al., 2020,). The nonprofit sector plays an important and relevant role in disaster response after a large-scale human-made or natural disaster.

By examining how resilience is defined in the literature, this review explores how nonprofit organizations can define and build resilience. Examining resilience from the perspective of nonprofits is vital because of the role they play in serving a public need. Governments cannot solve every problem, and nonprofits can answer niche needs. This is not to say that government(s) do not play or have the potential to serve in a supportive role. Nonprofit organizations are unique as they operate in the third sector or third space, where public-private partnerships led by government and private interests work together to address a need. Nonprofits are answering the needs of the public, and there are opportunities for them to work alongside and with the government to solve problems. From an organizational perspective, individual nonprofits offer unique expertise and make decisions on how they function and carry out their mission. Self-governing is one of the five characteristics of a nonprofit that Salamon and Anheier (1997) used to define organizations in the sector, and it implies that they

have their own governance procedure and have a degree of autonomy (Anheier, 2014, p. 73). This structural-operational quality of nonprofits further supports why resilience is relevant to the study of them as an organization.

This review argues alongside Houston (2018) that recovery does not look like a return to 'normal' or homeostasis. What is normal is also up to interpretation in each instance because sometimes "normal" is no longer obtainable. This experience is consistent with the current COVID-19 pandemic, as we have referred to many things as the "new normal" (Corpuz, 2021). Corpuz (2021) argues that if society is open to change, the degree to which we are open will depend on our "capacity to adapt, manage resilience in the face of adversity, flexibility, and creativity without forcing us to make changes" (p. 1). Each disturbance provides a lesson, experience in coping, and/or failure that can be used in the future. A growing body of literature examines organizational resilience in the context of nonprofits, but most scholars discuss resilience from a community perspective. One cannot jump from the individual to the community level without recognizing an intermediate level where organizations are involved and contribute to the community (Figure 1). What is also unclear is what is happening or the contribution of the resilience of individuals to community resilience and vice versa. The path from individual to community resilience is not linear; the literature names public and private organizations involved in communities and their efforts, yet it fails to discuss what resilience looks like for organizations. Figure 1 shows multiple levels and shareholders of resilience (dashed lines) connected and contributing to each other (solid black lines). Just as individual resilience contributes or is related to an organization's resilience, the opposite is true where organizational resilience contributes or is related to individual resilience.

Figure 1. Resilience Shareholders Framework
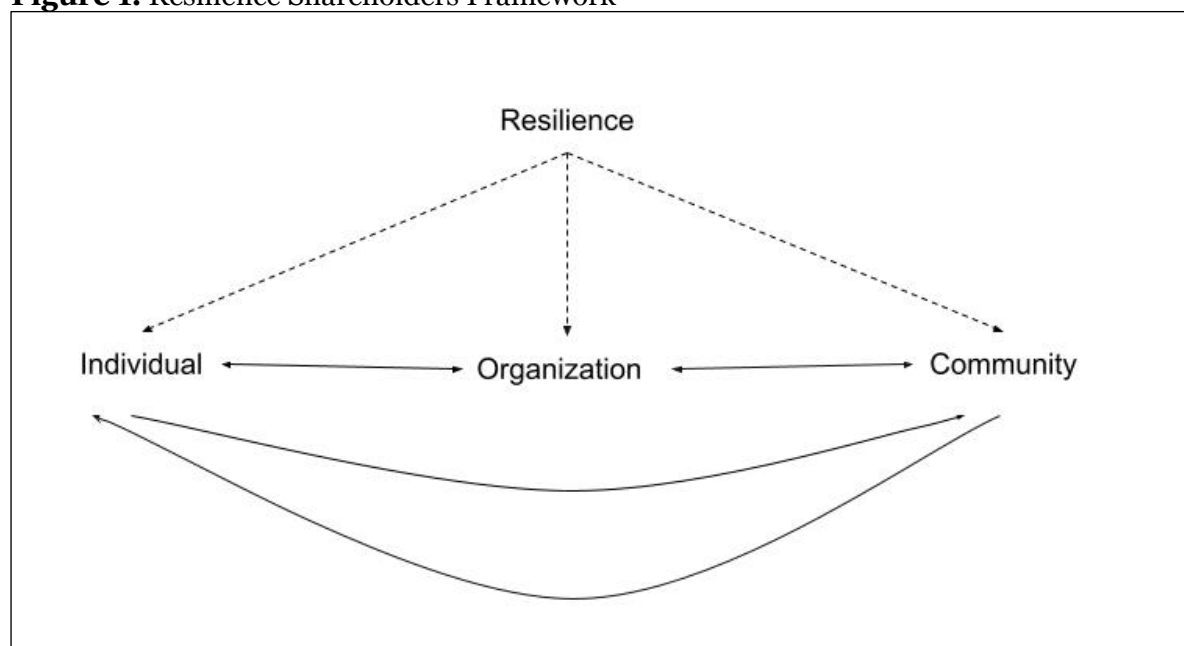
Literature Review

Aside from distinguishing individual and organization-level resilience, it is also important to define the difference between resiliency and resilience. Resiliency is an individual personality trait (derived from ego-resiliency), whereas resilience is a process (Bonilla, 2015, p. 10).

Establishing a clear definition of resilience is critically important to this analysis. A consistent definition should be used in the decision-making process to facilitate the formulation of compatible policy goals by a wide range of actors (Carlson et al., 2012). Researchers,

academics, practitioners, and policymakers need to be mindful of definitions that vary depending on the object of analysis. Biases can form based on a chosen definition, and it is our duty to recognize them during this process.

Conceptual frameworks and research models related to resilience are most common in psychiatry, developmental psychology, human development, medicine, epidemiology, and, more recently, risk management (Ledesma, 2014; Renschler et al., 2010; Rose, 2009). For several billions of years, biological systems have been resilient, with feedback systems focused on the ability to cope with stressors, not on the ability to predict or avoid them. Ecosystems are strengthened by having encountered stress (Lovins & Lovins, 1982). The Canadian ecologist Holling (1973) introduced resilience to understand the capacity of ecosystems that could persist in their original state even when affected by an outside variable (Lovins & Lovins, 1982). Wildavsky (1988) interpreted Holling's work on the "control of risk with the capacity to cope resiliently" as "low stability seems to introduce high resilience" (p. 78).

In 1996, Holling identified a broader kind of resilience, which he called ecological resilience, where "the magnitude of a disturbance that can be absorbed before a system changes the structure and transitions to another state" (p.33). Gunderson and Holling (2001) defined resilience as the capacity of a system to experience disturbance and maintain its functions and controls. Carpenter et al. (2001) extended the research by examining the magnitude of disturbance an ecological system could tolerate before fundamentally changing into a different region with a new set of controls. Meanwhile, Hamel and Valikangas (2003) stress that while recovery, flexibility, or crisis preparedness are important values of resilience, it is not the end-all; rather, they point to a distinct source of "sustainable competitive advantage" to achieve economic resilience. What begins to emerge is that resilience is a function of a system's vulnerability and its adaptive capacity (Dalziell & McManus, 2004).

Aaron Wildavsky, an American political scientist known for his pioneering work in public policy, government budgeting, and risk management, stated that "resilience is the capacity to cope with an anticipated danger after they become manifest, learning to bounce back" (1988, p. 77). He argues that societies face risks every day, and that social and technical systems are becoming more intertwined and interdependent over time (Wildavsky, 1988). Ultimately calling for an investment in resilience, not just prevention (Wildavsky, 1988; Wukich, 2013). Wildavsky is credited with introducing resilience to public administration, and he is the first to use the phrase "bounce back." In his book *Searching for Safety*, he hypothesized that "the growth of resilience depends upon learning how to deal with the unexpected" (Wildavsky, 1988, p. 77).

Dealing with the unexpected and adversity is difficult to predict. The anticipatory work of experts and governing authorities often falls short. While plans and theories look good on paper, in practice, we cannot accurately predict how many or what kind of disturbances we might encounter. Since the 2010s, many peer-reviewed articles and publications define resilience, resilient organizations, and resiliency. Most borrow some version or interpretation of Wildavsky's (1988) "bounce back" theory (Boin & van Eten, 2013; Okamoto, 2020; Valero et al., 2015). However, few studies properly cite or recognize Wildavsky's work, and the "bounce back" theory is accepted. Houston (2015) challenges this by offering his view of "bounce forward," representing a return to the pre-crisis baseline level on one or more measures. The emphasis on "forward" represents the passing time that occurs and is necessary for the return. It also recognizes that the return to the baseline is not simply how things were before (Houston, 2015, 2018).

Theoretical frameworks focus on examining an organization's financial health and vulnerability to determine if they are resilient and ultimately will survive. For example, the RISE Model used by Maher et al. (2020) strongly emphasizes financial capacity and financial impacts. While the model emphasizes a multi-response stage and a thinking-forward recovery

model, there is a strong message that financial capacity is a measurement of recovery and stabilization (Maher et al., 2020). Fewer studies focus on proactive recovery or a nonlinear model. Cyclical models, or those with feedback loops, include elements like realignment and reorganization with collective and collaborative efforts. Hutton's (2018) Proactive Recovery Transition Model shows that resilient organizations can exhibit leadership after a crisis to capture expanded services or audiences. It also identifies levels of integration with partners that combine resources to transition to longer-term recovery. As an organization becomes more integrated with personal connections, collective action, and collaborative oversight, it archives more sustainable resilience. Forces within the model also operate in reverse to make resilience sustainable. Evaluations conducted collaboratively by most integrated organizations might lead to the realignment of collective action groups and individual nonprofits (Hutton, 2018). A shortcoming in the literature is the use of recovery, sustainability, and resilience interchangeably. These terms are conceptually discussed in different ways throughout the literature when discussing organizational resilience.

The literature raises two questions: about nonprofits' practices and decision-making processes and how scholars evaluate nonprofit organizations. Practice questions aim to understand how the practitioner and the nonprofit organization make decisions or practices regarding resilience. Answers to these questions might include the organization's managerial or strategic practices. Meanwhile, the evaluation questions are posed to scholars or outside parties who aim to evaluate organizations. These questions aim to think critically about the framing and perspective used to approach an evaluation on resilience.

Practice Question
1) How do nonprofit organizations anticipate the unexpected while trying to develop and grow resilience?
2) What do "bounce back" or "bounce forward" practices look like for nonprofit organizations? Are they attainable in the long term?

Evaluation Question
1) How can we minimize the gap between theory and nonprofit organizations' practice?
2) What model (e.g., linear or feedback loop model) is most appropriate for modeling recovery where resilience is either a product and/or a part of an organization's functions?
3) What steps can we take as scholars to push beyond financial-based solutions when examining resilience?

Methodology

This analysis was designed to systematically review the current literature on resilience, from the perspective of nonprofit organizations. To minimize selection bias, this study avoids strategies that would limit searches to specific publication years, theories, journals, or disciplines. This review has an interest in literature that considers the COVID-19 pandemic.

Selection Criteria

To ensure replicability, limits on the scope were used. The following criteria were imposed in this search:

1) Empirical: This analysis is interested in the state of the research on resilience in nonprofit organizations; therefore, eligible literature must be empirical, where the researchers used qualitative, quantitative, or mixed methods. Articles that did not include literature reviews, narratives, theoretical analyses, and theoretical models were not included.

2) Nonprofit *organizations / NGOs*: Research articles must include data and theories from nonprofit organizations or NGOs. Articles were included even if nonprofit organizations were a passing mention. There is no limit on whether these organizations are domestic (within the U.S.) or foreign (outside the U.S.).

3) Nonprofit *resilience or resiliency:* Research articles must reference resiliency or resilience from the perspective of nonprofit organizations. Articles that reference human resilience (psychology) and financial resilience in the context of the for-profit business sector are not included.

Data And Methods

The goal of this study is to produce reliable, systematic, and robust conclusions about the state of current research. To achieve this goal, the Cochrane-Campbell protocols, which include recommendations for systematic reviews, were utilized to establish a research question and keyword search protocols in advance. An eligibility criterion was established to produce a reproducible methodology to conduct a systematic review that attempts to identify the studies that meet the criteria while minimizing bias in the selection or interpretation process (Gazley & Guo, 2020). The data collection process took place in the spring of 2020 and included several stages, beginning with keyword searches of articles and abstracts only. Each article was reviewed further for criteria results shown in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) format (Figure 2).

Stage One of the article section used the online archive EBSCO through the University of Maryland, Baltimore County's Library search engines and interlibrary loan services using the following keywords as search criteria where, in the full text and keywords include nonprofit resilience, and or nonprofit organizations and resiliency or resilience or resilient. There is no limit in the sample in regard to whether organizations are domestic (with in the U.S.) or foreign (outside the U.S.). Nor was there a time range filter when searching for articles. The preliminary sample included 246 articles.

In Stage Two of article selection, the sample was examined further to review the article's full abstract. If the abstract did not meet the three criteria, an article was removed from the sample. After reviewing the abstracts, the sample was reduced to 23 eligible articles.

Stage Three of article selection began with downloading the articles from EBSCO and the interlibrary loan system. Twenty-Two articles were available in the full text and were easily accessible. One article was removed from the final systematic review sample for an invalid link.

The final sample is broken down into themes to better grasp the literature. These themes were based on trends in the literature. While they do not aim to address the questions posed from the literature review, they do assist in answering the larger questions around nonprofit practices and decision-making processes and how scholars evaluate nonprofit organizations.

Data Trends and Findings

More than half of the 22 articles (Table 1) were published in 2021 or 2022. The articles ranged over eight years from 2014 to 2022, with natural disasters like Hurricane Katrina and Hurricane Florence, as well as the COVID-19 pandemic, identified as a disturbance. The methodology of the articles included both quantitative and qualitative studies, with some using mixed methods in their analysis and models. An important observation is how each

article and its author(s) choose to define and or articulate resilience in their study. Since there are varying definitions and defining variables, the scope of resilience is broad even in this smaller sample. A synthesis based on trends and themes is utilized to combine these varying definitions to increase the generalizability and transferability of the research. After reviewing the articles, three main themes were identified: Disturbances to the System, Financial Resilience, and Leadership and Management trends in Nonprofit Organizations.

Figure 2. PRISMA Flow chart of the review protocol and results.
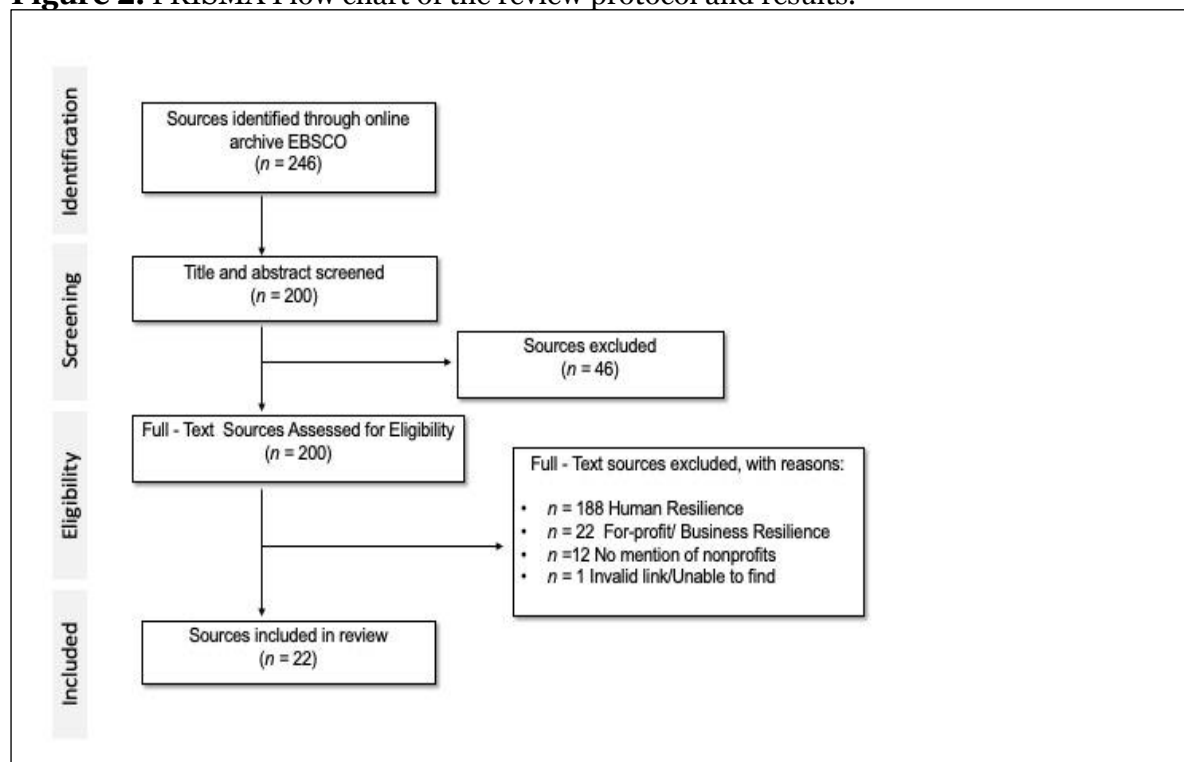
In the sample, resilience is referred to as an interdisciplinary concept with the most common definitions using terms like "bounce back," the "ability to absorb," and "positive adaptability" (Table 2) (Cerquetti & Cutrini, 2021; Chen, 2021; Hutton et al., 2021; Kim, 2022; Maher et al., 2020; Okamoto, 2020; Paluszak et al., 2021; Pena et al., 2014; Rochet et al., 2008; Whitman, 2021). Cerquetti and Cutrini (2021) explain that the most utilized definitions take an evolutionary approach based on adaptability, transformability, configuration, and reorganization when discussing resilience. This is consistent with early resilience scholarship from the study of ecosystems in zoology (Holling, 1973, 1996). Disaster literature focuses on organizational resilience as a measure of a "complex blend of behaviors, perspectives, and interactions that can be developed, measured, and managed" (Lengnick-Hall & Beck, 2005, as cited in Pena et al., 2014, p. 591). Because of a natural disaster, two variables are emphasized when discussing firm or sector resilience: the structure of the organization and the extent of participation in the community (Pena et al., 2014).

Twenty-first-century scholars warn that resilience is a buzzword in today's research and vocabulary (Martin & Sunley, 2015, as cited by Cerquetti & Cutrini, 2021). They warn scholars to be aware of its meaning and context since buzzwords can be used due to popularity with little to no understanding of the deeper meaning. The term is often used in discussions around financial resilience and vulnerability. Searing et al. (2021) call for moving beyond the use of vulnerability and survival analysis when looking at the resilience of a nonprofit organization because the results fail to provide answers on how service delivery continued and how to recover. This is a continued theme and critique where resilience is generally understood, but

from an analytic perspective, it lacks consistent qualitative or quantitative elements for analysis and use across disciplines.

Table 1. Literature Review Sample Summary

Authors	Publication Date	Article Title	Journal
Cerquetti, Mara; Cutrini, Eleonora	2022	Structure, People, and Relationships: A Multidimensional Method to Assess Museum Resilience.	Nonprofit & Voluntary Sector Quarterly
Chen, Xintong	2021	Nonprofit Financial Resilience: Recovery from Natural Disasters.	Voluntas: International Journal of Voluntary and Nonprofit Organizations
Chen, Xintong	2022	Nonprofit Resilience in a Natural Disaster Context: An Exploratory Qualitative Case Study Based on Hurricane Florence	Human Service Organizations Management Leadership & Governance
Fathalikhani, Somayeh; Hafezalkotob, Ashkan; Soltani, Roya	2020	Government intervention on cooperation, competition, and competition of humanitarian supply chains	Socio-Economic Planning Sciences
Geller, Stephanie Lessans; Salamon, Lester M.; Mengel, Kasey L.	2010	Museums and Other Nonprofits in the Current Recession: A Story of Resilience, Innovation, and Survival	The Journal of Museum Education
Horvath, Aaron; Brandtner, Christof; Powell, Walter W.	2018	Serve or Conserve: Mission, Strategy, and Multi-Level Nonprofit Change During the Great Recession.	Voluntas: International Journal of Voluntary & Nonprofit Organizations
Hutton, Nicole S.	2018	Sustaining Resilience: Modeling Nonprofit Collaboration in Recovery	Professional Geographer

Hutton, Nicole. S.; Mumford, Steve W.; Saitgalina, Marina; Yusuf, Juita-Elena. (Wie); Behr, Joshua G.; Diaz, Rafael; & Kiefer, John. J.	2021	Nonprofit capacity to manage hurricane-pandemic threat: Local and national perspectives on resilience during COVID-19.	International Journal Of Public Administration
Irvin, Renee A.; Furneaux, Craig W.	2021	Surviving the Black Swan Event: How Much Reserves Should Nonprofit Organizations Hold?	Nonprofit And Voluntary Sector Quarterly
Kim, Min-Hyu	2022	Framing Effects, Procedural Fairness, and the Nonprofit Managers' Reactions to Job Layoffs in Response to the Economic Shock of the COVID-19 Crisis.	Voluntas: International Journal of Voluntary And Nonprofit Organizations
Maher, Craig S.; Hoang, Trang; Hindery, Anne	2020	Fiscal Responses to COVID-19: Evidence from Local Governments and Nonprofits.	Public Administration Review
Moran, Kenneth A.	2016	Organizational resilience: Sustained institutional effectiveness among smaller, private, non-profit US higher education institutions experiencing organizational decline.	Work- A Journal of Prevention Assessment & Rehabilitation
Okamoto, Kristen E.	2020	'As resilient as an ironweed:' narrative resilience in nonprofit organizing	Journal of Applied Communication Research
Paluszak, Grzegorz Tadeusz; Wiśniewska-Paluszak, Joanna Alicja; Schmidt, Joanna; Lira, Jarosław	2021	The Organizational Resilience of Rural Non-Profits under Conditions of the COVID-19 Pandemic Global Uncertainty.	Agriculture
Pena, Anita A.; Zahran, Sammy; Underwood, Anthony; Weiler, Stephan	2014	Effect of Natural Disasters on Local Nonprofit Activity.	Growth & Change

Searing, Elizabeth A. M.; Wiley, Kimberly K.; Young, Sarah L.	2021	Resiliency Tactics During Financial Crisis: The Nonprofit Resiliency Framework	Nonprofit Management & Leadership
Țiclău, Tudor; Hințea, Cristina; Trofin, Constantin	2021	Resilient Leadership. Qualitative Study on Factors Influencing Organizational Resilience And Adaptive Response To Adversity.	Transylvanian Review of Administrative Sciences
Waerder R; Thimmel S; Englert B; Helmig B	2021	The Role of Nonprofit-Private Collaboration for Nonprofits' Organizational Resilience.	Voluntas: International Journal of Voluntary And Nonprofit Organizations
Whitman, John R.	2021	Does Covid Portend a Shocking Future for Nonprofits?	Canadian Journal of Nonprofit And Social Economy Research
Witmer, Hope; Mellinger, Marcela Sarmiento	2016	Organizational resilience: Nonprofit organizations' response to change	Work-A Journal of Prevention Assessment & Rehabilitation
Routhieaux, Robert L.	2015	Shared leadership and its implications for nonprofit leadership	Journal of Nonprofit Education and Leadership

Table 2. Resilience Definition Themes in Systematic Literature Review

Definition Theme	Author, year
Bounce Back	Chen 2022
	Chen, 2021
	Paluszak et al., 2021
	Țiclău et al., 2021
	Waerder et al., 2021
	Horvath et al., 2018
	Pena et al., 2014
Ability to Absorb	Cerquetti et al., 2022
	Whitman, 2021
	Witmer and Mellinger, 2016

Positive Adaptability	Hutton et al., 2021
	Paluszak et al., 2021
	Țiclău et al., 2021
	Waerder et al., 2021
	Maher et al, 2020
	Moran, 2015
	Pena et al., 2014
Does not explicitly define resilience but uses the term, provides examples, and cites sources that do define resilience.	Kim, 2022
	Fathalikhani et al., 2021
	Searing et al., 2021
	Irvin et al., 2021
	Hutton, 2018
	Routhieaux, 2015
	Geller et al., 2010

Disasters and Disturbances to the System

Resilience refers to persistence in the face of disturbance across the literature. While some disturbances can cause temporary changes in environmental conditions, they cause pronounced changes in an ecosystem (Holling, 1973, 1996). Disasters like the COVID-19 pandemic have demonstrated large and small shocks throughout all levels of society and the globe. The Great Recession (2007-2009) is a prominent example of nonprofit resilience research on responses to external threats in the sample in the context of economic downturns (Horvath et al., 2018; Moran, 2015; Ticlău et al., 2021).

Literature from earlier years in the sample reference major events like the 9/11 Terrorist Attack, Hurricane Katrina, and the Great Recession (Horvath et al., 2018; Moran, 2015; Pena et al., 2014). These events were highly destructive, as they continue to have long-term effects and influence policy and our lives more than ten years later (Horvath et al., 2018; Moran, 2015). While many disasters are challenging to prepare for because of unpredictability, hurricanes provide a unique circumstance. The common roles of nonprofits in disaster relief include assisting in providing up-to-date information, assisting the community with FEMA-related processes, and distributing goods (Pena et al., 2014). Preparation for major weather events includes an extensive system of infrastructure, data, and technology that nonprofit organizations, the public, and governments can use to better prepare for the initial shock and recovery. For example, in a qualitative study on nonprofit resilience and natural disasters, those who experienced Hurricane Matthew reported that learning from the outcome helped them prepare better to recover more effectively and efficiently from Hurricane Florence just two years later (Chen, 2022).

Nonprofits remain on the frontline of natural disasters like hurricanes by providing immediate response and long-term assistance to individuals and communities, unlike FEMA, which will deploy and eventually dissolve their efforts (Pena et al., 2014). Hutton (2018) argues that after a disaster, the preexisting nonprofit organizations that exhibit the most resilience do so by altering their operations "to share the burden." Typically, no single agent has enough resources to manage and respond to a disaster. Thus, coordination and interactions among various aid agencies are necessary (Fathalikhani et al., 2021). This supports the claim that governments cannot answer all public needs, and that there is a niche role for nonprofit organizations. Other actors who support the government may include aid agencies like local and international NGOs, the military, and private sector organizations (Fathalikhani et al., 2021; Pena et al., 2014).

The COVID-19 pandemic exposed many weaknesses including the healthcare system and threatened the financial capacity of public and nonprofit organizations (Maher et al., 2020). The pandemic is a unique disturbance to the system because the unexpected outbreak of COVID-19 caused a "dramatic increase in uncertainty" across the globe (Paluszak et al., 2021). Unlike our experiences with hurricanes the only major event the United States experienced comparable to the COVID-19 pandemic is the Spanish flu pandemic in 1918.

The shocks or results of COVID-19's disturbances have an immediate and abrupt effect on the normal function of nonprofit organizations (Ticlău et al., 2021). With growing pressures to maintain an organization's functions, adaptation and change are necessary (Maher et al., 2020; Ticlău et al., 2021). Instability or changes in external funding sources, as well as the supply chain, remain consistently identified by the nonprofit. Although during this time, funding sources ranged from individual donor donations, grants, federal funding, and local government funding, there was a significant concern about the financial status of organizations (Maher et al., 2020; Searing et al., 2021; Ticlău et al., 2021).

Leadership and Management Trends

Nine articles focused specifically on nonprofit leadership and personnel resilience rather than the organization. A focus on leadership and personnel is salient because all organizations are made up of individuals with a particular purpose or goal in mind. It is vital to examine internal and external factors, including revenue, leadership, and partnerships. Hutton et al. (2020) call for focusing on "process, practice, and people" through financial management, strategic planning, and staff management. It is important to remember that nonprofits are embedded in their communities and play a role in the community's adaptive capacity. Scholars should examine nonprofits as an organism because they cannot provide the services and meet their missions without the individuals that keep them alive and the system they exist in.

Adopting a more holistic approach to evaluation is essential to understand how organizations are resilient and foster resilience. While it may seem more pressing to identify the external factors and threats facing a nonprofit organization, it is the internal factors, resources, and tactics contributing to the organization's main functions. Searing et al. (2021) identify five tactical themes based on resources utilized: financial, human resources, outreach programs, services, management, and leadership. While the tactical themes are diverse, the resiliency tactics mentioned focus primarily on financial affairs and administrative tasks when considering resilience. The literature supports and argument for nonprofit organizations dedicating resources and time to resiliency, starting with organizational management fundamentals.

Additionally, organizations are the product of the routines and practices they adopt (Horvath et al., 2018). This is crucial for collective action. Searing et al. (2021) attest that there is a need for nonprofit managers to balance the goals of organizational persistence with those that allow continued delivery of human services during a time of crisis. To sustain resilience, nonprofit collaboration in recovery should also be considered (Hutton, 2018). With the goal of a sustainable recovery process, there is a need for mitigation and preparation phases to strategically plan for sustainable resilience and, secondly, a long-term continuation of resilience post-disaster (Hutton, 2018).

While using survival analysis and financial vulnerability is common in analyses, it does not answer the practical question of how organizations adopt, foster, and implement resilience (Horvath et al., 2018; Searing et al., 2021). One strategy includes mapping out an organization's strengths and weaknesses while also paying attention to the perceptions and insights of stakeholders outside of the organization's management. Leadership and

management are important because the individuals involved in decision-making processes and day-to-day activities are crucial data points as the boots-on-the-ground and first-person perspectives.

Managers and Management Styles

The attitudes of nonprofit managers, specifically optimism, are noted as an explanation for relative success in nonprofit organizations. Managers who "do not take defeat easily" and who remain positive are accustomed to "beating the odds" even when they are not in their favor (Geller et al., 2010, p. 138). In the Geller et al. (2010) study, about half of the respondents believed their "future is bright" because they used the crisis as an opportunity to redesign their organizations to improve conditions in both the short and long term. One organization in this study explained its growth as being a "leaner, healthier organization" thanks to examining the procedures and programs that were in place (Geller et al., 2010, p. 138).

Nonprofit organizations were urged to adopt managerial practices associated with the business sector in the 1990s and early 2000s to articulate goals and the specific processes to achieve them (Horvath et al., 2018). The use of strategic planning to develop organizational goals and the specific steps to achieve them started to be utilized by hiring consultants, undergoing financial audits, and using quantitative performance. This strategic planning empowers nonprofit organizations to practice planning through everyday activities for the organization's members at all levels, from the board to volunteers. It ensures that mission-driven organizations do not pursue activities vaguely related to their goals without planning strategically (Geller et al., 2010; Horvath et al., 2018).

An organization's mission is not static and serves as the lens through which it will work within its area of focus. It often motivates organizations through values and morals, and it strongly disapproves of deviations from them. This does not mean nonprofit organizations do not change their mission over time. Influences by external factors create new needs and require new services, such as in the case of disasters.

Nonprofit organizations that are invested put considerable thought and practice into plans over time, which include personnel updates, seeking outside input, and making all parts of the strategic plan actions of the staff, departments, and volunteers and ultimately shape how they respond to crisis (Horvath et al., 2018). Being able to stay the course allows for the organization to standardize routines and practices. Rather than spending a considerable amount of time trying to "bounce back" they are working proactively. Pauluszak et al. (2021), argues that a disruption is the beginning of transformation and "bouncing forward." Transformation resulting from external disturbances lead to a "thriving organization" that has a new resilience (Pauluszak et al., 2021). So, while a disruption is not something a nonprofit organization wants, there is an opportunity for a positive outcome in the context of resilience. Ultimately, strategic planning offers the organization, its managers, and the individuals that make up the organization a future orientation.

Characteristics of Nonprofit Organizations

When looking at an organization as a whole there are significant variables like size and age to consider. Larger organizations can better overcome and manage the situations due to higher levels of formal management and crisis response (Ticlău et al., 2021). For smaller organizations, regardless of the sector, the focus is the organization's survival. While smaller organizations tend to rely on their "ad-hoc solutions" because they tend to work and keep them afloat, they do not translate to a higher level of resilience long-term. Mature and more established nonprofit organizations have the advantage of integrating and implementing mechanisms through management and leadership. These mechanisms can be used more efficiently and effectively to respond to shocks through their learned experiences. However,

learned experiences are only as helpful as the leadership that chooses to adopt or innovate in the future. Learned experience is especially helpful for emergency and disaster aid nonprofits. Chen (2022) calls for future studies exploring recovery strategies that address staffing, needs for services, and the development of best practices. Nevertheless, having a plan does not make all nonprofit organizations immune to disaster impact. It can help them cope with the impacts, especially for those whose services were impacted (McManus et al., 2008, as cited by Chen, 2022).

Financial Resilience

Financial health is examined to determine the success or failure of an organization from the nonprofit sector to the business sector. In the case of nonprofits, financial vulnerability is most analyzed and critiqued. Research and literature stress how an organization can and should reduce its financial vulnerability. It is important to recognize that in trying to build financial resilience, there is competition among nonprofit organizations. Nonprofits are often financially independent of the government and usually are funded in other ways. When challenges or tragedies exist, the competition increases, and as there are limited numbers of donors or funds, there is competition for financial resources (Fathalikhani et al., 2021). While competition creates more challenges, competition in nonprofit organizations allows social welfare to be provided at a higher level with fewer financial resources needed (Fathalikhani et al., 2021). What has not been explored is how nonprofits "bounce back" to their original performance level (Chen, 2021). When they are already in a poor financial state due to internal or external reasons, does the organization ever fully bounce back? This offers an opportunity for exploration from a managerial and theoretical perspective.

Chen (2021) defines nonprofit financial resilience as "the ability or capacity of an organization to bounce back to at least its original financial performance level after a disruptive event" (p. 1010). Organizational resilience characteristics such as nonprofit size and financial resources are associated with total expense recovery. Meanwhile, the diversity in revenue and the organization's age were not statistically significant. Early results from research that focuses on the COVID-19 pandemic and the nonprofit sector showed that organizations with financial reserves exhibited more resilience in the wake of extraordinary financial shock and fewer disruptions to their mission-related programming compared to other organizations lacking significant savings (Kim & Mason, 2020 as cited in Irvin & Furneaux, 2021).

The post-disaster recovery time varies between 3 to 5 years; some even fail to reach their original financial levels after 10 years (Chen, 2021). This shows how the effects of a disaster can cause long-term damage to an organization and even lead to its mortality. Data suggests that around three months of operating reserves is best to ensure the stability of programming in the event of financial shocks; however, this is only true for the most prominent nonprofits (Irvin & Furneaux, 2021). For medium to small nonprofits, three months of operating reserves are hard to accumulate and sometimes brutally inadequate to ensure stability. Meanwhile, medium-sized organizations need about five and a half months of reserves.

Why one strategy is selected over another is not always clear. However, there is some evidence that an organization's mission may play an important role as a call to action or constraint to act (Horvath et al., 2018). Declining revenues, increased costs, declining endowments, increased competition for resources, and increased demand all contribute to the fiscal stresses nonprofit organizations face (Geller et al., 2010). Despite these findings, a 2010 study found that although the sample of nonprofit organizations and museums were facing daunting fiscal pressures, they were considerably resilient (Geller et al., 2010). To achieve resilience in these difficult economic conditions, nonprofit organizations and museums use coping strategies such as fundraising, belt-tightening, entrepreneurial expansions, and optimism (Geller et al., 2010).

Furthermore, fundraising, while a common practice of nonprofit organizations, is not always the answer. Geller et al. (2010) found that targeted efforts focused heavily on donations from individual donors, foundations, and corporations rather than on state, local, and federal funds that they usually were dependent on. Nonprofits were also more likely to use creative fundraising approaches like applications to smaller local banks for smaller loans to support specific projects.

Belt-tightening is used to cut administrative costs, while others may tap into collaborative efforts and relationships with other nonprofits to increase financial resources. Nonprofits often lean heavily on their social capital, especially those rooted in communities (Pena et al., 2014). They benefit heavily from their local knowledge, networks, and credibility within the community to gain support both financially and otherwise. Other belt-tightening strategies include hiring freezes, paring down programming, increasing reliance on volunteers, and shifting to cheaper products or services (Geller et al., 2010). Many eco-friendly tactics ended up being the most innovative and cost-effective by just simply cutting the use of gas and electricity because office spaces were not being used, along with fewer travel expenses. Innovative and entrepreneurial expansions also served as defense measures to help cope with cuts. Improving and expanding market efforts and implementing or expanding advocacy efforts are among the most successful coping strategies.

Discussion

The dialog around resilience in the context of nonprofit organizations is new and emerging in the literature. This study has limitations as the sample was retrieved in the spring of 2022, includes only English text, and focuses specifically on the organizational perspective. Grey literature was not included, and newly published studies consistently joined the conversation. However, a plethora of literature focuses on community and individual resilience from a physiological perspective, as well as emergency management literature. While these contribute to the literature, this review focused specifically on an organizational perspective, focusing on nonprofit organizations. Selection bias and human error should be considered, as I am the sole investigator who screened the titles and abstracts of each piece of literature. Word choice or poorly written abstracts and titles could have also affected the sample as they were utilized during screening.

Resilience in public administration was defined by Arron Wildavsky in 1988; only after examining multiple decades of published research can we begin to accumulate enough literature to begin the conversation about how we can define resilience and build a holistic framework. A holistic framework would bring together structures, people, and relationships in the discussion of fostering and being resilient. Scholars need to move beyond measuring success based on financial status. While financial resilience is an essential pillar of organizational resilience, there are assets of equal importance. Social assets are integral to understanding resilience, including building and sharing knowledge among the organization and networks. Managers and leaders who have lived experiences are essential to building resilience. The literature supports that "bounce back" is nearly impossible for smaller organizations, who are often in vulnerable financial situations to begin with, and yet they still manage to evolve and survive (Chen, 2021, 2022; Geller et al., 2010; Horvath et al., 2018; Ticlău et al., 2021; Pauluszak et al., 2021).

Figure 3. The Adaptive Process of Resilience with Shareholder Framework
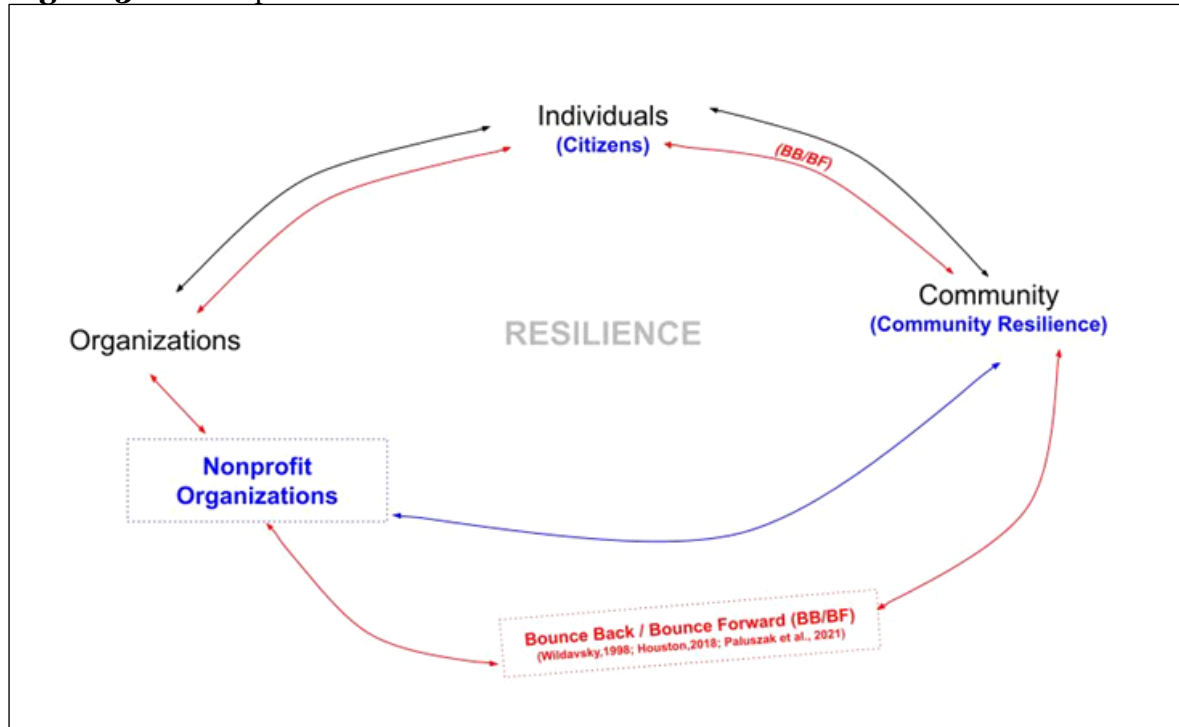
The Adaptive Process of Resilience with Shareholders Framework (Figure 3) builds off the Resilience Shareholders Framework (Figure 1) to better depict resilience as a process. Each shareholder (individual, organization, and community) is connected by a black arrow, showing how resilience from one shareholder contributes to another. The blue arrow shows the "how" or the process that nonprofits are contributing to community resilience; the blue arrow only goes one direction, unlike the other arrows in the figure. It is the "how," the process, and the work of nonprofits to build and contribute to community resilience. The red arrows show "bounce back" and "bounce forward" in both directions as part of the resilience process.

The adoption of the "bounce back" or "bounce forward" approach has advantages and disadvantages; however, when looking at the sample, the literature published in the last few years shows that bouncing forward is a better approach to building resilience. While these approaches may be an active decision through management and strategic planning, I do not think a baseline recovery is always possible. It is also critical to remember that strategic planning is a double-edged sword, as strategic plans can be constraining when they are being followed conservatively. Other times there could be a need to throw out any strategic plan all together calling for innovation. Disturbances provide unique circumstances that require the organization to innovate and transform in its attempt to survive and thrive.

Nonprofit organizations could benefit from surveying their financial and social assets. By identifying these assets internally or having a third party assist in the process, nonprofits can identify the tools in their toolkit that can be strategically used during crises. Community asset mapping is a tool used to collect information about the strengths and resources of a community and can help uncover solutions. Community asset mapping is utilized in public health and education policy, where the goal is to improve the community (Lou et al., 2022). Asset mapping is helpful when you want to start a new program, make program decisions, or mobilize and empower a community (Lou et al., 2022). An advantage of asset mapping is that it promotes community involvement, ownership, and empowerment, which is a strong point for nonprofit organizations. Community asset mapping can be applied and translated to nonprofit organizations to prioritize resilience.

Based on this systematic review, an asset I believe we should explore is how organizations may have resilience, mainly because people are resilient by nature. Missions have a role in an organization's resilience because they can be more strategic through management and leadership. We must remember that organizations are made up of people; without their values and morals, the work that brings missions alive would not be possible. Collaborative efforts and coalitions with other organizations are crucial to building resilience to expand our relationships and social capital. Resiliency is an individual personality trait, whereas resilience is a process (Bonilla, 2015).

Conclusion

Disasters like the COVID-19 pandemic and recent natural disasters are increasingly putting pressure on nonprofit organizations to join forces in solving social and humanitarian problems. Their role as primary actors in the communities devastated by disasters makes them especially important to understand when trying to grasp resilience from an organizational perspective.

This study contributes to both research and practice, first by highlighting the role of resilience in the nonprofit sector in the literature by building on the work of Carlson et al.'s (2012) literature review on resilience. The analysis also supports the argument for a more unified or standardized approach to defining resilience. We need consistency and an understanding of the definition to measure resilience and develop a way to increase resilience. Literature published in the past three years is becoming more consistent with using themes and terms around the definition of resilience.

This study provides insights to practitioners on when managers should consider short-term and long-term recovery plans and how to build a surplus beyond financial means. For some nonprofit organizations, short-term goals are only obtainable based on internal and external variables. External forces like disasters and disturbances show us just how resilient some nonprofit organizations can be in the short term through creative thinking, optimism, and entrepreneurial spirit. Future research should examine the organizational choices or changes due to external or internal forces. This will help to understand the organization's priorities and mindsets when experiencing these forces. You might hypothesize that these new experiences and perspectives will require a short- or long-term change. Asking about short-term impacts can help identify the organization's immediate priorities. At the same time, long-term impacts tend to require more reflection and are sometimes not as apparent in the moment.

Future research areas could include the interchangeability with terms such as sustainability and recovery. In trying to define resilience understanding the terms at the practice level will help us to understand if there are differences. This research could be better supported by the emergency management and disaster response literature, which is present in this analysis but not at a level to support the distinction in this analysis. Another area for future research is specific organizational assessment tools that nonprofits could adapt or employ from other disciplines. The literature offers asset mapping as one tool, but exploring others to frame resilience from an organizational perspective is unclear. Lastly and more abstractly, the concerns around resilience as a buzzword; there are dangers and concerns regarding buzzwords in both policy and practice. A comparison of use in policy and practice might spark stakeholders to ask: Is resilience the word we want to use? Or should we use resilience in this way?

When managers think of resilience as a process and a verb in their organizations, they can build it into their daily practices. We cannot cite or call for resilience without a robust comprehension of its meaning in the context. If we can agree that resilience is a process and

is only possible through the work of individuals first and second organizations, we can continue to work towards strengthening our frameworks and theories of analysis.

Disclosure Statement

The author declares that there are no conflicts of interest that relate to the research, authorship, or publication of this article.

References

Anheier, H. K. (2014). In Nonprofit organizations: Theory, management, policy. Abingdon, Oxon: Routledge.

Boin, A., & van Eeten, M. J. (2013). The Resilient Organization. Public Management Review, 15(3), 429–445. https://doi.org/10.1080/14719037.2013.769856

Bonilla, M. (2015). Building Resilience in Small Nonprofits. Organization Development Practitioner, 47(1), 8–14.

Carlson, J. L., Haffenden, R. A., Bassett, G. W., Buehring, W. A., Collins III, M. J., Folga, S. M., ... & Whitfield, R. G. (2012). Resilience: Theory and Application (No. ANL/DIS-12-1). Argonne National Lab.(ANL), Argonne, IL (United States).

Carpenter, S., Walker, B., Anderies, J. M., & Abel, N. (2001). From metaphor to measurement: resilience of what to what?. Ecosystems, 4(8), 765-781.

Cerquetti, M., & Cutrini, E. (2022). Structure, people, and relationships: A multidimensional method to assess museum resilience. Nonprofit and Voluntary Sector Quarterly, 1–23. https://doi.org/10.1177/08997640211068409

Chen, X. (2021). Nonprofit Financial Resilience: Recovery From Natural disasters. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, 32(5), 1009–1026. https://doi.org/10.1007/s11266-021-00415-w

Chen, X. (2022). Nonprofit resilience in a natural disaster context: An exploratory qualitative case study based on Hurricane Florence. Human Service Organizations: Management, Leadership & Governance, 46(2), 145–161. https://doi.org/10.1080/23303131.2021.1946455

Corpuz, J. C. (2021). Adapting to the culture of 'new normal': An emerging response to covid-19. Journal of Public Health, 43(2). https://doi.org/10.1093/pubmed/fdab057

Dalziell, E. P., & McManus, S. T. (2004). Resilience, vulnerability, and adaptive capacity: implications for system performance.

Fathalikhani, S., Hafezalkotob, A., & Soltani, R. (2020). Government intervention on Cooperation, competition, and cooperation of humanitarian supply chains. Socio-Economic Planning Sciences, 69. https://doi.org/10.1016/j.seps.2019.05.006

Finchum-Mason, E., Husted, K., & Surez, D. (2020). Philanthropic Foundation Responses to COVID-19. Nonprofit and Voluntary Sector Quarterly, 089976402096604.

Gazley, B., & Guo, C. (2020). What do we know about nonprofit collaboration? A systematic review of the literature. Nonprofit Management and Leadership, 31(2), 211–232. https://doi.org/10.1002/nml.21433

Geller, S. L., Salamon, L. M., & Mengel, K. L. (2010). Museums and other nonprofits in the current recession. Journal of Museum Education, 35(2), 129–140. https://doi.org/10.1080/10598650.2010.11510659

Gunderson, L. H., and Holling, C. S. (2001). Panarchy: understanding transformations in human and natural systems. Washington, D.C., USA Island Press pp.(25-62)

Hamel, G., & Valikangas, L. (2003). The Quest for Resilience. Harvard Business Review, 81(9), 52-63.

Holling, C.S. (1973). Resilience and stability of ecological systems. Annual Review of Ecology and Systematics 4:1-23.

Holling, C.S. (1996). Engineering resilience versus ecological resilience. Pages 31- 44 in P. Schulze, editor. Engineering within ecological constraints. National Academy Press, Washington, D.C., USA.

Horvath, A., Brandtner, C., & Powell, W. W. (2018). Serve or conserve: Mission, strategy, and multi-level nonprofit change during the Great Recession. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, 29(5), 976–993. https://doi.org/10.1007/s11266-017-9948-8

Houston, J. B. (2015). Bouncing forward: Assessing Advances in Community Resilience Assessment, Intervention, and Theory to Guide Future Work. American Behavioral Scientist, 59(2), 175–180. https://doi.org/10.1177/0002764214550294

Houston, J. B. (2018). Community Resilience and Communication: Dynamic Interconnections between and among individuals, families, and organizations. Journal of Applied Communication Research, 46(1), 19–22. https://doi.org/10.1080/00909882.2018.1426704

Hutton, N. S. (2018). Sustaining resilience: Modeling nonprofit collaboration in recovery. The Professional Geographer, 70(4), 655–665. https://doi.org/10.1080/00330124.2018.1443479

Hutton, N. S., Mumford, S. W., Saitgalina, M., Yusuf, J.-E. (W., Behr, J. G., Diaz, R., & Kiefer, J. J. (2021). Nonprofit capacity to manage hurricane-pandemic threat: Local and national perspectives on resilience during COVID-19. International Journal of Public Administration, 44(11-12), 984–993. https://doi.org/10.1080/01900692.2021.1922439

Irvin, R. A., & Furneaux, C. W. (2021). Surviving the black swan event: How much reserves should nonprofit organizations hold? Nonprofit and Voluntary Sector Quarterly, 51(5), 943–966. https://doi.org/10.1177/08997640211057405

Kim, M. H. (2022). Framing effects, procedural fairness, and the nonprofit managers' reactions to job layoffs in response to the economic shock of the COVID-19 crisis. VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, 33(5), 1035–1050. https://doi.org/10.1007/s11266-021-00439-2

Ledesma, J. (2014). Conceptual frameworks and research models on resilience in leadership. SAGE Open, 4(3), 1–8. https://doi.org/10.1177/2158244014545464

Lovins, A. B., & Lovins, L. H. (1982). Brittle Power: Energy Strategy for national security. Brick House.

Luo, Y., Ruggiano, N., Bolt, D., Witt, J.-P., Anderson, M., Gray, J., & Jiang, Z. (2022). Community asset mapping in public health: A review of applications and approaches. Social Work in Public Health, 1–11. https://doi.org/10.1080/19371918.2022.2114568

Maher, C. S., Hoang, T., & Hindery, A. (2020). Fiscal Responses to COVID -19: Evidence from Local Governments and Nonprofits. Public Administration Review, 80(4), 644-650. doi:10.1111/puar.13238

Moran, K. A. (2016). Organizational resilience: Sustained institutional effectiveness among smaller, private, non-profit US higher education institutions experiencing

organizational decline. Work, 54(2), 267–281. https://doi.org/10.3233/wor-162299

Okamoto, K. E. (2020). 'as resilient as an ironweed:' narrative resilience in nonprofit organizing. Journal of Applied Communication Research, 48(5), 618–636. https://doi.org/10.1080/00909882.2020.1820552

Paluszak, G. T., Wiśniewska-Paluszak, J. A., Schmidt, J., & Lira, J. (2021). The organizational resilience (or) of rural non-profits (RNPOs) under conditions of the COVID-19 pandemic global uncertainty. Agriculture, 11(7), 670–693. https://doi.org/10.3390/agriculture11070670

Pena, A. A., Zahran, S., Underwood, A., & Weiler, S. (2014). Effect of natural disasters on local nonprofit activity. Growth and Change, 45(4), 590–610. https://doi.org/10.1111/grow.12056

Renschler, C.S., A.E. Frazier, L.A. Arendt, G. Cimellaro, A.M. Reinhorn, and M. Bruneau, (2010), A Framework for Defining and Measuring Resilience at the Community Scale: The PEOPLES Resilience Framework, U.S. Department of Commerce National Institute of Standards and Technology, NIST GCR 10-930.

Rochet, C., Keramidas, O., & Bout, L. (2008). Crisis as change strategy in public organizations. International Review of Administrative Sciences, 74(1), 65–77. https://doi.org/10.1177/0020852307085734

Routhieaux, R. L. (2015). Shared leadership and its implications for nonprofit leadership. *Journal of Nonprofit Education and Leadership*, *5*(3).

Rose, A., (2009). Economic Resilience to Disasters, CARRI Research Report 8. Available at http://www.resilientus.org/library/Research_Report_8_Rose_1258138606.pdf,

Salamon, L. M., & Anheier, H. K. (1997). Defining the nonprofit sector: A cross-national analysis. University Press.

Searing, E. A., Wiley, K. K., & Young, S. L. (2021). Resiliency tactics during financial crisis: The nonprofit resiliency framework. Nonprofit Management and Leadership, 32(2), 179–196. https://doi.org/10.1002/nml.21478

ȚICLĂU, T., HINȚEA, C., & TROFIN, C. (2021). Resilient leadership. qualitative study on factors influencing organizational resilience and adaptive response to adversity. Transylvanian Review of Administrative Sciences, (SI 2021), 127–143. https://doi.org/10.24193/tras.si2021.7

Valero, J. N., Jung, K., & Andrew, S. A. (2015). Does transformational leadership build resilient public and nonprofit organizations? *Disaster Prevention and Management*, *24*(1), 4–20. https://doi.org/10.1108/dpm-04-2014-0060

Waerder, R., Thimmel, S., Englert, B., & Helmig, B. (2021). The role of nonprofit–private collaboration for nonprofits' organizational resilience. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *33*(4), 672–684. https://doi.org/10.1007/s11266-021-00424-9

Whitman, J. R. (2021). Does Covid-19 portend a shocking future for nonprofits? Canadian Journal of Nonprofit and Social Economy Research, 12(2), 8–12. https://doi.org/10.29173/cjnser.2021v12n2a547

Witmer, H., & Mellinger, M. S. (2016). Organizational resilience: Nonprofit organizations' response to change. *Work*, *54*(2), 255–265. https://doi.org/10.3233/wor-162303

Wildavsky, A. B. (1988). Searching for Safety. Transaction Books.

Wukich, C. (2013). Searching for Resilience [Review of Building Resilience: Social Capital in Post-Disaster Recovery; Designing Resilience: Preparing for Extreme Events, by D. P. Aldrich, L. K. Comfort, A. Boin, & C. C. Demchak]. Journal of Public

Administration Research and Theory: J-PART, 23(4), 1013–1019.

Author Biographies

Grace De Oro is a graduate of the University of Maryland Baltimore County's School of Public Policy where she completed this research and defended her dissertation in July of 2024 titled, *Bouncing Forward: Exploring Nonprofit Resilience through Emergency Management.* Grace works in higher education supporting students in historically minority groups and first-generation college students. She designed and instructed a first-year seminar in Fall of 2024 based on her dissertation research and this article on resilience.

Evaluation and Transformative Social Innovation: A Call for New Habits

Mieke Berghmans - KU Leuven
Joke Vandenabeele - KU Leuven

In this article we argue that dominant measurement and evaluation methods reduce the transformative potential of social innovation initiatives. Taking a pragmatist perspective on transformative social innovation and drawing on a commonly used distinction between simple, complicated and complex intervention logics, we demonstrate that habitually used monitoring and evaluation methods are often supported by a simple or complicated intervention logic. These intervention logics and related understandings of social change and knowledge production are incompatible with the ambition to realise transformation. Less dominant methods, which often are supported by a complex intervention logic, seem to be more apt, especially when they do not focus on adaptation alone, for the monitoring and evaluation of transformative social innovation initiatives. On the basis of this analysis, we plead for new habits of monitoring and evaluation and formulate an agenda for further action and research.

Keywords: Transformation – Social innovation – Monitoring and evaluation Pragmatism

Introduction

Practices and theories of social innovation are not new. Historically, social innovation can for example be observed in the development of cooperatives as an alternative to rampant 19th century capitalism, the initiation of neighborhood committees in response to the new urban question around the 1970's, or commoning initiatives, like community supported agriculture at the end of the 20th century. People have, in other words, time and again - without necessarily using the term 'social innovation' – deliberated, pursued, and achieved sustainable change (McGowan & Westley, 2015; Moulaert et al., 2017; Moulaert & MacCallum, 2019). While in the past social innovation was predominantly initiated and stimulated at the civil society or local community level, in recent decades government policy makers have become strong supporters of social innovation (Moulaert et al., 2017; Reynolds, et al., 2017). In a world that is confronted with "wicked problems', limited resources and ingenuity gaps" (McGowan & Westley, 2015, p. 54), social innovation has become an attractive pathway for dealing with complex societal challenges to which existing practices, rules and institutions do not seem to find sustainable answers. In this context, government policy makers have started to promote policies, regulatory frameworks, funding mechanisms (Nicholls & Edmiston, 2018) as well as research agendas (Moulaert & Mehmood, 2020) which support and stimulate social innovation.

The adoption of social innovation in governmental policy agendas on regional, national and local levels is met with scholarly scrutiny and critique. Scholars suggest that the transformative potential of social innovation initiatives is likely to weaken when these initiatives become part

and parcel of neoliberal policy strategies. The potential of bottom-up social innovation by civil society easily gets overlooked (Moulaert et al., 2017; Wright & Nyberg, 2019) as many of the contemporary social innovation policies focus on market creation and social business. Social innovation becomes co-opted in a hegemonic market discourse which depoliticizes problem framings and promotes a neoliberal agenda of economic growth (Fougère et al., 2017; Tesfaye & Fougère, 2022; Wittmayer et al., 2019). Moreover, as policy administrators promote dominant accountability mechanisms, which focus on metrics and indicators, social innovation initiatives are forced into dominant cognitive frameworks and institutional structures thus running the risk of inscribing themselves into the dominant order rather than transforming it (Moulaert et al., 2017; Novy, 2017; Oosterlynck et al., 2019; Reynolds et al., 2017).

In this article we delve deeper into this last issue: the idea that the transformative potential of social innovation initiatives gets reduced through the use of dominant monitoring and evaluation tools. This concern is often mentioned in the emerging body of scholarly work on transformative social innovation. Scholars agree that there exists a mismatch between dominant measurement and evaluation methods and the transformative potential of social innovation initiatives (Kok et al., 2023, Moulaert et al., 2017; Novy, 2017; Oosterlynck et al., 2019; Reynolds et al., 2017). In spite of a general consensus that this mismatch exists, little attention has however been paid to explaining precisely how these dominant measurement and evaluation tools reduce or inhibit the transformative potential of social innovation initiatives. Some scholars have pointed out that this mismatch results from an incompatibility between social innovation projects and conventional monitoring and evaluation tools. They indicate that the former consist of uncertain and complex processes that have their own "dynamics and multidimensional impacts" (Antadze & Westley, 2012, p. 143), whereas the latter rely on linear, cause-effect relationships and focus on the realisation of a plannable result (Antadze & Westley, 2012; Cunha & Benneworth, 2020; Oosterlynck et al., 2019). Others have suggested that conventional monitoring and evaluation tools promote a knowledge production which is "external to social innovation dynamics" and thus "reduce[s] the latter's transformative potential" (Oosterlynck et al. 2019, p. 225).

This article builds further on these observations. It presents a comprehensive analysis of how dominant measurement and evaluation methods may reduce the transformative potential of social innovation initiatives. On the basis of this analysis this article develops a plea for alternative habits of monitoring and evaluation that could contribute to (rather than hinder) transformative change in social innovation projects. It also formulates an action and research agenda for those who engage with monitoring and evaluation in transformative social innovation projects.

This article is structured as follows. We first briefly describe transformative social innovation. We point out the difference between transformative and non-transformative social innovation and identify some key characteristics and conditions for transformative social innovation. Drawing on a pragmatist perspective, we argue that transformative social innovation implies changing social habits and infrastructures, and, as such, requires intricately connected processes of collective action and collective learning. Next, we make a detailed exploration of the many monitoring and evaluation tools that are habitually used to assess the progress and impact made in social intervention projects today. We show how these monitoring and impact evaluation tools support and are supported by three different change logics (simple, complicated, and complex) and can be categorized accordingly. We confront these three types of evaluation methods with the ambition to realise transformative social innovation. In doing so, we show how dominant monitoring and evaluation tools that support and are supported by a simple and complicated change logic, are unlikely to support transformative change. We also suggest that even less dominant monitoring and evaluation tools which are founded on a complexity paradigm, might reduce the transformative potential of social innovation initiatives when these tools focus more on adaptation than on transformation. On the basis of this analysis we conclude with an action and research agenda for those who engage with the monitoring and evaluation of transformative social innovation.

Transformative Social Innovation

Transformative social innovation is "a process of changing social relations that involves the emergence and spread of new knowledge and practices that challenge, alter or replace the established institutions in a specific context" (Haxeltine et al., 2017, p. 61). In adopting this definition, we agree with other scholars (e.g., Dias & Partidario, 2019; Haxeltine et al., 2017) who make a clear distinction between the broader category of social innovation and the more specific notion of transformative social innovation, and who argue that transformation is not an intrinsic quality of social innovation. In and on itself, social innovation - which can for example be defined as "the design and implementation of new solutions that imply conceptual, process, product, or organizational change, which ultimately aim to improve the welfare and wellbeing of individuals and communities" (OECD, n.d.) - does not necessarily aim for change at the institutional level. It pertains to processes of changing social relations that can develop, spread and interact with established institutions. When however social innovation aims for transformation, it intends to alter existing institutions and related practices, organisations, and norms that "both constrain and enable social relations and established patterns of doing, organising, framing and knowing" (Strasser et al., 2019, p. 1,304). These institutions can refer to both formal institutions like organisations, laws, regulations, or policies, as well as to less formal institutions like customs, beliefs, norms, habits, or values (Strasser et al., 2019). Historical examples of transformative social innovation initiatives are – even though they were at the time not necessarily given that name – cooperatives and credit unions, neighbourhood committees, or new models of education, participation, and governance (Moulaert et al., 2017). Contemporary examples of social innovation initiatives with a transformative agenda are hackerspaces, transition towns, community supported agriculture, participatory budgeting, or Via Campesina (Avelino et al., 2019). These examples also demonstrate that contemporary social innovations are developed through a diversity of organizations and collaborations (Unceta et al., 2017). They can take many different forms and may emerge in public, private, and nonprofit organizations as well as in complex cross-sectoral collaborations (Sadabadi & Rahimi Rad, 2022; Shier & Handy, 2016).

There exists a widely shared scholarly consensus that transformative social innovation is the result of intricately connected processes of collective action and collective learning (De Blust et al., 2019; Dumitru et al., 2016; Jessop et al., 2013). Transformative social innovation cannot be achieved by change of individuals alone but requires the emergence of new values, new imaginaries, new practices, new ways of interacting, and new understandings at the collective level so that dominant institutions can be recreated and provided with new meanings (De Blust et al., 2019). In practice this often means that transformative social initiatives:

> start when a group of individuals come together to develop a common vision for social and institutional change, responding to perceived deficits or failures in current societal arrangements. Endorsing a set of alternative values, they set out to co-shape a reflexive and experimental space in which their vision may be realized as new social relations and configurations of practices […]. As the SI initiative develops, it provides a space in which these new values can take root, new interpersonal relations can be shaped and enabled […] and wherein both individual and collective empowerment can take place […] (Pel et al., 2020, p. 7).

The intricate connection between collective learning and collective action to achieve transformative change is further elaborated in Kotov and Pedanik (2016) who suggest that transformative social innovation requires "rehabiting." They argue that transformative social innovation requires changing (sets of) intricately connected conceptual, emotional, and behavioural social habits as well as the infrastructures that shape them.

It is important to underline that, when making this claim, Kotov & Pedanik (2016) do not rely on a "narrow" Cartesian (Pedersen & Dunne, 2020) notion of habit which refers to "repetitive

mindless routine" (Kotov & Pedanik, 2016, p. 139) at the level of the individual. Instead they rely on a Deweyan, pragmatist understanding of reality and identify habits as "general tendencies or predispositions acquired through inquiry and learning or (more commonly) through social interaction manifest in specific action while, typically, the actor is not aware of the predisposition" (Kotov & Pedanik, 2016, p. 143). Habits are in other words "dispositions that coordinate the relationship between human organism and his/ her environment" (Kotov & Pedanik, 2016, p. 144). In this sense habits exist at once at the individual and at the collective level. They are shaped in transaction with the environment, and thus also in transaction with collectively shared and intricately connected conceptual, emotional, and behavioural social habits as well as the infrastructures that support and are supported by these habits. In this, conceptual habits refer to "systems of meanings, beliefs, knowledge, discourses, along with cultural self-descriptions, norms, concepts of what a situation is about, what it means or how it should be dealt with [...], patterns of reasoning and argumentation that, once acquired, become unconscious, unquestioned, and are hard if not outright impossible to change" (Kotov & Pedanik, 2016, p. 150). Emotional habits are "general tendencies to emotionally react in a certain way in specific situations" (Kotov & Pedanik, 2016, p. 151). Behavioural habits are a) "related to the physical body, including the culturally regulated functions of the physical body; and b) skill, a way of doing or handling things and the infrastructure that supports these habits" (Kotov & Pedanik, 2016, p. 151). Infrastructures consist of "tools and artefacts..." that shape our life worlds and environment, which is "structured by a habit that supports the continuation of the habit and its manifestation through individual behaviour" (Kotov & Pedanik, 2016, p. 151) .

To transform these intricately connected social (conceptual, emotional and behavioural) habits and infrastructures, which generate "stability and preservation" (Kotov & Pedanik, 2016, p. 145) but can also be changed, it is not sufficient to target the individual alone. Indeed, Kotov and Pedanik (2016) refer to what Dewey has written on the importance of the transactional relationship between individuals and their environment. "No amount of preaching good will or the golden rule or cultivation of sentiments of love and equity will accomplish the results. There must be change in objective arrangements and institutions. We must work on the environment not merely on the hearts of men" (Dewey 1922:22 as cited in Kotov & Pedanik, 2016, p. 145). Nor will it be sufficient to formulate an independent rational argument that is not connected to emotional, behavioural habits or infrastructures; "thought which does not exist within ordinary habits lacks means of execution" (Dewey, 1922, 67).

The desire and need to evaluate and transform habits, which generally remain unnoticed and invisible, emerges in real life situations when a new factor – e.g. an emerging social issue, a new need, or a newly acknowledged problem – arises which disturbs or challenges a habit or generates a conflict between habits. In these situations the values that are embodied by the habit or habit sets, become visible and the question rises whether what is "at present valued – the intentional object of any habit involved – is really valuable" (Kennedy, 1955, p. 90). What is needed for transformative change are then collective processes of value judgment critical inquiry, new imaginations, and experimentation with new hypotheses to be tested in minds and in actions. This process can contribute to the transformation of collectively shared (sets of) conceptual, emotional, behavioural habits as well as the infrastructures that support and are supported by these habits.

Monitoring and Evaluation of Social Initiatives: it's not Neutral

Just like other social initiatives, transformative social innovation initiatives are confronted with both external and internal requests for monitoring and evaluation. When they are financed or supported by a government or external funders, transformative social innovation initiatives are likely to be confronted with external requests to make their impact and progress demonstrable for accountability reasons. At the same time these initiatives may also have internal drivers to systematically monitor and evaluate. They may for example want to engage

in an evaluation process to stimulate organizational learning or to demonstrate internal accountability. Evaluation, often in a more encompassing manner referred to as 'monitoring and evaluation' or 'monitoring, evaluation, accountability, and learning' can, in other words be done at the demand of different actors (e.g. funders, local authorities, scientists, Board members, etc.) and with different purposes (e.g. internal or external accountability, evidence building, comparison, fundraising, performance improvement, stimulating organizational learning and reflection processes, etc.) in mind. Tensions exist between these different evaluation goals and purposes (Chelimsky, 2006; Ebrahim, 2005; Regeer et al., 2016; Reinertsen et al., 2022).

Today the field of monitoring and evaluation is filled with a diversity of approaches, methods, and techniques. A selection of dishes on the contemporary monitoring and evaluation menu (Berghmans et al., 2014) are: social return on investment, randomised controlled trial, (quasi-) experimental methods, most significant change (Davies & Dart, 2005), principle-focused evaluation (Patton & Johnson, 2017), success case method (Brinkerhoff, 2003), innovation history (Douthwaite & Ashby, 2005), sensemaker (Cognitive Edge, 2023), process tracing (Beach & Pedersen, 2019), outcome harvesting (Wilson-Grau, 2018), outcome mapping (Earl et al., 2001), and contribution analysis (Mayne, n.d. & 2011). On this menu some dishes are more popular than others. There seems to be a broad societal preference for methods like randomised controlled trial or social return on investment (Archer & Tritter, 2000; Benjamin et al., 2023), but also theory-based evaluations have gained popularity (Mackenzie & Blamey, 2005; Van der Knaap, 2004). Other methods, like for example sensemaker or outcome harvesting, seem to have the status of 'connoisseur dishes'. They are methods that are not widely known (yet) but predominantly enjoyed and used by experts.

Many of these tools are presented as neutral and suitable for any setting or initiative. Consequently, the assessment of progress and impact of a transformative social innovation project might, at first sight, seem to simply be a matter of routinely choosing one tool from this plethora of evaluation tools and implementing it correctly. Evaluations are however not neutral or value-free (House & Howe, 1999). They can affect the social interventions that are the object of evaluation (Benjamin et al., 2023; Ebrahim, 2019; Ruff, 2021). Evaluations can for instance contribute to goal displacement. A strong evaluation focusing on specific outcomes can affect the actions of involved stakeholders who may feel that the quality of their performance is reflected in the evaluated outcome and who may therefore only start working towards the realisation of this outcome whilst overlooking other relevant aspects of the intervention (Benjamin et al. 2023; De Lancer Julnes, 2006). Participation in evaluation may also affect the cognitive understandings of involved stakeholders. The language used in evaluation may affect their perceptions, meanings, and understandings of the evaluated intervention and of social reality (Benjamin et al. 2023; Kirkhart, 2000; Patton, 2000). Participation in evaluation may also have a positive or negative effect on the participants' individual and collective feelings and affect their views of merit and worth (Benjamin et al., 2023; Froncek & Rohmann, 2019; Kirkhart, 2000). Equally it may impact the power dynamics and perceptions of privilege and power within a program (Gregory, 2000; McKegg, 2019; Weiss, 1993). It is therefore important to carefully consider what kind of monitoring and evaluation methods are suitable and compatible with the evaluated intervention.

Different monitoring and evaluation methods all have their own particularities and characteristics. Nevertheless, drawing on a commonly used distinction between simple, complicated, and complex intervention logics (Glouberman & Zimmerman, 2002; Patton & Johnson, 2017; Rogers, 2008; Stern et al., 2012; Westley et al., 2007), they can arguably, in an ideal type manner, be categorised as monitoring and evaluation methods that support and are supported by a simple, complicated or complex intervention logic. Below we further elaborate on this categorisation. We demonstrate how different methods are supported by different intervention logics and thus reflect different understandings of how social change is established and what kind of knowledge production is required to pursue, monitor, and evaluate this social change.[i] We also show how these logics, and the related understanding of
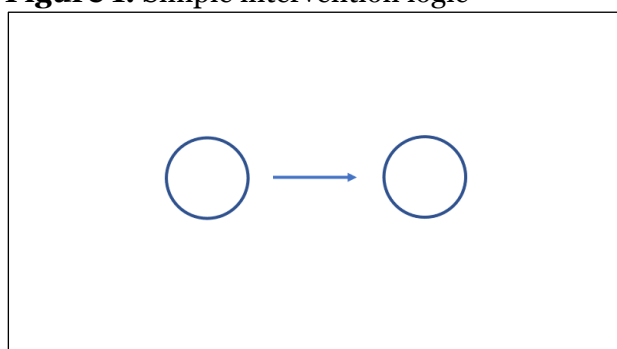
social change and of knowledge production, are not necessarily compatible with the ambition to realise transformative change.

> *Simple' and 'complicated' monitoring and evaluation methods and social innovation for transformation: not a perfect match*

Within the variety of monitoring and evaluation methods, a first group of methods that can be identified are the methods that support and are supported by a simple understanding of social change, such as randomised controlled trials, quasi-experimental methods or (social) return on investment methods. These methods necessitate establishing and measuring specific quantitative performance indicators prior to the project's commencement, enabling before-and-after comparisons to determine if and to what extent the project or program has achieved the anticipated or desired outcomes. Evaluation in these methods means measuring to which extent predefined and expected performance indicators have been realised as an effect of the planned intervention.

These methods start from the assumption that indicators of success can be set beforehand and that there exists a linear and proportional causal relationship between the intervention and the effect (Kurtz & Snowden, 2003; Snowden, 2005). They produce and reproduce an understanding of social change as the result of a single, predictable linear cause-effect relationship. Social change is seen as the result of a simple, 'if A, then B' intervention logic (see figure 1).

**Figure 1.** Simple intervention logic

These methods also produce and reproduce a specific understanding of knowledge and knowledge production for social change. The process of measuring, processing, and analysing the extent to which the predefined indicators have been realised is external to the social intervention. Knowledge production about the project happens in other words outside the dynamics of the project. More precisely, knowledge is produced by people, often methodological experts, who take the perspective of a spectator and observes objective facts. If the demonstrated performance on predefined indicators is satisfying, then the social intervention is likely to be understood as a 'recipe' for success. It is in other words assumed that an identical intervention could, in comparable conditions, lead to a similar success. If the measured performance on predefined indicators is not satisfying, then project organizers are invited to make changes and strategically test and evaluate another 'recipe'.

A second group of monitoring and evaluation methods are methods like, for example, mediation analysis and clustered randomised controlled trial (Lemire et al., 2020; Peck, 2012), contribution analysis[ii] (Mayne, 2011), or theory-based evaluations in which the theory underlying to an intervention – i.e. the "causal chain from inputs to outcomes and impact" (White, 2009, p. 271) – is mapped out and underlying assumptions are tested. This type of methods starts from the idea that the purposeful realisation of social change is complicated. They focus on building a theory of change that indicates how different causal mechanisms, conditions, and contextual factors might contribute to a desired outcome and on collecting

evidence about the results and the causal links between the different elements in these cause-effect chains. They are supported by and support an idea of social change as the result of causal chains of different elements that together, under specific conditions, contribute to an intended effect (Rogers, 2008) (see figure 2). Effects are in these methods, at least in principle, understood as predictable and pre-identifiable outcomes of a complicated chain of linear cause-and-effect relationships between different constitutive elements. If unintended effects are realised or if the projected theory of change does not work, one can try to adjust some of the elements in the chain or adjust the existing intervention theory (Berghmans et al., 2014; Poli, 2013; Rogers, 2008).

Figure 2. Complicated intervention logic
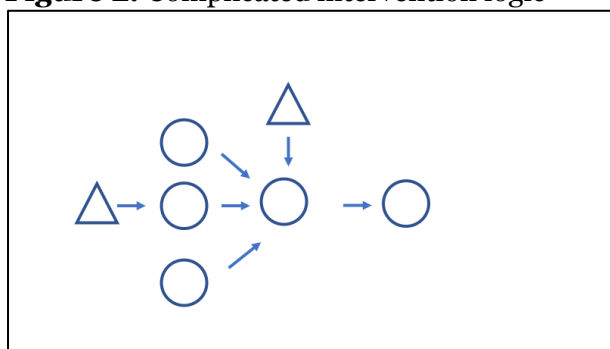
Knowledge production is, also in these methods supported by a complicated intervention logic, external to the social intervention itself. Theories are produced before the social intervention takes place and data about the causal chains are collected after implementation of the intervention.[iii] Knowledge is also produced strategically, in function of developing a more plausible or more effective set of causal chains.

As we already suggested, these two sets of monitoring and evaluation tools, respectively supported by a simple or complicated change logic, are very popular dishes in the monitoring and evaluation kitchen. With a Nobel prize for Duflo, Banerjee and Kremer, the presentation of randomized controlled trials as the 'gold standard', and the rise of 'simple' monitoring and evaluation methods in evidence based policy making and in non-medical research (e.g. Connolly et al., 2018), methods that are based on a simple intervention logic have become very popular and are increasingly institutionalized in the broader nonprofit sector (Benjamin et al., 2023). At the same time, the habitual use of these methods to measure social change is increasingly subject to criticism. Apart from pointing at the general methodological, ethical, and practical limitations of these methods, critical questions are raised as to whether these methods are really suitable for the evaluation of social interventions as these interventions are in most cases not standardized, have limited predictability, and are strongly influenced by contextual factors (Ebrahim, 2019; Rogers, 2008; van der Meulen Rodgers et al., 2020). In response to these criticisms, the methods which support and are supported by a complicated intervention logic, have become increasingly popular for the monitoring and evaluation of social interventions (Van der Knaap, 2004). It is generally felt that this second group of methods is more apt for the evaluation of social interventions as these methods allow one to work with intended as well as unintended effects and approach change as the result of a combination of activities and factors.

But does this also imply that these methods are more suitable for the evaluation of social interventions that aim for transformative social innovation? On the basis of the above analysis of methods that support or are supported by a complicated causal logic, we are inclined to say no. Of course transformative social innovation initiatives may, in practice and for a variety of reasons and purposes – such as for example to attract financial investment (Antadze & Westley, 2012) or to attain symbolic power (Nicholls, 2015) – choose to adopt a diversity of evaluation tools. This however does not mean that all these methods are fully conducive for

and compatible with the ambition to realise transformative social innovation. In fact, it can be argued that methods founded on a complicated causal logic as well as methods founded on a simple causal logic, are not a well-tuned match for transformative social innovation initiatives. First, whereas transformative social innovation requires processes in which collective action and collectively learning processes are intricately connected, learning is, in these evaluation methods, rather situated outside of the dynamics of the intervention and change processes. It implies a distant look of a spectator who, from a fixed reference point (of a preset theory of change or a preset indicator) neutrally observes the performance on an indicator, a contribution, or the causal link between an intervention and an effect. In this process of learning little room is given to the possibility of questioning existing habits and infrastructures, creating new imaginaries, or experimenting in mind and action with new hypotheses.

Second, and related to this, these evaluation methods are unlikely to stimulate innovation. They are based, as we demonstrated above, on the assumption that desirable effects and the pathway(s) towards this effect can, at least in principle, be identified beforehand and be predicted. This assumption is at odds with the "innovation uncertainty principle" (Morris, 2011) which refers to the idea that:

> the pursuit of innovation necessarily involves a venture into the unknown and if we try to pin these unknowns down too early in our process we may make it more difficult to recognize and realize good opportunities or solutions. If attempts are made to calculate the impact of every idea very early on in the process of its development, the result could be a meaningless and misleading number that may have disproportionate influence on the emergent process at precisely the wrong time (Morris, 2011, p. 3).

The use of 'simple' or 'complicated' evaluation methods implies pinning down measurable expected outcomes and the expected linear causal pathway(s) towards these outcomes. These predicted outcomes and pathways also acquire prescriptive value, as in these monitoring and evaluation methods success means optimally realising the desired effect through the planned causal pathways. In this manner, these evaluation methods tend to reproduce dominant standards, and dominant ways of thinking, feeling, or behaving through the intervention process. They are unlikely to stimulate risk taking, creativity, and innovation, let alone transformation. In this manner, the mentioned monitoring and evaluation methods might, in spite of their popularity, reduce the transformative potential of social innovation initiatives rather than supporting it.

> *'Complex' monitoring and evaluation methods and social innovation for transformation: the need to safeguard the transformative potential?*

The "complexity turn" in the social sciences (Urry, 2005) has contributed to the emergence of a third category of monitoring and evaluation methods which are supported by a complex intervention logic. According to this logic, social interventions – and the wider system in which these interventions are situated – have to be understood as open, dynamic networks of relationships between multiple elements which do not exert a linear influence on each other, but rather interact and adapt to the changing environment (Poli, 2013) (see figure 3).

Figure 3. Complex intervention logic
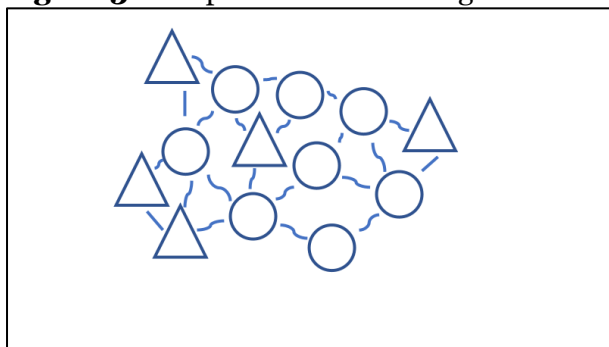
Key concepts in this complexity perspective on social reality are: adaptiveness and adaptive management, feedback loops, tipping points, iteration, emergence, equilibrium and so forth. According to this complex understanding of change, small inputs can lead to unplanned disproportionate effects. Effects can thus not be predefined, controlled, or managed in a planned way. However, change can be influenced by "learning to dance" with the complex system (Poli, 2013).

Drawing on this complex change logic, authors have developed and used different methods that can support the purposeful pursuit, monitoring, and evaluation of complex change (e.g. Derbyshire, 2019; Kurtz & Snowden, 2003; Rogers, 2009; Van der Merwe et al., 2019; Van Mierlo et al., 2010; Westley et al. 2007). These methods are, for example, most significant change (Davies & Dart, 2005; Ohkubo et al., 2022), sensemaker (Cognitive Edge, 2023; Van der Merwe et al., 2019) or outcome harvesting (INTRAC, 2017; Wilson-Grau, 2015, 2018). These methods are less widely known. It is mostly connoisseurs of the monitoring and evaluation kitchen who know, use, and appreciate these methods. In contrast to the previous two types of methods (which rely on a simple or complicated intervention logic) these monitoring and evaluation methods do not focus on evaluating whether the intervention went as planned or on evaluating whether desired effects were realised. Rather, they start from a given situation, try to identify important past changes, and make sense of how these changes have emerged. This process of real-time monitoring and evaluating the impact of certain interventions can then inform future courses of action.

In spite of their shared inscription into a general complexity logic, these methods seem to give different interpretations to the meaning of 'learning to dance with the complex system'. These interpretations move on a continuum between two poles of interpretation, namely a pole that focuses on adaptation and a pole that stresses transformation.[iv]

The first pole of interpretation, which stresses adaptation, can for example be observed in the works of Rogers (2009), Derbyshire (2019), and Dotson et al. (2008) but also in the popular examples of "organizing a children's party' (Cognitive Edge, n.d.) or "raising a child" (Glouberman & Zimmerman, 2002). This interpretation starts from the assumption that the system with which one has to learn to dance is relatively stable and knowable. It is, as Funtowicz and Ravetz (1994) would say, an "ordinary complex system" with a relatively stable[v] and clear finality or purpose. Though pathways towards this purpose are variable and cannot be predicted nor prescribed, there is a generally shared understanding of what is valued in regard to a successful social intervention. Social change can, according to this interpretation, purposefully be pursued by running small scale experiments. What is needed is regularly gathering feedback about changes caused by these experiments, and also collectively reflecting, together with involved actors and with the support of monitoring and evaluation methods, about emerging patterns and emerging outcomes. If these requirements are fulfilled small adaptations can be made in the direction of the shared finality or purpose. Monitoring and evaluation pertains, in this interpretation, to collectively making sense of and reflecting about

past changes and experimentations in order to make strategic adaptations towards a shared purpose.

A second pole of interpretation, which stresses *transformation*, is reflected in for example Kurtz and Snowden (2003), Lamboll and others (2021), and Fernandez-Gimenez et al. (2008). This orientation starts from the idea that the social intervention is an 'emergent complex system' (Funtowicz & Ravetz, 1994). In this interpretation the focus is not (only) on making strategic adaptations towards a shared purpose. The focus is on collaborative dialogue and social learning to further discover what a shared purpose could consist of. In this interpretation, it is assumed that people involved in the complex intervention have different values, different loyalties, and different perspectives; they do not have a shared purpose (yet). There might even exist contradictions and tensions between the interests of the individual people on the one hand and what is good for the emergent complex system on the other. Monitoring and evaluation in this second pole of interpretation is all about collectively making sense of one's own implicit assumptions and ideas and reflecting the norms and values that underpin these assumption and ideas. This collective reflection may then contribute to developing a new language, new values, new meanings, and new norms, thus creating a new common ground that may further inform and shape the finality of the social intervention.

Methods supported by a complex change logic allow for new discoveries that may emerge throughout the innovation process, and in this manner they stimulate the creativity and risk taking necessary for transformative social innovation. The iterative processes of collective reflection, experimentation, monitoring, and sensemaking, which may involve many stakeholders, may contribute to creating an open space of collective change and learning in which new infrastructures and new (sets of) collective conceptual, emotional, and behavioural habits can emerge.

It might however be a too generalised statement to say that all these methods, which all support social innovation, also support and stimulate transformative change. In fact, it can be wondered to what extent monitoring and evaluation methods that align more with the first pole of interpretation and focus predominantly on strategic adaptations in view of a shared purpose) will stimulate and support transformative change. Indeed, there seems to be a tension between (processes of) adaptation and transformation. Adaption and transformation, do not only imply a different understanding of change but are also supported by a different type of reflexivity (Dias & Partidario, 2019; Moore et al., 2018). Adaptation requires a strategical reflexivity which consists of rapid feedback loops to strategically steer the intervention into the 'right direction'. This type of reflexivity is more strategical and requires quick responses and is therefore different from the reflexivity that is needed for transformation. Transformation processes, on the other hand, require more and longer time "to understand what aspects of the existing system to break down and which to leverage to build an entirely new alternative system" (Moore et al., 2018, p. 11). They require a deep reflexivity at the institutional level – i.e. a willingness of all involved stakeholders to critically look at and reflect about dominant formal and informal institutions. In doing so they have a focus on developing new imaginaries and testing new collective ideas, emotions, behaviours, and infrastructures in mind and in action. This "deeper systemic reflexivity" (Moore et al., 2018), might not be stimulated in monitoring and evaluation processes that focus on strategic adaptation towards a shared finality. In fact, if such reflection processes are not generating discomfort, they are probably not contributing to transformative change:

> In a sense then, a transformative learning space is not a safe space at all. It is actually a training ground to move into these more dangerous spaces. It relies on creating a kind of temporary psychological safety in order for people to do the necessary work of unlearning, crossing scales, confronting diversity, and acknowledging positive and negative dynamics, but it also relies on making people uncomfortable enough to prepare them to move through these contested, unknowable systems with courage, resilience, and grace (Moore et al., 2018, 38).

If one adopts a 'complex' monitoring and evaluation method to assess a social innovation initiative that aims for transformation, it is therefore important to be attentive to the nature of the collective reflection, experimentation, and sensemaking processes in this method. It is important to make sure that these processes touch at the institutional level and take a critical stance towards dominant ways of understanding, reacting to, and interacting with and in the world. Only then can these processes, which are likely to also generate some discomfort among the involved stakeholders (Moore et al., 2018), bring to the surface a "re-habiting" of dominant (sets of) habits and infrastructures (Kotov & Pedanik, 2016) and stimulate the transformative potential of the social innovation initiative.

Conclusion and Agenda for Future Action and Research

In this article we have demonstrated how dominant monitoring and evaluation methods may reduce the transformative potential of social innovation initiatives. We have pointed out that many habitually used evaluation methods – such as randomized controlled trial, return on investment, or theory based evaluation – are supported by a simple or complicated intervention logic and rely on the idea of plannable and prescribable change. As such, these methods are at odds with Morris's (2011) innovation uncertainty principle. They do not give room for uncertainty, risk taking, creativity, or innovation, let alone for transformation. Next to that knowledge production processes in these methods are situated outside of the intervention. Learning in these evaluation processes implies taking the position of a spectator who, from a fixed reference point (of a preset theory of change or a preset indicator), observes the performance on an indicator or on the causal link between an intervention and an effect. This form of 'spectator learning' leaves little room for actively and collectively questioning social habits and infrastructures, for creating new imaginaries, and for collectively experimenting in mind and action with new imaginations and pathways to change. The intricate connection between collective learning and collective action, which is a condition for transformation, is inhibited in these (simple and complicated) evaluation methods. In this manner, these habitually used monitoring and evaluation methods tend to contribute to reproducing dominant institutions and habits, rather than challenging, altering, and questioning them.

Methods supported by a complex intervention logic, which are often known and used by 'connoisseurs' only, seem to be more suitable for the monitoring and evaluation of transformative social innovation initiatives. When these methods do not focus on strategic adaptation in view of a fixed purpose but generate – often uncomfortable and unsafe – processes of "deeper systemic reflexivity" (Moore et al., 2018), they may have the potential of enhancing rather than reducing the transformative potential of social innovation initiatives. They then hold the possibility of creating an open space of collective action and learning, involving all stakeholders, so that new infrastructures and new (sets of) collective conceptual, emotional and behavioural habits may emerge.

On the basis of the analysis made in this paper, we make a plea for what we call a 're-habiting of monitoring and evaluation' in the field of transformative social innovation. If we really want to stimulate a process of transformative social innovation or "a process of changing social relations that involves the emergence and spread of new knowledge and practices that challenge, alter or replace the established institutions in a specific context" (Haxeltine et al., 2017, p. 61), we should not just adopt 'old' monitoring and evaluation habits. It is crucial that we also re-habit the conceptual, emotional, behavioral habits and infrastructures that shape monitoring and evaluation itself.[vi]

It can however be wondered if the development of new complexity-based evaluation methods and approaches alone will lead to new monitoring and evaluation habits and infrastructures. As we have demonstrated above, a plethora of alternative monitoring and evaluation methods

which rely on a complexity paradigm, already have been developed in order to do more justice to complex interventions. Next to that scholars like Bamberger and colleagues (2015), Williams and Hummelbrunner (2010), and Patton (2019) have already proposed wider evaluation approaches which do more justice to complex interventions. Bamberger et al. (2015) for example propose a complexity-aware approach to evaluation. They claim that complexity-aware evaluations demand unpacking a complex intervention into different evaluable components which have to, in first instance and supported by appropriate evaluation designs, be assessed separately to then, in second instance, be reassembled for the evaluation of the overall program. These authors also plead for the adoption of new technologies, big data, and data analytics in complex evaluations. Williams and Hummelbrunner (2010) propose a systemic approach to evaluation. They claim that evaluators who engage in systemic evaluation need three orientations. They need to be reflective about "assumptions, mental models and values – and how they affect what we see or hear" (Reynolds et al., 2016, p. 668). They also have to pay "due attention to emergence and value differences from plans, as these can provide useful clues for improvements" and they need to adopt "think–act-think circularity" (Reynolds et al., p. 668). Patton (2019), with the Blue Marble evaluation, has, at his turn developed an evaluation approach oriented towards global systemic change and transformation. In fact, Patton's call to "transform evaluation to evaluate transformation," seems be a call for a re-habiting of evaluation habits. His suggestion that the transformation of evaluation requires "moving from a theory of change to a theory of transformation," moving from "an external, independent stance to a skin-in the game stance," and moving "beyond the evaluator competence to becoming world savy through ongoing learning" (Blue Marble Evaluation, n.d.) seems to very much in line with this paper's plea for 're-habiting monitoring and evaluation'.

The mere existence of complexity-aware evaluation methods and approaches might however not be enough to change existing social habits about monitoring and evaluation. It might also require a broader conscientization. Indeed, up until now, evaluation methods that are founded on a complex change logic, seem to be reserved for 'connoisseurs'. Also, new complexity-aware evaluation approaches, like the ones developed by Bamberger and colleagues (2015), Williams and Hummelbrunner (2010), and Patton (2019), seem to be discussed predominantly by evaluation professionals. At the same time, there continues to be a broad societal preference for classical evaluation methods that are founded on a specific understanding of change and produce and reproduce a specific 'spectator' understanding of knowledge and knowledge production. If we really want to re-habit monitoring and evaluation and change dominant ways of thinking, feeling, or behaving in monitoring and evaluation a more general, critical conscientization – of policy makers, practitioners as well as of the general public – about monitoring, evaluation, and transformation may be necessary.

Next to this call for action, or more specifically for critical conscientization, we end this article with two suggestions for future research. A first suggestion pertains to the differences and maybe even incompatibilities (Dias & Partidario, 2019) between adaptation and transformation in complex evaluation methods and approaches. We think that this is an insufficiently explored and under-theorized aspect in complexity-aware monitoring and evaluation methods and approaches. A second suggestion is to fully integrate a pragmatist philosophical perspective into future theory formation about monitoring and evaluation for transformative social innovation. Indeed, it can be observed that scholarly theories about monitoring and evaluation of uncertain and complex interventions, are often grounded in a complexity or systems perspective. As these complex or systems perspectives are rooted in natural sciences, these theories often seem to pay relatively little attention to questions that are also central in transformative social innovation, such as 'how do we collectively decide what is of value?', 'how is collective change and learning established?', or 'how can purposeful collective action be made possible when ends and values are unclear or contested?'. Though pragmatist notions like single, double, and triple loop learning and the "reflective practitioner" (Argyris, 1976; Schön, 1983) have already been integrated in reflections about transformative evaluation (e.g. Regeer et al., 2009) we think, in line with the argument formulated by Ansell and Geyer (2017), that a more serious engagement with (Deweyan) pragmatist notions such as

'habit', 'value', 'community of inquiry', 'intelligent action', and 'imagination' may contribute to a more profound theory formation about evaluation for transformation.

Notes

1. [i]In contrast to others (like for example Glouberman & Zimmerman, 2002) we do not see simplicity, complicatedness, or complexity as inherent characteristics of interventions or of social systems. Rather we approach simplicity, complicatedness, and complexity as "pragmatic perspectives" (Petticrew, 2011, p. 397) adopted by evaluation researchers to help describe and understand the interventions in question. Starting from this premise, we then argue that these pragmatic perspectives might not always be compatible with the ambition to realise transformative social innovation.

2. [ii]Contribution analysis aims to demonstrate to what extent observed results or outcomes are the result of a specific initiative rather than of other factors. By developing a theory of change, looking for evidence for that theory of change, and paying attention to other factors which may also have contributed to the realisation of specific outcomes, contribution analysis aims to generate evidence about the contribution made by a specific initiative (Mayne, n.d.).

3. [iii]In some cases, theories are produced in the course of or after an intervention. It could be argued that when these theories-in-use are being constructed, the knowledge production is not external to the social intervention itself.

4. [iv]It needs to be stressed that both poles are only poles on a wider continuum of interpretation. Further research is needed to explore whether theses poles point at the different ways in which monitoring and evaluation tools are used or if they point at the internal logics of monitoring and evaluation tools.

5. [v]This goal cannot be exactly defined beforehand and translated in measurable indicators. Nevertheless, it is relatively stable and unlikely to change when the social intervention unfolds. The stability of the goal can emerge in different manners. The goal can for example be stable because it is set and guarded by decision makers who 'own the project and the program theory' (e.g. professionals who are determined to pursue a specific goal). But it can also be stable because there exists a general consensus within the system of the social intervention on what precisely has to be achieved through the social intervention.

6. [vi]Kotov & Pedanik (2016) make a similar point for habits of problem solving. They argue that transformative social innovation often requires "a rehabiting also in the ways and methods of problem solving" (Kotov & Pedanik, 2016, p.147).

Disclosure Statement

References

Ansell, C. & Geyer, R. (2017). 'Pragmatic complexity' a new foundation for moving beyond 'evidence-based policy making'? *Policy Studies, 38*(2), 149-167. https://doi.org/10.1080/01442872.2016.1219033

Antadze, N., & Westley, F. R. (2012). Impact metrics for social innovation: Barriers or bridges to radical change? *Journal of Social Entrepreneurship, 3*(2), 133-150. https://doi.org/10.1080/19420676.2012.726005

Archer, M. S., & Tritter, J. Q. (Eds.). (2000). *Rational choice theory: Resisting colonization*. Psychology Press.

Argyris, C. (1976). Single-loop and double-loop models in research on decision making. *Administrative Science Quarterly, 21*(3), 363-375.

Avelino, F., Wittmayer, J. M., Pel, B., Weaver, P., Dumitru, A., Haxeltine, A., Kemp, R., Jorgensen, M. S., Bauler, T., Ruijsink, S., & O'Riordan, T. (2019). Transformative social innovation and (dis)empowerment. *Technological Forecasting and Social Change, 145*, 195-206. https://doi.org/10.1016/j.techfore.2017.05.002

Bamberger, M., Vaessen, J., & Raimondo, E. (Eds.). (2015). *Dealing with complexity in development evaluation: A practical approach*. Thousand Oaks: Sage Publications.

Beach, D., & Pedersen, R. B. (2019). *Process-tracing methods: Foundations and guidelines*. Michigan: University of Michigan Press.

Benjamin, L. M., Ebrahim, A., & Gugerty, M. K. (2023). Nonprofit organizations and the evaluation of social impact: A research program to advance theory and practice. *Nonprofit and Voluntary Sector Quarterly, 52*(1_suppl), 313S-352S. https://doi.org/10.1177/08997640221123590

Berghmans, M., Deprez, S., Celis, H., Van Hemelrijck, A., & Vandenabeele, J. (2014). Peilen naar de impact van sociaal-culturele praktijken. Brussels: Socius.

Blue Marble Evaluation (n.d.) Transformative alignment principle. Retrieved at https://bluemarbleeval.org/principles/operating-principles/transformational-alignment-principle/

Brinkerhoff, R. (2003). *The success case method: Find out quickly what's working and what's not*. Berrett-Koehler Publishers.

Chelimsky, E. (2006). The purposes of evaluation in a democratic society. In I. Shaw, J.C. Green & M.M. Mark.(Eds.) *The Sage handbook of evaluation* (33-55). London: Sage.

Cognitive Edge. (n.d.). How to organise a children's party? Retrieved from https://www.youtube.com/watch?v=Miwb92eZaJg

Cognitive Edge. (2023). About sensemaker. Retrieved from https://thecynefin.co/about-sensemaker/

Connolly, P., Keenan, C. & Urbanska, K. (2018). The trials of evidence-based practice in education: A systematic review of randomised controlled trials in education research 1980–2016. *Educational Research, 60*(3), 276-291.https://doi.org/10.1080/00131881.2018.1493353

Cunha, J. & Benneworth, P. (2020). How to measure the impact of social innovation initiatives? *International Review on Public and Nonprofit Marketing, 17*, 59–75.

Davies, R. & Dart, J. (2005). The 'Most Significant Change' (MSC) technique: A guide to its use.

De Blust, S., Schreurs, J., & Devisch, O. (2019). Towards a situational understanding of collective learning: A reflexive framework. *Urban Planning, 4*(1), 19–30. https://doi.org/10.17645/up.v4i1.1673

De Lancer Julnes, P. (2006). Performance measurement: An effective tool for government accountability? The debate goes on. *Evaluation, 12*(2), 219-235. https://doi.org/10.1177/1356389006066973

Derbyshire, J. (2019). Use of scenario planning as a theory-driven evaluation tool. *Futures & Foresight Science*, *1*(1), e1. https://doi.org/10.1002/ffo2.1

Dewey, J. (1922). *Human nature and conduct*. New York: Henry Holt.

Dias, J., & Partidário, M. (2019). Mind the gap: The potential transformative capacity of social innovation. *Sustainability, 11*(16), 4465. https://doi.org/10.3390/su11164465

Dotson, Y., Folkman, D., & Syam, D. (2008). Welcome to Cynefin: A framework for learning and action in the face of complexity and chaos. Conference proceedings of the 27th Annual Midwest Research-to-Practice Conference in Adult, Continuing, Community,

and Extension Education . Retrieved at
http://digitalcommons.wku.edu/cebs_pubs/17

Douthwaite, B. & Ashby, J. (2005). Innovation histories: A method for learning from experience, ILAC Brief 5. Retrieved at
https://www.betterevaluation.org/sites/default/files/ILAC_Brief05_Histories_0.pdf

Dumitru, A., Lema-Blanco, I. & García Mira, R. (2016). Social learning for transformative social innovation, TRANSIT Deliverable 2.3.

Earl, S., Carden, F., & Smutylo, T. (2001). Outcome mapping: Building learning and reflection into development programmes. Ottawa: International Development Research Centre (IDRC).

Ebrahim, A. (2005). Accountability myopia: Losing sight of organizational learning. *Nonprofit and Voluntary Sector Quarterly, 34*(1), 56-87.

Ebrahim, A. (2019). *Measuring social change. Performance and accountability in a complex world.* Stanford: Stanford Business Books.

Fernandez-Gimenez, M. E., Ballard, H. L., & Sturtevant, V. E. (2008). Adaptive management and social learning in collaborative and community-based monitoring: A study of five community-based forestry organizations in the western USA. *Ecology and Society, 13*(2), 4. http://www.jstor.org/stable/26267955

Fougère, M., Segercrantz, B., & Seeck, H. (2017). A critical reading of the European Union's social innovation policy discourse:(Re)legitimizing neoliberalism. *Organization, 24*(6), 819-843. https://doi.org/10.1177/135050841668517

Froncek, B., & Rohmann, A. (2019). "You get the great feeling that you're being heard but in the end you realize that things will be done differently and in others' favor": An experimental investigation of negative effects of participation in evaluation. *American Journal of Evaluation, 40*(1), 19-34.

Funtowicz, S. & Ravetz, J.R. (1994). Emergent complex systems. *Futures, 26*(7), 568-582. https://doi.org/10.1016/0016-3287(94)90029-9

Glouberman, S. & Zimmerman, B. (2002). Complicated and complex systems: What would successful reform of Medicare look like. *Discussion Paper,* 8. Commission on the Future of Health Care in Canada.

Gregory, A. (2000). Problematizing participation: A critical review of approaches to participation in evaluation theory. *Evaluation, 6*(2), 179-199. https://doi.org/10.1177/13563890022209208

Haxeltine, A., Pel, B., Wittmayer, J., Dumitru, A., Kemp, R., & Avelino, F. (2017). Building a middle-range theory of transformative social innovation: Theoretical pitfalls and methodological responses. *European Public & Social Innovation Review, 2*(1), 59–77. https://doi.org/10.31637/epsir.17-1.5

House, E., & Howe, K. R. (1999). *Values in evaluation and social research.* Thousand Oaks: Sage Publications.

INTRAC (2017). Outcome harvesting. Retrieved at
https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=web&cd=&ved=0CAIQ
w7AJahcKEwigr_jJ-
oaAAxUAAAAAHQAAAAAQAg&url=https%3A%2F%2Fwww.intrac.org%2Fwpcms%2
Fwp-content%2Fuploads%2F2017%2F01%2FOutcome-
harvesting.pdf&psig=AOvVaw0q7D-
FcoEgQgTtHF2tjRVN&ust=1689175163270577&opi=89978449

Jessop, B., Moulaert, F., Hulgård, L., & Hamdouch, A. (2013). Social innovation research: A new stage in innovation analysis? In F. Moulaert, D. MacCallum, A. Mehmood, & A. Hamdouch (Eds.), *The international handbook on social innovation. Collective action, social learning and transdisciplinary research* (pp. 110–130). Cheltenham: Edward Elgar.

Kennedy, G. (1955). The hidden link in Dewey's theory of evaluation. *Teachers College Record*, *56*(8), 1–7.

Kirkhart, K.E. (2000), Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation, 88*, 5-23. https://doi.org/10.1002/ev.1188

Kok, K. P. W., van der Meij, M. G, Wagner, P. Cesuroglu, T., Broerse, J. E. W., & Regeer, B. J. (2023). Exploring the practice of Labs for sustainable transformation: The challenge of 'creating impact'. *Journal of Cleaner Production, 388*, 135994. https://doi.org/10.1016/j.jclepro.2023.135994.

Kotov, K. & Pedanik, R. (2016). Social innovation as rehabiting. In C. J. Rodriguez Higuera & T.J. Bennett (Eds.), *Concepts for semiotics* (pp. 133-156). Tartu: University of Tartu Press.

Kurtz, C.F. & Snowden, D.J. (2003). The new dynamics of strategy. Sense-making in a complex and complicated world. *IBM Systems Journal, 42*(3), 462-448.

Lemire, S., Peck, L.R., & Porowski, A. (2020). The growth of the evaluation tree in the policy analysis forest: Recent developments in evaluation. *Policy Studies Journal, 48*(S47-S70). https://doi.org/10.1111/psj.12387

Lamboll, R., Nelson, V., Gebreyes, M., Kambewa, D., Chinsinga, B., Karbo, N., Rukonge, A., Sekeleti, M., Litaba Wakun'uma, W. Gutema, T.H., Henjewele, M., Kampanje-Phiri, J., Masikati-Hlanguyo, P., Quaye, W., Duah, S., Kivuyo, M., Nyanga, P., Akuffobea Essilfie , M., Asafu-Adjaye, N.Y., Clottey, V. & Martin, A. (2021). Strengthening decision-making on sustainable agricultural intensification through multi-stakeholder social learning in sub-Saharan Africa. *International Journal of Agricultural Sustainability*, *19*(5-6), 609-635. https://doi.org/10.1080/14735903.2021.1913898

Mackenzie, M., & Blamey, A. (2005). The practice and the theory: Lessons from the application of a theories of change approach. *Evaluation*, *11*(2), 151–168. https://doi.org/10.1177/1356389005055538

Mayne, J. (2011). Contribution analysis: Addressing cause and effect. In K. Forss, M. Marra & R. Schwartz (Eds.), *Evaluating the complex* (pp. 53-96). Transaction Publishers.

Mayne, J. (n.d.). Contribution analysis: An approach to exploring cause and effect. ILAC . Retrieved at https://nonprofitbuilder.org/storage/377/Contribution-analysis-An-approach-to-exploring-cause-and-effect-ILAC.pdf

McGowan, K. & Westley, F. (2015). At the root of change: The history of social innovation. In A. Nicholls, J. Simon & M. Gabriel (Eds.), *New frontiers in social innovation research* (pp. 52-68). London: Palgrave Macmillan.

McKegg, K. (2019). White privilege and the decolonization work needed in evaluation to support indigenous sovereignty and self-determination. *Canadian Journal of Program Evaluation, 34*(2), 357- 367.

Moore, M.-L., Olsson, P., Nilsson, W., Rose, L., & Westley, F. (2018). Navigating emergence and system reflexivity as key transformative capacities: Experiences from a Global Fellowship program. *Ecology and Society, 23*(2), 38.

Morris, L., (2011). *The innovation master plan: The CEO's guide to innovation*. Walnut Creek, CA: Innovation Academy.

Moulaert, F. & MacCallum, D. (2019). *Advanced Introduction to Social Innovation*. Cheltenham: Edward Elgar Publishing.

Moulaert, F. & Mehmood, A. (2020). Towards a social innovation (SI) based epistemology in local development analysis: lessons from twenty years of EU research. *European Planning Studies, 28*(3), 434-453. https://doi.org/10.1080/09654313.2019.1639401

Moulaert F., Mehmood, A., MacCallum, D., & Leubolt B. (2017). *Social innovation as a trigger for transformations. The role of research*. Luxembourg: Publications Office of the European Union.

Nicholls, A. (2015). Synthetic grid: A critical framework to inform the development of social innovation metrics. *CRESSI Working papers, 14*. Oxford: Creating Economic Space for Social Innovation (CRESSI) project. Retrieved at http://www.transitsocialinnovation.eu/resource-hub/synthetic-grid-a-critical-framework-to-inform-the-development-of-social-innovation-metrics-cressi-working-papers-142015

Nicholls, A. & Edmiston, D. (2018) Social innovation policy in the European Union. In R. Heiskala & J. Aro (Eds.), *Policy design in the European Union: An empire of shopkeepers in the making?* (pp. 161-190). Palgrave Macmillan.

Novy, A. (2017). Transformative social innovation. *SRE-Discussion Papers, 2017/05.* WU Vienna University of Economics and Business. Retrieved at https://EconPapers.repec.org/RePEc:wiw:wus009:5619

OECD (n.d.). Social innovation. https://www.oecd.org/cfe/leed/social-economy/social-innovation.htm

Ohkubo, S., Mwaikambo, L., Salem, R. M., Ajijola, L., Nyachae, P., & Sharma, M. K. (2022). Lessons learned from the use of the most significant change technique for adaptive management of complex health interventions. *Global Health: Science and Practice, 10*(1), e2100624. https://doi.org/10.9745/GHSP-D-21-00624.

Oosterlynck, S., Novy, A., & Kazepov, Y. (2019). Conclusion: Local social innovation and welfare reform. In S. Oosterlynck, A. Novy & Y. Kazepov (Eds.), *Local social innovation to combat poverty and exclusion* (pp. 217-228). Bristol: Policy Press.

Patton, M.Q. (2000). Overview: Language matters. *New Directions for Evaluation, 86,* 5-16.

Patton, M.Q. & Johnson, N.M. (2017). *Principles-focused evaluation: The GUIDE*. Guilford Press.

Patton, M.Q. (2019). *Blue marble evaluation. Premises and principles.* Guilford Press.

Peck, L. R. (2012). Laura R. Peck on "Failure of Intervention or Failure of Evaluation: A Meta-Evaluation of the National Youth Anti-Drug Media Campaign Evaluation." *Substance Use & Misuse, 47*(13–14), 1425–1426. https://doi.org/10.3109/10826084.2012.723553

Pedersen, M., & Dunne, S. (2020). Dewey's broad account of habit and its relevance for change management: A conceptual clarification with pragmatic illustrations. *Journal of Change Management, 20*(3), 247–263. https://doi.org/10.1080/14697017.2020.1755342

Pel, B., Haxeltine, A., Avelino, F., Dumitru, A., Kemp, R., Bauler, T., Kunze, I., Dorland, J., Wittmayer, J. & Jørgensen, M. S. (2020). Towards a theory of transformative social innovation: A relational framework and 12 propositions. *Research Policy*, 49(8), 104080. https://doi.org/10.1016/j.respol.2020.104080

Petticrew, M. (2011). When are complex interventions 'complex'? When are simple interventions 'simple'? *European Journal of Public Health, 21*(4), 397–398. https://doi.org/10.1093/eurpub/ckr084

Poli, R. (2013). A note on the difference between complicated and complex social systems. *Cadmus Journal, 2*(1), 142-147.

Regeer, B. J., de Wildt-Liesveld, R., van Mierlo, B., & Bunders, J. F. (2016). Exploring ways to reconcile accountability and learning in the evaluation of niche experiments. *Evaluation, 22*(1), 6-28.

Regeer, B. J., Hoes, A.-C., van Amstel-van Saane, M., Caron-Flinterman, F. F., & Bunders, J. F. G. (2009). Six guiding principles for evaluating mode-2 strategies for sustainable development. *American Journal of Evaluation, 30*(4), 515-537. https://doi.org/10.1177/1098214009344618

Reinertsen, H., Bjørkdahl, K., & McNeill, D. (2022). Accountability versus learning in aid evaluation: A practice-oriented exploration of persistent dilemmas. *Evaluation*, 28(3), 356-378.

Reynolds, M., Gates, E., Hummerbrunner, R., Marra, M., & Williams, B. (2016). Towards systemic evaluation. *Systems Research and Behavioral Science, 33*(5), 662-673.

Reynolds, S. Gabriel, M. & Heales, C. (2017). Social innovation policy in Europe: where next? Retrieved at https://www.siceurope.eu/sites/default/files/field/attachment/social_innovation_policy_in_europe_-_where_next.pdf

Rogers, P. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation, 14*(1), 29-48. https://doi.org/10.1177/1356389007084674

Rogers, P. (2009). Matching impact evaluation design to the nature of the intervention and the purpose of the evaluation. *Journal of Development Effectiveness*, 1(3), 217-226. https://doi.org/10.1080/19439340903114636

Ruff, K. (2021). How impact measurement devices act: The performativity of theory of change, SROI and dashboards. *Qualitative Research in Accounting & Management, 18*(3), 332-360. https://doi.org/10.1108/QRAM-02-2019-0041

Sadabadi, A. & Rahimi Rad, Z. (2022). How can cross-sector partnership promote social innovation? *Systemic Practice and Action Research, 35,* 1-20.

Schön, D. (1983). *The reflective practitioner: How professionals think in action.* New York: Basic Books.

Shier, M. L., & Handy, F. (2016). Cross-sector partnerships: Factors supporting social innovation by nonprofits. *Human Service Organizations: Management, Leadership & Governance, 40*(3), 253–266. https://doi.org/10.1080/23303131.2015.1117556

Snowden, D. (2005). Strategy in the context of uncertainty. *Handbook of Business Strategy, 6*(1), 47-54.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). Broadening the range of designs and methods for impact evaluations. *Working Paper, 38.* London: Department for International Development (DFID).

Strasser, T., de Kraker, J., Kemp, R. (2019). Developing the transformative capacity of social innovation through learning: A conceptual framework and research agenda for the roles of network leadership. *Sustainability,* 11(5), 1304. doi.org/10.3390/su11051304.

Tesfaye, L.A., & Fougère, M. (2022). Frugal innovation hijacked: The co-optive power of co-creation. *Journal of Business Ethics 180,* 439–454. https://doi.org/10.1007/s10551-021-04883-4

Unceta, A., Castro-Spila, J., & García Fronti, J. (2017). The three governances in social innovation. *Innovation: The European Journal of Social Science Research, 30*(4), 406–420. https://doi.org/10.1080/13511610.2017.1279537

Urry, J. (2005). The complexity turn. *Theory, Culture & Society, 22*(5), 1-14. https://doi.org/10.1177/0263276405057188

Van der Knaap, P. (2004). Theory-based evaluation and learning: Possibilities and challenges. *Evaluation,* 10(1), 16-34. https://doi.org/10.1177/1356389004042328

Van der Merwe, S. E., Biggs, R., Preiser, R., Cunningham, C., Snowden, D. J., O'Brien, K., Jenal, M., Vosloo, M., Blignaut, S., & Goh, Z. (2019). Making sense of complexity: Using SenseMaker as a research tool. *Systems, 7*(2), 25. http://dx.doi.org/10.3390/systems7020025

van der Meulen Rodgers, Y., Bebbington, A., Boone, C., Dell'Angelo, J., Platteau, J. P., & Agrawal, A. (2020). Experimental approaches in development and poverty alleviation. *World Development, 127,* 104807. https://doi.org/10.1016/j.worlddev.2019.104807

Van Mierlo, B., Arkesteijn, M., & Leeuwis, C. (2010). Enhancing the reflexivity of system innovation projects with system analyses. *American Journal of Evaluation, 31*(2), 143-161. https://doi.org/10.1177/1098214010366046

Weiss, C.H. (1993). Where politics and evaluation research meet. *Evaluation Practice, 14*(1), 93-106.

Westley, F., Zimmerman, B. & Patton, M. (2007). *Getting to maybe. How the world is changed.* Penguin Random House.

White, H. (2009). Theory-based impact evaluation: principles and practice. Journal of Development Effectiveness, 1(3), 271–284. https://doi.org/10.1080/19439340903114628

Williams, B. & Hummelbrunner, R. (2010). *Systems concepts in action: A practitioner's toolkit.* Stanford Business Books.

Wilson-Grau, R. (2015). Outcome harvesting. Better evaluation. Retrieved from http://betterevaluation.org/plan/approach/outcome_harvesting.

Wilson-Grau, R. (2018). *Outcome harvesting. Principles, steps, and evaluation application.* Information Age Publishing.

Wittmayer, J.M, Backhaus, J. Avelino, F., Pel, B, Strasser, T, Kunze, I., & Zuijderwijk, L. (2019). Narratives of change: How social innovation initiatives construct societal transformation. *Futures, 112,* 102433. https://doi.org/10.1016/j.futures.2019.06.005.

Wright, C., & Nyberg, D. (2019). Climate change and social innovation. In G. George, T. Baker, P. Tracey & H. Joshi (Eds.), *Handbook of inclusive innovation: the role of organizations, markets and communities in social innovation* (pp. 47-60). Cheltenham: Edward Elgar Publishing.

Author Biographies

Mieke Berghmans
Mieke Berghmans is a postdoctoral assistant affiliated to the research unit Education, Culture, and Society of the KU Leuven (Belgium). She has a broad interest in contemporary governance and performance issues in education and the third sector. She works, teaches and does research on the design and development of innovative educational practices.

Joke Vandenabeele
Joke Vandenabeele is an associate professor social and cultural pedagogy at the Education, Culture and Society department, KU Leuven (Belgium). Her research focuses on how transformative innovation is enacted within practices of non-formal education, community education and public pedagogy.

The Evaluation Capacity of Czech Nonprofit Social Services Organizations

Salvatore P. Alaimo – Grand Valley State University
Jakub Pejcal – Masaryk University
Simona Smrčková – Charles University

This study explores the capacity of Czech nonprofit social services organizations to have their services evaluated. These organizations are often required to evaluate their services by their national, regional governments, and/or the EU Social Investment Fund, however they are challenged with evaluation capacity building (ECB). This study organizes the ECB literature around six dimensions and draws on ECB studies, frameworks, and checklists. It examines the exploratory question of evaluation capacity through a national survey of, and semi-structured interviews with, directors of these organizations. Results demonstrate that the challenges for evaluation capacity are similar to previous ECB studies in other countries. However, in this study they are somewhat driven by directors' social constructions of evaluation which are influenced by government guidelines and standards and their limited understanding of ECB and perception for their role in it. Conclusions are accompanied by recommended practice.

Keywords: Evaluation capacity building; Evaluation; Czech Republic; Nonprofit management; Social services

Introduction

Nonprofit organizations are challenged with demonstrating program effectiveness from government, philanthropic foundations, accreditation organizations, and other stakeholders. Many lack capacity, "the human capital (skills, knowledge, experience, etc.) and financial/material resources" necessary for evaluation (Boyle, Lemaire, & Rist, 1999, p. 5), thus requiring evaluation capacity building (ECB), "the intentional work to continuously create and sustain overall organizational processes that make quality evaluation and its uses routine" (Stockdill, Baizerman, & Compton, 2002, p. 1). Czech nonprofit management and evaluation continue in their professionalization, largely driven by the evolution of Czech civil society, sector growth since the Velvet Revolution in 1989, and an evolving relationship with government. The Ministry of Labor and Social Affairs (MoLSA) establishes, regulates, and evaluates social services against their fifteen standards (MoLSA, 2016).

The approach to this study is deductive by drawing from existing evaluation and ECB literature, and frameworks used for ECB research, to develop survey and interview instruments used to

assess ECB in participating organizations within the six dimensions. There has not been such a study that examines ECB in Czech nonprofit organizations, nor many, in other former Soviet Bloc nations. This study seeks to answer the exploratory research question, within the Czech context: What influences the state of evaluation and ECB in nonprofit social services organizations?. Supporting questions include, What challenges do these organizations face for the evaluation of their services? and How does the relationship with government impact their understanding of evaluation as a concept and process, and ultimately impact ECB? For consistency in participant understanding of evaluation, we defined it in the survey and interview as any assessment of the impacts and overall benefits or gains provided by services to their recipients. In this study we did not consider mere monitoring (indicator tracking), employee appraisals, feedback monitoring, audits, compliance with standards, etc., as evaluations in and of themselves. While these activities can be important components of an overall evaluation, they themselves do not constitute a comprehensive assessment of the impact of services provided on individual clients/consumers (service outcomes), which is primarily how we generically frame evaluation in this study, with the understanding that there exist many specific types of evaluation i.e. outcome, participatory, empowerment, etc. The phrase "program evaluation" is typically used in the U.S., Canada, Australia and other nations, however, in the Czech language and context, the term "program" typically refers to an employee training program, hence they instead use the term "services." This study refers to "program evaluation," because that's what the literature refers to but it is used interchangeably here because it examines the evaluation of Czech social services.

This article organizes the literature on ECB around six dimensions, then presents the Czech context including background on the nonprofit sector, social services, and the extent to which evaluation has progressed. This is followed by the methods section which describes the study population, survey development and distribution, semi-structured interviews, along with data collection and analysis. The results are presented within the guide of the six dimensions. The discussion section follows, and the authors provide limitations of this study, directions for future research, and conclusions accompanied by corresponding recommended practice.

Evaluation Capacity Building

Organizational Context

ECB remains a relatively new concept, as its body of knowledge, mostly from the United States and Canada, is about twenty-five years old, but it has substantially evolved and expanded in that time. It is context dependent and stakeholder driven, and therefore is multi-dimensional; dynamic; has economic, sociological, and psychological implications; and is different in size, scope, factors, and components from one organization to another. Figure 1 shows contextual factors for ECB:

While Figure 1 may not be exhaustive in displaying such factors, it is important to note that all impact each other as well as ECB. For example, if an organization has *historically* prioritized program evaluation, then it is likely to have worked towards embedding it *culturally* over time and therefore ensured it has had the *financial* resources and proper *structures* necessary to *practically* engage in it. An understanding of these factors is necessary for managing the internal and external organizational contexts for evaluation, and being aware of "power hierarchies, administrative culture, and decision-making processes" (Volkov & King, 2007, p. 1); acquiring stakeholder understanding and agreement on the demand for evaluation (Festen & Philbin, 2007; Volkov & King, 2007;); balancing the demand and supply for evaluation (Boyle, LeMaire, & Rist, 1999); and determining how receptive the organization is to change (Grudens-Schuck, 2003;

Volkov & King, 2007). This thorough and comprehensive understanding is necessary for meaningful stakeholder participation in ECB, planning the evaluation, and ensuring that results and process are used for learning and improvement (Bryson & Patton, 2015; Fitzpatrick, Sanders, & Worthen, 2011; Patton, 2022).

Figure 1. Contextual Factors for ECB
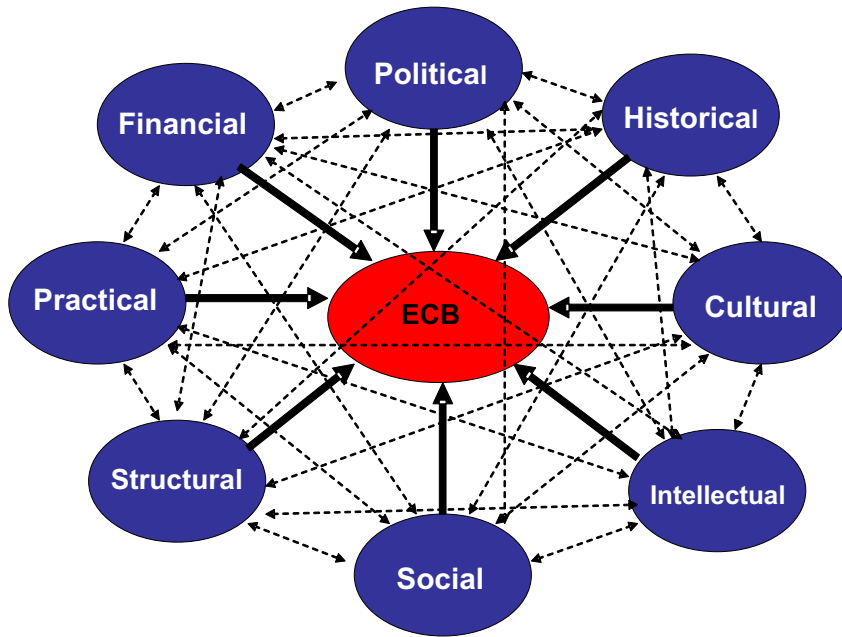
Dimensions for ECB

The existing literature on ECB including ECB research studies, ECB frameworks, ECB checklists and evaluability assessments all used to guide this study, and serve as the basis for the survey and interview instrument development, are shown with their sources below in Table 1:

Table 1. ECB Dimensions

Dimension	Recommended Practice	Sources
Leadership	• Understanding program evaluation • Taking responsibility for program evaluation • Managing political frame for program evaluation • Influencing organizational culture	Alaimo, 2008; Bamberger & Mabry, 2020; Compton, Glover-Kudon, Smith, & Eden Avery, 2002; Cousins, Goh, Elliott, & Bourgeois, 2014; Duigan, 2003; Festen & Philbin, 2007; Fitzpatrick, Sanders, & Worthen, 2011; Grudens-Schuck, 2003; Imas & Rist, 2009; Newcomer, Hatry, & Wholey, 2015; Preskill & Boyle, 2008; Sonnichsen, 1999; Stockdill, Baizerman, & Compton, 2002; Stufflebeam, 2002; Volkov & King, 2007; Wade & Kallemeyn, 2019;

		Williams & Hawkes, 2003
Organizational Culture	• Developing and maintaining an organization culture conducive to program evaluation • Developing and maintaining an organization culture conducive to learning	Baizerman, Compton, & Stockdill, 2002; Cousins, Goh, Elliott, & Bourgeois, 2014; Grudens-Schuck, 2003; Preskill & Boyle, 2008; Sanders, 2003; Sonnichsen,1999; Williams & Hawkes, 2003
Organizational Learning	• Ensuring program evaluation is a learning process where results are used for change and improvement • Ensuring ECB is a learning process where results are used for building capacity towards sustainable evaluation practice • Integrating the program evaluation and ECB processes so they inform each other • Learning takes place from both results and the evaluation and ECB processes	Bamberger & Mabry, 2020; Bastoe, 1999; Bourgeois & Cousins, 2013; Compton, Glover-Kudon, Smith, & Eden Avery, 2002; Cousins, Goh, Elliott, & Bourgeois, 2014; Duigan, 2003; Festen & Philbin, 2007; King, 2007; Mattesich, 2003; Monroe, M.C. et al., 2005; Newcomer, Hatry, & Wholey, 2015; Patton, 2008; Preskill & Boyle, 2008; Preskill & Torres, 1999; Sonnichsen, 1999; Stockdill, Baizerman, & Compton, 2002; Taut, 2007; Torres & Preskill, 2001
Resources	Providing the necessary human, financial, physical, informational resources and time dedicated to program evaluation	Bamberger & Mabry, 2020; Bourgeois & Cousins, 2013; Compton, Glover-Kudon, Smith & Eden Avery, 2002; Cousins, Goh, Elliott, & Bourgeois, 2014; Festen & Philbin, 2007; Stockdill, Baizerman, & and Compton, 2002; Stufflebeam, 2002; Volkov & King, 2007; Wade & Kallemeyn, 2019
Systems and Structures	Establishing systems and structures to effectively and efficiently evaluate programs	Bamberger & Mabry, 2020; Bastoe, 1999; Bourgeois & Cousins, 2013; Compton, Glover-Kudon, Smith, & Eden Avery, 2002; Duigan, 2003; Festen & Philbin, 2007; Preskill & Boyle, 2008; Volkov & King, 2007
Program Evaluability	• Assessing program readiness • Establishing a program theory of change • Developing a program logic model	Bamberger & Mabry, 2020; Chen, 2015; Imas & Rist, 2009; Knowlton & Phillips, 2013; McLaughlin & Jordan, 2015; Patton, 2008; Trevisan, 2007; Wholey, 2015
Existing ECB frameworks, studies, surveys, checklists, and evaluability assessments	• Collective recommended practice for what comprises ECB Note: Combined with above literature to develop survey	Bourgeois & Cousins, 2013; Carman & Fredericks, 2008; Fierro & Christie, 2017; Hudib & Cousins, 2022; Stufflebeam, 2002; Taylor-Ritzler et al., 2013; Trevisan, 2007;

Volkov & King, 2007; Wade &
Kallemeyn, 2020

These dimensions are not mutually exclusive, with some relationships being co- dependent and others being symbiotic, but they all collectively drive ECB. For each of them identified in column one, the authors identified the recommended practice in column two conveyed by the sources in column three. Each of these dimensions for ECB are discussed in more detail in the following sections.

Leadership

Leadership in the context of this study refers to the actions and decisions of top nonprofit executives, typically in the Czech Republic referred to as directors. It plays an important role in managing both the internal and external organizational context by initiating, catalyzing, and ensuring ECB takes hold in their organization (Alaimo, 2008; Preskill & Boyle, 2008; Volkov & King, 2007; Wade & Kallemeyn, 2019). One of the most important aspects is managing evaluation's existence within a political environment. Program evaluation involves making a value judgment for a program and decisions on changing it, improving it, etc. and involves multiple stakeholders each with their own assumptions and values, so it is therefore inherently a political process (Bamberger & Mabry, 2020a; Fitzpatrick, Sanders, & Worthen, 2011; Taylor & Balloch, 2022). Bamberger and Mabry sum it up well, "Evaluation is the most politically challenging of all approaches to inquiry because it often confirms or confronts values, personal preferences, and political agendas" (2020a, p. 95). Leadership also decides whether an organization will attempt to satisfy external demands solely for compliance, accreditation, reporting, or funding, or they will balance such responses of those external pulls by integrating them with an internal push (Alaimo, 2008) for evaluation. If such demands are not integrated into ECB, then organizations run the risk of not learning, not improving their programs, and rendering the evaluation results not worth the cost (Newcomer, Hatry, & Wholey, 2015).

Organizational Culture

Organizational culture is defined as, "A pattern of shared basic assumptions that the group has learned as it solved problems of external adaptation and internal integration, that has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think and feel in relation to these problems" (Schein, 1992, p. 12). If an organization has the proper resources, skills and other capacity issues covered to engage in program evaluation, but their culture is not conducive to change, does not embrace evaluation and does not prioritize learning, then program evaluation and ECB are not likely to occur. Some scholars have focused on the importance of mainstreaming program evaluation in the organization to where its priority, resourcing, attention, and engagement are on par with the other routine management functions (Duigan, 2003; Runnels, Andrew, & Rae, 2017: Sanders, 2003; Williams & Hawkes, 2003) and it becomes institutionalized (Stufflebeam, 2002). One indicator of when this happens is when program evaluation becomes an uncontested activity within the organization, stakeholders view it as a part of sound management, and it is reflected through the governance of the organization (Toulemonde, 1999).

Organizational Learning

Organizational learning occurs when "…individuals within an organization experience a problematic situation and inquire into it on the organization's behalf" (Argyris & Schon, 1996, p. 16). Scholars agree that program evaluation and ECB must be learning processes for them to be

effective (Monroe, et al., 2005; Parsons, Lovato, Hutchinson, & Wilson, 2016; Preskill & Boyle, 2008; Torres & Preskill, 2001). That effectiveness is driven by learning what works and what does not work in their programs and using information for improvement. Organizational learning is the catalyst for the inherent, potential, symbiotic relationship between ECB and program evaluation. This is where program evaluation results and learning from the process of evaluation are used to inform the work in ECB. In this relationship the learning from the program evaluation process is called process use (Patton, 2022, p. 214). Process use can contribute to evaluation capacity building through stakeholder participation resulting from increased knowledge and changed attitudes. ECB in turn improves and enhances program evaluations to be more efficient, relevant, robust, holistic, and comprehensive. Ongoing learning of this kind helps drive a culture of continuous improvement. The use of results also influences organizational and program thinking and decision making through revisions, improvements, and judgement of program worth (Bourgeois & Cousins, 2013; Cousins, Goh, Elliott, & Bourgeois, 2014).

Resources

Program evaluation like any other management function within an organization requires human, financial, physical, and informational resources. The typical question when it comes to staffing to conduct the evaluation is whether to build capacity from within and dedicate a full-time person or persons to evaluation or hire an external evaluator. Often this decision is driven by the size of the organization, its ability to afford outside help and whether funds have been dedicated to the evaluation function. In some cases, a hybrid form of staffing is employed where the internal and external evaluators work together on the evaluation. ECB's concerns also include building evaluation expertise within the organization through staff training and professional development (Sonnichsen, 1999). Bamberger and Mabry (2020b) recommend strengthening evaluation knowledge and skills in several groups including organizations that fund evaluations; evaluators; government; and other organizations that use the results of evaluations.

Insufficient funds and the need for dedicated funding for program evaluation are often lamented (Bamberger & Mabry, 2020a; Bourgeois & Cousins, 2013; Volkov & King, 2007). One reason may be that program evaluation simply is not a priority in the organization. Organizations tend to budget for things they deem important. Another reason is the organization's belief that allocating funds from their budget for program evaluation takes money away from their core service and their ability to serve their clients/consumers (Carman & Fredericks, 2008). Lastly, sometimes the very external stakeholders that demand evaluation inadequately fund it, if at all.

Systems and Structures

Systems and structures within the organization that support evaluation practice are necessary for ECB to be successful. One example is to develop and implement a purposeful long-term, ECB strategy and plan for the organization (Bamberger & Mabry, 2020b; Duigan, 2003; Volkov & King, 2007). Accompanying that plan should be an evaluation plan developed in consultation with senior managers including a needs assessment (Bourgeois & Cousins, 2013). These plans should be incorporated into the organization's strategic plan to ensure priority, implementation, learning and improvement, and such inclusion utilizes program evaluation results as a feedback loop to inform planning (Compton, Glover-Kudon, Smith & Eden Avery, 2002; Festen & Philbin, 2007). Communication systems and peer learning structures for socialization around evaluation are also recommended (Volkov & King, 2007).

Program Evaluability

ECB also exists at the program level, as not all programs are ready to be effectively evaluated. Important factors include clarification of the program design and possible need to redesign it; understanding of the demand for and feasibility of program evaluation; consensus on program goals and desired outcomes (Wholey, 2015); increased stakeholder understanding of the program; the specification of the program's theory (Patton, 2022); and the development of a program logic model (Knowlton & Phillips, 2013; Trevisan, 2007). Conceptualizing the program's theory of change impacts how it is implemented, understood, talked about, and improved and typically is a positive contributor to the use of evaluation results (Patton, 2022). While a theory of change model is conceptual, it is an important foundation for the program logic model which has practical application (Knowlton & Phillips, 2013). Program logic models are visual portrayals, often using columns with shapes and directional arrows, like a flow chart, of how programs operate and how they are intended to result in desired outcomes for participants.

These dimensions are not mutually exclusive, and their significance is realized in how they all influence each other. Organizations are recommended to address all dimensions for a comprehensive approach to ECB.

The Czech Context

Nonprofit Sector

The Czech Republic's nonprofit sector has grown substantially since the Velvet Revolution in 1989 that ended Communist one-party rule in the country and preceded the self-determined split of Czechoslovakia into the independent countries of the Czech Republic and Slovakia on January 1, 1993. There were approximately 2,000 registered nonprofits in 1989 serving households (the core of the nongovernmental nonprofit sector), rapidly growing to more than 150,000 in 2021 (Czech Statistical Office, 2022a). This growth is seen by some as part of the evolution of Czech civil society (Green, 1999; Pospisil, Navratil, & Pejcal, 2015; Potucek, 2000). It is also driven in part by significant laws governing the nonprofit sector since the early to mid 1990's. The legal forms of Czech nongovernmental, nonprofit organizations are primarily regulated by the Civil Code, and they include registered association, branch of association, foundation, philanthropic fund, public benefit company, church organization, and registered institution (Ministry of the Interior, 2012). According to the Satellite Account of Non-profit institutions, the revenue sources for Czech nonprofit organizations constitute public resources (36.4%); revenues from the non-market production (21.8%); revenues from the market production (10%); the work of volunteers (9.1%); corporate donations (8%); foreign (including EU) funding (7%); membership fees (4.7%); and personal donations (4%) (Czech Statistical Office, 2022b). Czech nonprofits were presented with new project and funding opportunities when the country joined the European Union in 2004, and their significance is seen in their share of the sector's revenue. The importance of the sector is noted in official Czech government documents including *Strategic Framework Czech Republic 2030* (Office of the Government of the Czech Republic, 2017) focused on sustainable development, and the Czech Government's Strategy for Cooperation Between Public Administration and Nongovernmental Non-profit Organizations 2021-2030 who's vision is one where nonprofit organizations are "a stable and strong partner of public administration in meeting the needs of the Czech society" (Office of the Government of the Czech Republic, 2021, p. 37).

Social Services

The Act on Social Services number 108 in 2006 regulates "conditions governing assistance and support to physical persons (individuals) in adverse social situations provided through social services, conditions governing the issue of the authorization for the social services provision, execution of public administration in the field of social services, inspection of the social services provision and prerequisites for performance of social services activities" (MoLSA, 2022a). Furthermore, the Act regulates "prerequisites for execution of the profession of a social worker performing activities in social services (a social services worker)" (MoLSA, 2022a). More than 8,750 registered social services are currently provided in the Czech Republic (MoLSA, 2022b), and most (more than 5,400) are provided by 1,055 non-governmental, nonprofit organizations. A smaller part of social services is provided by organizations established by the public sector (more than 2,880), while only a fraction is provided by for profit organizations (approximately 410) (MoLSA, 2022b).

In spite of growth, advancements and improvements by government and the social services providers over the past ten years, the MoLSA states, "Further development of social services, however, is hindered by outdated legislation, division of competencies, methods of distribution of funds, and professional abilities of social workers in the public administration" and that "It is necessary to go ahead with fundamental transformation of the system" (MoLSA, 2022c). They cite the most pressing challenges as a lack of consistency in implementing social policies across regions and municipalities; inadequate skills of social workers in the public administration and lack of continuous, current training; and inequities in access to funding (MoLSA, 2022c). Other challenges include high inflation, increasing upward pressure on wages, and inadequate and late public sector payments (Stejná odměna, 2023). The 300,000 Ukrainian refugees remaining in the Czech Republic as of February 2023, one year after Russia's invasion of Ukraine, has added demand for social services (Skacel, 2023). The MoLSA however, states "Probably the most import role in the field of social services is played by the strong non-profit, non-governmental organization sector striving to provide modern social services" (MoLSA, 2022c).

These challenges impact the evaluation capacity of Czech social services nonprofit organizations evidenced by the ministry stating that "the quality of the provided services is not adequately checked," and "there do not exist comprehensive data on social services and quality analyses are not performed" (MoLSA, 2022c). While the MoLSA's strategy supports quality improvement of the services driven by their 15 standards, it appears they don't collect or sufficiently use information on the performance and outcomes of those services (MoLSA, 2022d). Their *Standards for Quality in Social Services* in Decree on Implementing Certain Provisions of the Act on Social services 2006, state, "The most important indicator for the evaluation of services, however, is how the provided service projects into the life of the people who receive it" (MoLSA, 2006). The criteria for Standards 5 and 15 are the only ones that explicitly reference "evaluation." Standard 5, Individual Planning of the Course of Social Service, states "The provider has written internal rules according to the type and mission of the social service, which govern the planning and method of re-evaluating the service provision process" and "Together with the person, the provider continuously assesses whether their personal goals are being met." Standard 15, Improving the Quality of Social Services, states "The provider continuously checks and evaluates whether the method of providing social services is in accordance with the defined mission, goals and principles of the social service and the personal goals of individual persons" and "The provider also involves employees and other interested natural and legal persons in the evaluation of the provided social service" (MoLSA, 2006). While all social services organizations are regulated by these standards, how organizations satisfy them and engage in evaluation typically varies across types of organizations. For example, how an organization that serves senior citizens evaluates its

services would likely differ than one who serves people rehabbing from substance abuse because each is answering different evaluative questions, measuring different outcomes and indicators, and inquiring into different criteria about their services, which in turn require different data collection methods. What is measured significantly drives how to best measure it.

Evaluation in the Czech Republic

According to the International Organization for Cooperation in Evaluation (IOCE), there are approximately 165 national and regional evaluation associations around the world (IOCE, 2024). The Czech Evaluation Society (CES), founded in 2007, "is a voluntary, self-governed, non- profit, apolitical, and independent association of professional evaluators" whose "members evaluate public and private-sector development projects and programs" (Czech Evaluation Society, 2022). CES adopted *The Evaluator´s Code of Ethics* in 2011 and *Formal Standards of Conducting Evaluations* in 2013. While CES does not engage in all evaluations of social services, they serve as a pipeline of external evaluators that can be hired for evaluations required by the MoLSA or EU Social Investment Fund. CES also provides professional development with occasional workshops and an annual conference, therefore building evaluation capacity among its member evaluators. The Czech journal *Evaluation Theory and Practice* began publishing in 2013, and its mission is "to create a space for professional discourse about evaluation issues in the Czech and Slovak Republics and thus help to expand the evaluation capabilities between commissioners, evaluators, and other stakeholders" (Evaluation Theory and Practice, 2022).

These developments over the past sixteen years indicate that the evaluation profession is still rather new, and that evaluation is still evolving in the Czech Republic. Remr and Potluka (2020) remind us that the Czech culture and use of terminology around evaluation today is still influenced by the centrally planned economy of the Soviet era evidenced by the dominance of monitoring; the narrow purpose of evaluation for reporting; and lack of trust in evaluation data. They also reference the universally common ECB challenges of funding, time, knowledge and skills, training, and methodological guidance. The supply of experienced external evaluators and university courses in evaluation lag behind growing demand. Despite these challenges, the evaluation requirements from EU financed projects and assistance from the Czech Developmental Agency have been significant catalysts in the professionalization and evolution of evaluation (Remr & Potluka, 2020). If the promise of moving evaluation as a means for organizational learning and improvement is to be realized, Remr and Potluka remind us "Cooperation of all stakeholders is an essential issue for evaluation capacity building, especially in the Civil Society sector" (2020, p. 372).

Materials and Methods

The above description illustrates that the Czech Republic represents an interesting environment suitable for a closer examination of evaluation and ECB. As was mentioned, this study seeks to answer the exploratory research question, within the Czech context: What influences the state of evaluation and ECB in nonprofit social services organizations?. Supporting questions include, What challenges do these organizations face for the evaluation of their services? and How does the relationship with government impact their understanding of evaluation as a concept and process, and ultimately impact ECB? The exploratory question of evaluation capacity was answered by means of a national survey, and semi-structured interviews with directors of these organizations.

Survey

Table 1 serves as the basis for the development of a national, online survey sent to the directors of 1,032 (out of 1,055) registered social services provided by Czech nonprofit organizations, using Qualtrics software. The survey was developed from the literature on evaluation, ECB, frameworks and checklists, evaluability assessments, and existing ECB surveys used in various studies. The authors attempted to balance comprehensiveness with the feasibility to complete it given the time limitations of busy nonprofit directors. The initial draft of the survey based on contemporary literature was adapted to the Czech context, followed by a translation by the two Czech authors into Czech (the lead author only knew English, and this research fulfilled his Fulbright-Masaryk Award in NGO Management), pre-tested by three Czech social services nonprofit directors, and then revised again based on their feedback.

The Tailored Design Method (TDM) (Dillman, 2014) was used to promote and distribute the survey to maximize the response rate. Participants could enter a drawing of five random winners of a gift card worth 1,200CZK ($50 USD and 48 euros as of July 4, 2024). The following organizations endorsed the research: The Association of the Providers of Social Services in the Czech Republic, Czech Evaluation Society, Centre for Nonprofit Sector Research at Masaryk University, Faculty of Social Studies at Charles University, and the University of Ostrava.

The survey response was 41.3%, and the sample profile data is shown below in Tables 2 and 3:

Annual Budget (CZK)	Number	%
0 - 2,499,999	38	9%
2,500,000 - 4,999,999	67	16%
5,000,000 - 9,999,999	103	24%
10,000,000 - 14,999,999	54	13%
15,000,000 - 24,999,999	60	14%
25,000,000 - 49,999,999	60	14%
50,000,000+	44	10%
Total	426	100%

Table 2. Annual Budget in 2022 (n=426)

Table 3. Sources of Revenue (n=426)

Sources of Revenue	%
Public sector through state	65%
Revenues of the organization (payments)	14%
EU Social Investment Fund	8%
Funders - legal entities	7%
Donors - individuals	4%
Other	3%
Total	100%

One important point to note is the fact that 49% of the organizations are small with annual budgets ranging from 0 to 9,999,999 CZK ($0-$430,130 USD and 0-398,055 euros as of July 4, 2024). Another is that most of their revenue comes from government, common to social services, with 73% from the Czech government and the EU combined. Of the 426 organizations, 77.6% provided one to three different social services while the remaining 23.4% provided four or more different social services.

The sample represents the population as shown below in Figures 2-4:

Figure 2. Sample to Population Comparison by Legal Form (n=426)
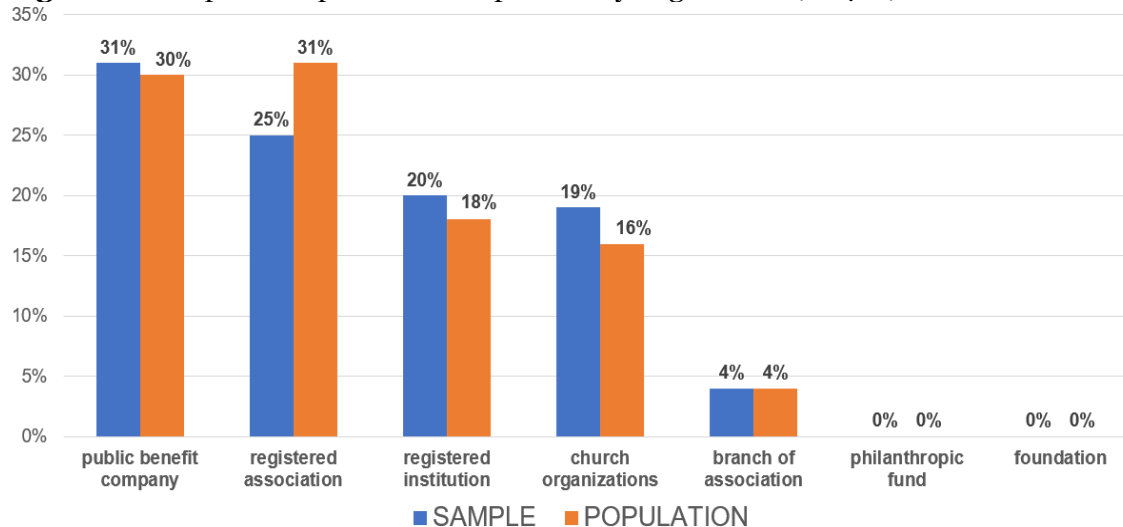
Figure 3. Sample to Population Comparison by Organization Age (n=391)
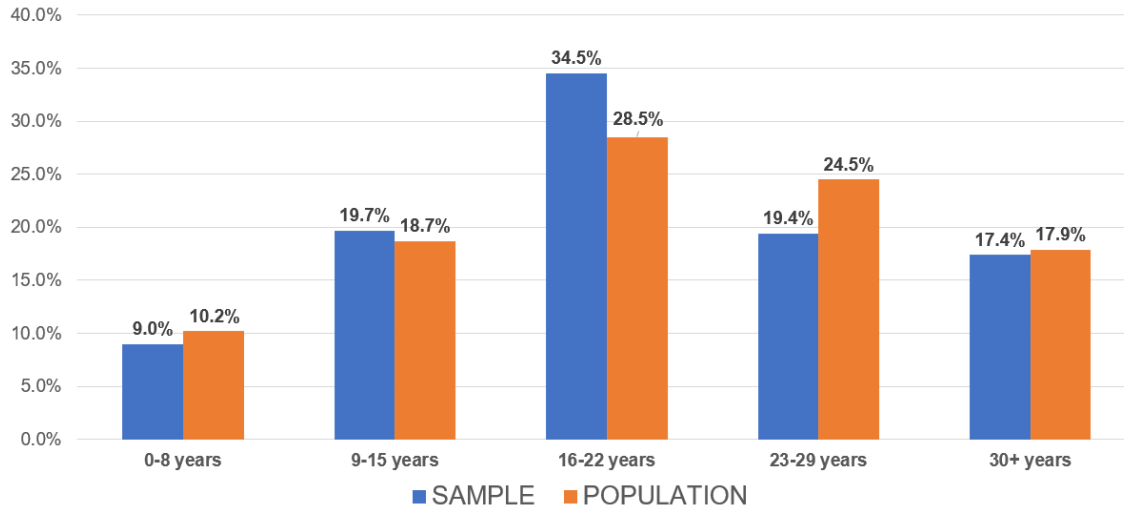
Figure 4. Sample to Population Comparison by # Full-time Employees (n=426)
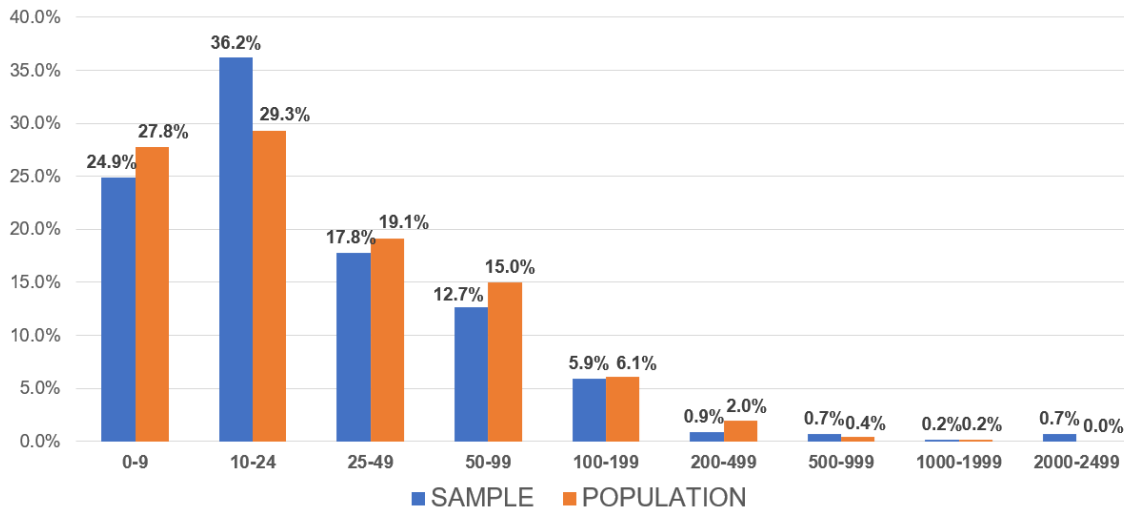
Interviews

Table 1 also informed the development of a complementary, semi-structured interview conducted with twelve directors of nonprofit social services organizations. Organizations were randomly chosen and contacted. Those directors who agreed to participate represented a variety of organizations regarding region, size, and age. Table 4 shows the profile information of the twelve organizations, represented by their directors who participated in the semi-structured interviews.

Table 4. Interview Organization's Profile Information (n=12)

Region	Legal form	Annual budget (CZK)	Age (years)
Ústecký Region	Registered association	0 - 2,499,999	9
Prague	Registered association	5,000,000 - 9,999,999	28
Pilsen Region	Registered association	10,000,000 - 14,999,999	27
South Bohemian	Public benefit company	2,500,000 - 4,999,999	12
Zlín Region	Public benefit company	5,000,000 - 9,999,999	12
Hradec Králové Region	Public benefit company	5,000,000 - 9,999,999	9
Pardubice Region	Public benefit company	5,000,000 - 9,999,999	27
Moravian-Silesian	Registered institution	5,000,000 - 9,999,999	25
Vysočina Region	Registered institution	15,000,000 - 24,999,999	23
Hradec Králové Region	Registered institution	25,000,000 - 49,999,999	19
Liberec Region	Church organization	15,000.000 - 24,999,999	17
Central Bohemian	Church organization	50,000,000+	31

The purpose of the interviews was to complement the survey by gleaning insights into the dynamics of evaluation within their organizations, including how they saw their role as leaders and decision-makers for evaluation, and the challenges they faced for it. The two Czech authors who conducted the interviews used clarifying, confirmatory, descriptive and explanatory probes, when necessary, based on participant responses to extract more thorough and clear responses from the participants. The interviews were recorded, transcribed using Transkriptor software, corrected by the two Czech authors, and translated for the English-speaking lead author for his participation in the data analysis. The analysis of the interviews was conducted in two stages. In the first stage two of the authors, one who conducted the interview and one who did not, both coded all interviews and then met to discuss discrepancies and come to consensus for inter-rater reliability, and then all authors grouped codes into common themes.

Given that ECB is a relatively new phenomenon in the Czech context of social services, a hermeneutic analysis with an inductive approach using open inquiry for coding was applied. The authors did not have a set of codes representing things they looked for and instead interpreted the data. Both coding authors instead just let the data do the talking and compared codes with each other for the same interviews. There may have been some bias from the ECB framework which might have subconsciously emphasized what they were looking for; however, both authors did their best to wipe the slate clean and coded based on what the data presented. The author who conducted the interview would be more familiar with the data than the second coding author who did not. However, that can be an advantage in coding as researchers through the process get closer to the data, and in having the other author who did not interview that participant serve to complement the author who did in the comparison and discussion of codes from both. The second stage involved the authors taking the coded results and cross-referencing them with Table 1 for analysis against the ECB dimensions.

Lastly, both the survey and interview analyses were combined for a more robust understanding of what took place in these organizations regarding ECB from their directors' perspectives. While the information obtained through the survey offers a more general view of the whole population, the information from the interviews attempts to explain the individual phenomena observed.

Results

ECB Dimensions

While 426 participants completed the survey, not all answered all the questions which explains why some of the numbers of participants (the "n's") for each of the results are slightly different. Also, for those participants whose social services were not evaluated, they did not answer all the same questions as those whose were evaluated, as skip patterns were used. Of the 426 survey participants, 51% indicate all their social services were evaluated while 20% said most, 14% said some and 15% said none. Of those that had at least some of their services evaluated, their challenges for evaluating them are shown in Table 5:

Table 5. Challenges for Evaluation of Social Services (n=362)

Challenge	#	% responses	% cases
Not enough time planned and dedicated for evaluation	174	18.8%	48.6%
Not enough money to pay for the evaluation	161	17.4%	44.7%
Lacking internal knowledge and skills in evaluation	128	13.9%	35.8%
Lack staff to conduct the evaluation	90	9.7%	25.0%
No evaluation plan	90	9.7%	25.0%
Difficulties in collecting services data	75	8.5%	21.9%
Lack of help/support in evaluating our services	57	6.1%	15.8%
Information technology (IT) problems	44	4.7%	12.2%
Missing structures and systems for adequate communication	31	3.3%	8.6%
Difficulties in managing services data	28	3.0%	7.8%
Don't know	26	2.8%	7.2%
Other	20	1.3%	3.3%
Organizational culture not supportive of change	3	0.3%	0.8%
Organizational culture not supportive of evaluation	1	0.2%	0.6%
Total	928	100%	

The 64 directors indicating in the survey none of their organizations' services were evaluated provided their reasons shown below in Table 6:

Table 6. Reasons why Social Services were not Evaluated (n=64)

Reason	#	% responses	% cases
We are not required to evaluate our services	34	24.3%	53.1%
We lack the funds necessary to pay for the evaluation	29	20.7%	45.3%
We lack the skills necessary to conduct the evaluation	29	20.7%	45.3%
We don't have time to conduct the evaluation	21	15.0%	32.8%
We don't know where to find an evaluator	12	8.6%	18.8%
We don't believe the evaluation of our services is necessary	11	7.9%	17.2%
Other	4	2.9%	6.3%
Total	140	100%	

The top three cited challenges for those organizations who had their services evaluated were two through four most frequently chosen for those who did not. For them, the top reason was they were not required to have their services evaluated. This presents an interesting incongruence with the interview results where almost all 12 interview participants said they would still have their services evaluated even if they were not required. There was a variance in what they would do within the scope of evaluation due to size, capacity, knowledge, and other factors. Some examples of what they said include:

"So, I think we would do it, but we would probably do it more and more on our knees. I think if nobody wanted us to do it, I don't think I'm going to make any graphic charts to prove the mission of the organization."

"I would think so. It's just that it would never take that kind of form... like formal. It's kind of like, you're always doing it on the fly and you're not making deliverables out of it."

The importance of evaluation within the organization was consistent across all legal forms of organizations, as church organizations, public benefit companies, branch of associations, registered associations and registered institutions each deemed evaluation very important ranging from 55 to 67%, while a similar range covered those indicating evaluation was somewhat important from 31 to 39%.

The survey and interview results are now presented by each ECB dimension. This information provides insights into what is behind these challenges and an overall picture of ECB within these organizations. It is important to note that some redundancy is expected as they are not mutually exclusive, they impact each other, and some results can be relevant to multiple dimensions.

Leadership

Directors play a key role in the presence of evaluation, and it is a matter of accountability for them in their interest to keep their organizations in the provider network and on the register of social service providers. For who makes decisions to pay for evaluation in their organization they were the majority at 57%. The degrees to which they integrated external demands for evaluation into their organization's evaluation systems and processes were 67% somewhat and 19% completely. One director acknowledged the political frame for evaluation:

"So, my role is, I would say, partly controlling and partly motivating… So, also linking and networking within the services and possibly also by being involved in politics somehow in that external environment as well, so some comparison of what's happening in the region where we operate and therefore some sharing of trends, as it is in the community planning of the cities where we operate, because I'm there in some way as well."

The role directors see themselves play in, and the extent to which they are, supporting and driving the evaluation process is important for successful ECB. One director describes their role as a driver of the process:

"Then in terms of the effectiveness of services, for example, and coordination and so on. So that's what we're trying to do, who's evaluating that is me as the director and now we have a direct support team leader, so we're actually working with him to do that. I think that's our job - just to get it done in some way."

Organizational Culture

Results produced numerous indicators concerning whether the organization's culture was conducive to evaluation and ECB. For the importance of the evaluation of their social services, 60% said very important and 35% said somewhat important. Approximately 71% indicated their organization seeks ways to improve processes within their social services while about 74% indicated their organization views problems or issues as opportunities to learn. For indicating their organization's tolerance, openness, and receptivity to negative information, 79% used negative feedback from employees to implement changes; 76% encouraged employees to provide both negative and positive feedback; and 69% discussed negative feedback with employees in meetings.

About 62% of directors indicated their organization allows enough time to reflect on and discuss its successes, challenges, and failures while 70% of the directors indicated their organization discusses evaluation regularly. How organizations financially account for their evaluation expenses is a significant cultural indicator for how they view them and prioritize them in their budget. In this study 57% of the participants indicated they recorded evaluation expenses as expenses related to their services while 29% recorded them as administrative or "overhead" expenses. This also is important because nonprofits are pressured to keep administrative or "overhead" costs low.

In cross tabbing the data, the percentage difference between the 70% of the sample participants indicated that the evaluation of their social services was discussed regularly at meetings and approximately 14% who said it was not, were consistent across who typically evaluates the social services. As expected, the extent that evaluation was deemed important within the organization matched up well with how often evaluation was discussed at meetings. Of the approximately 60% of the sample said evaluation was very important to their organization, 74.4% indicated it was discussed regularly at meetings, while of the approximately 35% who indicated evaluation was somewhat important, only 25% of them said it was discussed regularly at meetings.

Organizational Learning

All 362 organizations that had their services evaluated indicated they used their results. Examples are shown below in Table 7:

Table 7. How Evaluation Results of Social Services are Used (n=362)

How evaluation results are used	#	% responses	% cases
Make changes in existing services	283	18.7%	78.6%
Establish service goals or targets	257	16.9%	71.4%
In strategic planning	197	13.0%	54.7%
Outreach and public relations	142	9.4%	39.4%
Make staffing decisions	129	8.5%	35.8%
Reporting compliance with external stakeholders	111	7.3%	30.8%
Develop new services	107	7.1%	29.7%
Budgeting decisions	101	6.7%	28.1%
Report to management of organization	94	6.3%	26.4%
Secure funding	84	5.6%	23.6%
Other	11	0.5%	2.2%
Don't know	2	0.1%	0.6%
Total	1,518	100%	

Directors were asked if their organizations learned from their evaluation results, and 52% indicted they learned a lot while 46% learned a little. Table 8 below shows examples of what was learned:

Table 8. What Organizations Learned from their Social Service Evaluations (n=357)

What was learned	#	% responses	% cases
How to improve our services	308	39.2%	86.3%
More about how our services work	184	23.4%	51.5%
What is necessary to evaluate services	133	16.9%	37.3%
About the evaluation process	92	11.7%	25.8%
Roles and responsibilities of stakeholders	57	7.3%	16.0%
Don't know	6	0.8%	1.7%
Other	7	0.6%	1.4%

Total	785	100%

Discussing evaluation regularly at meetings, while important, is not enough. Meaningful reflection, discussion, review, and use of results and closing feedback loops are all necessary for continuous learning and improvement. Of the approximately 70% of participants that indicated that the evaluation of their social services was discussed regularly at meetings, approximately 65% said there was enough time allowed for the organization to reflect on and discuss its successes, challenges, and failures while approximately 32% said there was not. Also, within that 70%, 76% indicated they often sought ways to improve processes within their social services while 24% said they sometimes sought them.

Resources

Table 5 shows that the four most frequent responses for challenges faced with evaluation are resources – time, funds, knowledge and skills and personnel. Time remains a challenge for the evaluation of social services due to most of the organizations being small, struggling to meet the demand for their services, staff turnover, and the opportunity costs involved when evaluations are conducted by internal staff. This is reflected in some of the directors' comments:

> "I think a smart evaluator also tries not to burden the team more than necessary."

> "How do we reduce the administrative burden and have the evaluation set up so that it's sort of like automatic, that we can just click somewhere or open something up and see how we're doing? So, I think that's kind of been a goal of ours for a number of years that we haven't quite succeeded in."

The interviews revealed that the costs for evaluation were typically paid for with the funds received from the external entity requesting the evaluation, such as the MoLSA, the EU Social Investment Fund, or regional government while internal evaluations were paid out of general operating budgets, or the funds received for the services provided. The mix of external and internal evaluations was reflected in top three percentages of sources of funds to pay for evaluations in the survey respondents' organizations, with 43% from the public sector (Czech government), 33% from their own internal funds, and 9% from the EU Social Investment Fund. Organizations with larger budgets typically had higher percentages of use of external evaluators whether they be independent, from the government or other external entities.

In cross tabbing the data, 57% of the directors or head/service managers were most likely to be the one evaluating the services across all budget categories, ranging from 61% in organizations with a budget of 25,000,000 - 49,999,999 CZK to 49% for the largest organizations with a budget of more than 50,000,000 CZK. Only 16% of the sample indicated they had an employee dedicated to evaluations, otherwise known as an internal evaluator, with not much difference across budget sizes ranging from 14.7% for organizations with a budget of 5,000,000 - 9,999,999 CZK to 18.5% for the largest organizations with a budget of more than 50,000,000 CZK. Organizations utilizing an external, independent evaluator comprised 11.3% of the sample with not much difference across budget categories except for the lowest category of 0 - 2,499,999 CZK for whom only 3 organizations utilized one. When cross tabbed against the number of full-time employees and the age of the organization, differences were not significant as the director or service manager garnered the highest percentages for who evaluated the services in every category.

The challenge for affording external evaluators was reflected in most of the interviews:

"Well, we don't have an external evaluator, and we don't get a penny for that at all, so it's just that we're a learning organization and we have to be able to do it ourselves and set the dynamics of the service."

"We would reach for an external (evaluator), or at least maybe even calmly more multiple people, who could give us again some even picture of how they do it, what good practice they have, so we would stand for maybe sharing as well. Now that the money is not open...we're kind of doing it here on our own."

The financial support of the MoLSA, EU Social Investment Fund, and regional governments factored into their ability to utilize an external evaluator:

"We actually brought in like an external worker, Mrs. Soňa K., that was a condition of the European Social Fund... we accepted the challenge of Mrs. Markéta P. of the Social Innovation Department of the MoLSA where there was a mandatory evaluation by an external evaluator. So, without those projects we would not have, we might never have got to that point at all."

"So, actually the only evaluation that we're involved in now - quite spiritedly I think - and it's being done for us by somebody else is actually in a project that we're doing now with support. Or actually it's the second project that we have, where we are implementing a social housing project from a call - before it was OPZ (European Social Fund - Operational Programme for Employment), now it's under the OPZ+ (European Social Fund - Operational Programme for Employment plus)."

The internal knowledge and skills to conduct evaluations was also an issue, as expertise in evaluation is not part of the training for social workers nor is it typically part of the background of directors of these organizations. Here is what a few directors had to say:

"I think the other thing that's missing is that... in that area of social work, it's not quite clear how to evaluate those social services. There's not some sort of, I think, widespread knowledge base. And people don't even have that information, they don't know how to do it."

"We don't quite know how to go about it. We like to have it written down some, the area of housing, like maybe gained employment, how many like can we have, how many women... what helped them the most. So, from that perspective, we would benefit from her (external evaluator's) professional perspective."

Systems and Structures

Approximately 25% of the surveyed directors (n=360) indicated their organizations did not have an evaluation plan, and none of the directors interviewed referenced anything resembling an ECB plan. However, some indicated their desires to expand their evaluation efforts to be more comprehensive, robust, and meaningful including those who specifically referenced the desire to measure the specific impact on individual clients. Some of the directors interviewed indicated their organizations did not have a strategic plan, while some had strategic plans, but evaluation was not included in them. Those indicating evaluation was in their strategic plans offered few specific examples except to inform the plan for making changes to services, but not necessarily for planning on how to expand or improve their evaluation process. Some directors also expressed

the desire to make evaluation more meaningful but felt limited by the parameters of the external stakeholders requiring it, as in these examples:

> "Then there's the system maybe at the level of the municipality or at the level of those foundations, which tend to be like more substantive...If it's just some reporting according to some terribly complicated and completely arbitrary template, then in that case maybe it's not really evaluation in the sense of the word evaluation, it's more about fulfilling some bureaucratic requirements and discipline mainly."

> "We are either paid from the budget of the Zlín Region - we didn't have an evaluation there, but when it's the projects from the OPZ (European Social Fund - Operational Programme for Employment), where there are more funds and more possibilities, they basically involve other entities. So, it is not a rule, it depends on how the Zlín Region writes the project, but in this last project it was the case that there was actually an evaluation, because they are of course interested in the how the funds are spent."

Approximately 70% of the directors indicated their organizations discussed evaluation regularly at meetings, and approximately 74% of the directors indicated their organizations had evaluation included in their policies and procedures. Other challenges related to systems and structures include difficulties in data collection (22%); information technology issues (12%); missing structures and systems for adequate communication of information and results (9%); and problems managing their data (8%). Of the approximately 60% of participants indicated evaluation was very important to their organization, 54% said it was included in their organization's policies and procedures while approximately 9% said it was not. Of the approximate 34% that said evaluation was somewhat important, approximately 19% included it in their policies and procedures while about 10% did not.

Program Evaluability

Approximately 89% of directors indicated their social services had clear goals and approximately 82% indicated their services have clear and well-defined outcomes. Approximately 88% of directors indicated there is agreement among the users of evaluation results for how those results will be used. However, concerns stem from the fact that approximately 65% of directors said none of their services had a theory of change and about 74% said none had a logic model. This raises the question of what the basis was for, and how they developed, clear and well-defined outcomes without a theory of change or logic model, or how they're framing "outcomes."

Discussion

The organizations' hard work, concern and care for clients shone through all interviews with their directors. A recurring theme, except from a few of the larger, more sophisticated organizations, is the lack of understanding and/or narrow framing and perspective for evaluation. Often the concept was used interchangeably with "inspections," "quality assurance," "meeting standards," "monitoring," and "audits." While they may be components of an overall evaluation effort, they alone do not comprise a holistic or comprehensive evaluation of the impact of the services on individual clients. This also presents the possibility that they have a limited understanding of client impact to be mostly client satisfaction. This dilemma seems largely driven by the fact that most formal, structured evaluations are externally driven by the MoLSA whose goal is to assess the organization's ability to meet their 15 standards. The standards are rigorous and seemingly a good way to assess quality and ensure the safety of clients, however they do not specifically extend

to measuring the desired outcomes for each client. That seems to be left to the organizations, most of whom do not know how to do it and/or don't have the capacity to do it. While most survey participants said their services have clear and well-defined outcomes, the measuring of service outcomes was not referenced in any of the interviews. Based on these results, it is possible that what they mean by "outcomes" is the ability to meet the MoLSA's standards. Deeper investigation into this would be necessary to fully understand this incongruence in results.

Resource dependency heavily influences the behavior and decisions of nonprofit organizations because they are not self-sustaining organizations and instead are open systems relying heavily on external financial support to deliver their services. It's no surprise that money was a recurring theme throughout this study within the context of funds to provide and/or expand services, conduct evaluations, and pay staff. The inability to adequately pay their social workers who interact directly with their clients was discussed in all twelve interviews, and the dilemma is common in other countries including the U.S. This should not be framed solely as a staffing issue, as turnover is a program issue in terms of quality of service delivery; consistency with clients in terms of knowledge of their case, relationships with them, etc., and the constant need for training of new staff. If there is instability with delivery of services it is likely to also impact their evaluation, especially in smaller organizations who cannot afford to hire an external evaluator. How can an organization build capacity to evaluate their services when they struggle simply to deliver them?

The tying of the evaluation to funding seemed embedded in how stakeholders understood evaluation and why it was conducted. Interview participants expressed the desire to have an external evaluator, but most cannot afford one. The confusion of evaluation with monitoring might contribute to a narrow perspective for the role of the evaluator as important only to securing funding which contributes to a culture of compliance in lieu of organizational learning. This misaligned prioritization of evaluation driven by resource dependency often runs counter to learning and instead develops more so a culture of compliance.

The top three challenges for evaluation are consistent with some of the more frequently cited challenges for evaluation in other ECB studies including Wade and Kallemeyn's qualitative study of 12 Chicago area nonprofits (2020); Bourgeois and Cousins' qualitative study of Canadian federal government organizations (2013); Fierro and Christie's study of 162 evaluators and program managers from U.S. federal public health programs (2017); Taylor-Ritzler et al.'s study of 169 Chicago area nonprofits (2013); Hudib and Cousins' global study of 52 bilateral and multilateral aid organizations (2022); and Carman and Fredericks's study of 189 Indiana nonprofits (2008). In the Indiana study, the challenges closely matched their most frequently cited "barriers" for evaluation that were "not enough time," "not enough trained staff," and "not enough evaluation expertise" (Carman & Fredericks, 2008). These other studies in other countries and contexts, including those in places where evaluation is arguably more advanced than in the Czech Republic, show that these challenges remain consistent across all because they are inherent to the dynamics of program evaluation and ECB and where they are in their evolution within nonprofit organizations, including social services.

For professional development and training of staff on evaluation, time is the biggest constraint. Most organizations are small and therefore do not have a dedicated person to conduct the evaluation, so the responsibility falls to the social worker or to the director. Most of the interview participants stated this responsibility was in the social workers' job descriptions. However, the lack of knowledge and skills on top of the lack of time result in opportunity costs for their core work. Survey participants indicated they provide professional development for evaluation including courses but given that very few university courses in evaluation exist in the Czech

Republic, it is not clear what they are referencing. The receptivity and support for evaluation internally from staff had mixed results from the interviews ranging from them understanding its importance and that it's part of their job to not believing it is helpful in the organization meeting its goals. This also may relate to their level of understanding for the concept and process, as understanding must come first before valuing something.

The role of the director in the evaluation process varied greatly across the interview participants, and only a few conveyed that they "drove" the process to make sure it happened. All understandably seemed unaware of the concept of ECB or what their role should be in it. A few in the larger, more advanced organizations desired the ability to expand their evaluation process to be more meaningful by measuring the impact of their services on individual clients. However, 1) they were unclear on how to build capacity to move in that direction, 2) they cited capacity issues as a reason for their inability, and/or 3) that the MoLSA or regional governments were not interested enough in individual client impact.

These organizations use evaluation results for important things such as planning, making changes, establishing goals and targets, however it is unclear whether they transform such use into continuous learning towards sustainable evaluation practice or if these examples of use are limited in scope and depth due to their incomplete understanding for evaluation.

Limitations and Further Research

This research discovered Czech nonprofit social services organizations' understanding and activity around evaluation. However, the results are through the lens and perspective of their directors who, while the likely choice for such inquiry, may also reflect some social desirability bias in their responses. Every attempt has been made to ensure accurate, two-way translation however, contextual nuances and the impact of the English-speaking evaluation lexicon may have impacted the participants' understanding of questions. For example, there are several terms in the Czech language that refer to evaluation - "evaluace" and "hodnocení" or "vyhodnocení." This inconsistency may result in different perceptions of the purpose, methodology, and overall process of evaluation.

The confusion between monitoring and evaluation and the variance in social constructions, understanding and perceptions around the concept and process of evaluation warrants a qualitative study to delve deeper into what these organizations specifically think evaluation is. Such a study should inquire with other key stakeholders such as social workers. Some of the interview participants reference "individual client plans" and "outcomes" as a part of meeting quality standards, but it was unclear what was meant by "outcomes" and what they did with either, warranting further investigation. In-depth case studies are recommended where the researcher would closely observe their work with evaluation, how their services are evaluated against the MoLSA's standards, and the dynamics between stakeholders around evaluation and the dimensions for ECB.

Conclusions and Recommended Practice

This study, while conducted in the Czech Republic, reflects some conclusions and implications for practice applicable for the rest of the world. The top three most cited challenges for the evaluation of their services in time, funds, and internal knowledge and skills are consistent with other ECB studies from around the world, including countries in which evaluation is advanced as a

profession. This seems to indicate that these challenges are universal and global regardless of the level of professionalization of evaluation in that country, and the type, size, and maturity of the organization. With the understanding, appreciation, concern and factoring in of differences in context and culture, this consistency still provides opportunities for organizations to learn from each other's challenges and successes in ECB.

The interview results point out two important factors for ECB not often discussed – stakeholders' social construction of the concept and process of evaluation and the organizations' directors lacking knowledge of ECB and what their role can and arguably should be for it. Organizations may respond affirmatively in studies they are engaging in "evaluation," but what do they mean by it, and what exactly are they doing for it. Confusion between monitoring and evaluation along with narrowly framing evaluation as quality assurance or client satisfaction shows the variance in understanding of evaluation. Education or training in ECB should be seriously considered in nonprofit management degree programs as well as non-academic training programs conducted by nonprofit centers around the world.

The state of evaluation within Czech nonprofit social services organizations indicates a solid foundation to build from. However, the dominant influence from MoLSA, regional governments, and EU Social Investment Fund seems to significantly form the organization directors' social constructions, understandings and perceptions of the concept and process of evaluation. This in turn impacts their lacking knowledge of ECB and what role they could play in it, reflected in the interviews when they deferred the addressing of ECB challenges to the external stakeholders requiring the evaluation. Client satisfaction, monitoring, and quality standards are important components of a robust and comprehensive evaluation process, but in this study they often were confused individually or collectively as "evaluation." The missing component appears to be the measuring of the impact of services on each individual client as outcomes, in terms of how their lives were changed or improved. Some directors expressed interest in expanding their evaluation efforts to include outcome measurement but cited lacking capacity as a deterrent. Organizations should conduct an evaluability assessment using the six dimensions for ECB to build long-term sustainable evaluation practice. With such an effort they can begin transforming their organizational culture into one driven by continuous learning and improvement.

We offer below in Table 9 conclusions with corresponding recommended practice by each key stakeholder group within the world of Czech nonprofit social services, conveying that ECB "takes a village" and should not solely be the responsibility of the social services organizations.

Table 9. Conclusions with Corresponding Recommended Practice for ECB

Conclusions	Recommended practice	Stakeholder group
Predominant focus on client satisfaction, MoLSA standards, monitoring and outputs	Build upon this foundation to move towards also measuring outcomes for individual clients • Develop theory of change model for services • Develop logic model for services • Develop outcome measurement framework	Social services organizations
Lacking funds to expand evaluation effort	• Budget for evaluation with dedicated line item • Negotiate with external funders for more funds for evaluation	Social services organizations
	Increase funding for expanded evaluation	MoLSA and regional governments

Demonstrated desire to expand and improve evaluation	Develop ECB plan and link to strategic plan	Social services organizations
Social worker staffing concerns and challenges • Insufficient compensation • Extreme workload – burnout/turnover • Professionalization • Lack of knowledge, skills and training in evaluation	• Advocate to MoLSA to set higher rate of pay for social workers • Budget and seek professional development opportunities	Social services organizations
	• Increase pay of social workers • Increase funding for professional development for evaluation and ECB • Advocate with Ministry of Education to universities to offer courses in evaluation	MoLSA and regional governments

	• Promote *Formal Standards for Conducting Evaluations* and *Evaluator's Code of Ethics*
	• Collaborate with universities to offer professional development and consulting to organizations in evaluation and ECB
	      o Internships
	      o Practicums
		Czech Evaluation Society

	Provide professional development opportunities in evaluation and ECB	The Association of the Providers of Social Services in the Czech Republic (APSS CR)

While these conclusions and recommended practice are specific to this Czech study, all can serve for learning and adaptation of human and social services organizations from around the world who are challenged with evaluation and/or seek to develop sustainable evaluation practice.

The dimensions for ECB can be applied to these organizations regardless of region, legal form, or size. Such application can provide a guide for building sustainable evaluation practice through training, workshops, hands-on learning, coaching and mentoring. In some ECB workshops, as those delivered by one of the authors, the dimensions have served as workshop components and collectively as an overall guide for recommended practice. Evaluation is nothing new to social services and the field of social work. For example, the National Association of Social Workers (NASW) has in its Ethical Standards, 3.03 a standard for Performance Evaluation that states, "Social workers who have responsibility for evaluating the performance of others should fulfill such responsibility in a fair and considerate manner and on the basis of clearly stated criteria." Standard 5.02, for Evaluation and Research has 17 sub standards stating social workers should "monitor and evaluate policies, the implementation of programs, and practice interventions"; "promote and facilitate evaluation and research to contribute to the development of knowledge" ; "critically examine and keep current with emerging knowledge relevant to social work and fully use evaluation and research evidence in their professional practice"; and fulfill other sub standards for how they evaluate programs, policies and interventions as well as report findings (NASW, 2021). The second example, worth mentioning because it represents the Czech context, is The Institute of Education of the Association of Social Service Providers of the Czech Republic which offers two evaluation-related courses, "Evaluation and motivation of social service workers" and "Social service quality assessment and change implementation" (Institute of Education APSS CR, 2025). Standards and how-to workshops are important; however, the authors feel the key

missing piece is the organizations' development of evaluation capacity, and therefore training and professional development in this area would be helpful for nonprofit managers.

Clients or consumers ultimately will benefit from more robust, meaningful, comprehensive and impactful evaluations driving the improvement of services they receive which ultimately enhances Czech civil society. The Czech context in this study can assist other former Soviet Bloc nations who are challenged with evaluation and ECB on similar trajectories, as they their civil societies evolve, and they advance relationships between their governments and nonprofit sectors. This study reminds us that the challenges for evaluation and ECB are global and universal, existing across nations; types, sizes, and ages of organizations; and contexts. Leaders of nonprofit organizations can benefit from learning about evaluation and ECB so they can go beyond the contextual grounding and compliance culture set by their external stakeholders and be catalysts for advancing and progressing the evaluation of their organizations' programs or services.

Declaration of interest statement

The authors have no conflicts of interest to disclose.

Ethics Approval

The Grand Valley State University Institutional Review Board (IRB) (Protocol #23-018-H) reviewed this research and issued a Certificate of Exemption. This study was also approved by the Ethical Commission for Research at University of Ostrava, Ostrava, Czech Republic.

Informed Consent

All participants in the study were informed regarding the purpose of the study and their role and rights associated with participation. All participants gave their informed consent.

References

Argyris, C., & Schon, D.A. (1996). *Organizational Learning II: Theory, Method, and Practice (2nd Edition)*. Reading, MA: Addison-Wesley Publishing Co.

Alaimo, S. P. (2008). Nonprofits and evaluation: Managing expectations from the leader's perspective. *New Directions for Evaluation*, 2008(119), 73–92. https://doi.org/10.1002/ev.268

Bamberger, M. & Mabry, L. (2020a). *Real World Evaluation: Working Under Budget, Time, Data, and Political Constraints 3rd edition*. Thousand Oaks, CA: Sage Publications, Inc.

Bamberger, M. & Mabry, L. (2020b). APPENDIX 19.6 EVALUATION CAPACITY DEVELOPMENT. *Real World Evaluation: Working Under Budget, Time, Data, and Political Constraints 3rd edition*. Thousand Oaks, CA: Sage Publications, Inc. https://edge.sagepub.com/system/files/bamberger_realworld_3e_online_appendix_proof_final_0.pdf . Accessed July 10, 2022, 266-273.

Bourgeois, I. and Cousins, J.B. (2013). Understanding Dimensions of Organizational Evaluation Capacity. *American Journal of Evaluation* 34(3), 299-319.

Boyle, R., Lemaire, D., & Rist, R. C. (1999). Introduction: Building evaluation capacity. In R. Boyle & D. Lemaire (Eds.), *Building effective evaluation capacity: Lessons from practice*. New Brunswick, NJ: Transaction, 1-19.

Bryson, J.M. and Patton, M.Q. (2015). Analyzing and Engaging Stakeholders. In Newcomer, Kathryn E., Hatry, Harry P., and Wholey, Joseph S. eds., *Handbook of Practical Program Evaluation, 4th edition*. Hoboken, NJ: John Wiley & Sons, 36-61.

Carman, J.G. and Fredericks, K.A. (2008). Nonprofits and Evaluation: Empirical Evidence From the Field. *New Directions for Evaluation*, 119, 51-71.

Compton, D.W., Glover-Kudon, R., Smith, I.E., and Avery, M.E. (2002). Ongoing Capacity Building in the American Cancer Society 1995-2001. *New Directions for Evaluation*, 93, 47- 61.

Cousins, J.B., Goh, S.C., Elliott, C.J., and Bourgeois, I. (2014). Framing the Capacity to Do and Use Evaluation. *New Directions for Evaluation* 141, 7-23.

Czech Evaluation Society. (2011). Evaluator´s Code of Ethics. https://czecheval.cz/Dokumenty%20%20archiv/ces_eticky_kodex_en.pdf. Accessed July 21, 2022.

Czech Evaluation Society. (2013). Formal Standards of Conducting Evaluations. https://czecheval.cz/Dokumenty%20%20archiv/ces_formalni_standardy_evaluaci_short_en.pdf . Accessed July 21, 2022.

Czech Evaluation Society. (2022). About us. https://czecheval.cz/en/About-us .
Accessed August 7, 2022.

Czech Statistical Office. (2022a). Public Database: Organizational structure by group of institutional sectors. https://vdb.czso.cz/vdbvo2/faces/en/index.jsf?page=vystup-objekt&pvo=ORG07&z=T&f=TABULKA&skupId=3775&katalog=33695&pvo=ORG07&str=v386&v=v7 null null null . Accessed December 13, 2022.

Czech Statistical Office. (2022b). Satellite Account of Non-profit Institutions. https://apl.czso.cz/pll/rocenka/rocenka.indexnu_sat. Accessed April 21, 2023.

Dillman, Don A., Smyth, Jolene D., Christian, Leah Melani. (2014). *Internet, Phone, Mail and Mixed-Mode Surveys: The Tailored Design Method, 4th edition*. Hoboken, NJ: John Wiley.

Duigan, P. (2003). Mainstreaming Evaluation or Building Evaluation Capability? Three Key Elements. *New Directions for Evaluation* 99, 7-21.

Evaluation Theory and Practice. (2022). About the Journal. https://en.evaltep.cz/. Accessed August 7, 2022.

Festen, M. & Philbin, M. (2007). *Level Best: How Small and Grassroots Nonprofits can Tackle Evaluation and Talk Results*. San Francisco: Jossey-Bass.

Fierro, L.A. and Christie, C.A. (2017). Evaluator and Program Manager Perceptions of Evaluation Capacity and Evaluation Practice. *American Journal of Evaluation* 38(3), 376-392.

Fitzpatrick, J.L., Sanders, J.R. & Worthen, B.R. (2011). *Program Evaluation: Alternative Approaches and Practical Guides, 4th edition*. Boston: Pearson Education.

Green, A.T. (1999). Nonprofits and Democratic Development: Lessons from the Czech Republic. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 10(3), 217-235.

Grudens-Schuck, N. (2003). The Rigidity and Comfort of Habits: A Cultural and Philosophical Analysis of the Ups and Downs of Mainstreaming Evaluation. *New Directions for Evaluation* 99, 23-32.

Hudib, H.A. and Cousins, J.B. (2022). Evaluation policy and organizational evaluation capacity building: A study of international aid agency evaluation policies. *New Directions for Evaluation* 173, 29-48.

Imas, Linda G. Morra and Rist, Ray C. (2009). *The Road to Results: Designing and Conducting Effective Development Evaluations*. Washington, DC: The World Bank.

Institute of Education - Association of Social Service Providers of the Czech Republic. "Turnkey courses." https://www.institutvzdelavani.cz/kurzy-na-klic. Accessed May 6, 2025.

International Organization for Cooperation in Evaluation (IOCE). (2024). VOPE Directory. https://ioce.net/vopes/vope-directory/ Accessed July 9, 2024.

King, J.A. (2007). Developing Evaluation Capacity Through Process Use. *New Directions for Evaluation* 116, 45-59.

Knowlton, L.W. and Phillips, C.C. (2013). *The Logic Model Guidebook: Better Strategies for Great Results, Second Edition*. Thousand Oaks, CA: Sage Publications.

Ministry of Labour and Social Affairs (MoLSA). (2016). Standards for Quality in Social Services. https://www.mpsv.cz/documents/20142/225517/Vyhlaska_c._505_2006.pdf/541be239 51fb-5924-bf4e-aecf544f318b . Accessed April 14, 2023.

Ministry of Labour and Social Affairs (MoLSA). (2022a). Social Services Act No. 108/2006 Coll. https://www.mpsv.cz/documents/20142/372809/Annex_3_social_services_act.pdf . Accessed July 21, 2022.

Ministry of Labour and Social Affairs (MoLSA). (2022b). Register of Social Services Providers. http://iregistr.mpsv.cz/socreg/hledani_sluzby.do?SUBSESSION_ID=1659701652781_2 . Accessed August 5, 2022.

Ministry of Labour and Social Affairs (MoLSA). (2022c). The Structure of Social Services – current status in the Czech Republic. https://www.mpsv.cz/documents/625317/625935/2646.pdf/ . Accessed July 22, 2022.

Ministry of Labour and Social Affairs (MoLSA). (2022d). Social Services. https://www.mpsv.cz/web/en/social-services. Accessed August 6, 2022.

Ministry of the Interior. (2012). Collection of Laws and Collection of International Treaties. https://aplikace.mvcr.cz/sbirkazakonu/SearchResult.aspx?q=89/2012&typeLaw=zakon &what=Cislo_zakona_smlouvy. Accessed July 12, 2023.

Monroe, M.C. et al. (2005). Evaluators as Educators: Articulating Program Theory and Building Evaluation Capacity. *New Directions for Evaluation* 108, 57-71.

National Association of Social Workers (NASW). (2021). Code of Ethics of the National Association of Social Workers. https://www.socialworkers.org/About/Ethics/Code-of-Ethics/Code-of-Ethics-English/Social-Workers-Ethical-Responsibilities-to-the-Social-Work-Profession. Accessed May 6, 2025.

Newcomer, Kathryn E., Hatry, Harry P., and Wholey, Joseph S. (2015). Planning and Designing Useful Evaluations. In Newcomer, Kathryn E., Hatry, Harry P., and Wholey, Joseph S. eds., *Handbook of Practical Program Evaluation, 4th edition*. Hoboken, NJ: John Wiley & Sons, 7-35.

Office of the Government of the Czech Republic, Department of Sustainable Development. (2017). Strategic Framework Czech Republic 2030. https://www.vlada.cz/assets/ppov/udrzitelny-rozvoj/projektOPZ/Strategic_Framework_CZ2030.pdf Accessed December 13, 2022.

Office of the Government of the Czech Republic. (2021). Strategy for Cooperation Between Public Administration and Non-governmental Non-profit Organizations 2021-2030.

Parsons, Beverly L., Lovato, Chris Y., Hutchinson, Kylie and Wilson, Derek. (2016). Building Evaluation Capacity Through CLIPs: Communities of Learning, Inquiry, and Practice. *New Directions for Evaluation* 151, 21-36.

Patton, M.Q. (2022). *Utilization-Focused Evaluation, 5th Edition*. Thousand Oaks, CA: Sage Publications.

Pospisil, M., Navratil, J. and Pejcal, J. (2015). Civil Society in the Czech Republic. In Christian Schreier, ed. *25 Years After: Mapping Civil Society in the Visegrad Countries*. Stuttgart: Lucius & Lucius Verlag, 143-188.

Potucek, M. (2000). The Uneasy Birth of Czech Civil Society. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 11(2), 107-121.

Preskill, H. & Boyle, S. (2008). A Multidisciplinary Model of Evaluation Capacity Building. *American Journal of Evaluation* 29 (4), pp. 443-459.

Remr, Jiří and Potluka, Oto. (2020). Czechia. In R. Stockmann et al. (eds.), *The Institutionalisation of Evaluation in Europe*, 351-375. https://doi.org/10.1007/978-3-030-32284-7_14 .

Runnels, Vivien, Andrew, Caroline and Rae, Jennifer. (2017). Building Evaluation Culture and Capacity in a Community-Level Program: Lessons Learned from Evaluating Youth Futures. *Canadian Journal of Program Evaluation* 32(1), 122-130.

Sanders, J.R. (2003). Mainstreaming Evaluation. *New Directions for Evaluation* 99, 3-6.

Schein, E. (1992). *Organizational Culture and Leadership*. San Francisco: Jossey-Bass Publishers.

Skacel, Jiri. (2023). Ukraine war: A year on, refugees in Czech Republic strive to integrate but miss home. *Euronews*. Retrieved from https://www.euronews.com/2023/02/22/ukrainians-find-refuge-in-the-czech-republic. Accessed July 5, 2023.

Sonnichsen, Richard C. (1999). Building Evaluation Capacity Within Organizations. In R. Boyle & D. Lemaire (Eds.), *Building effective evaluation capacity: Lessons from practice*. New Brunswick, NJ: Transaction, 53-73.

Stejná odměna (2023). What do we recommend? Retrieved from https://www.stejnaodmena.cz/co-doporucujeme/ . Accessed June 29, 2023.

Stockdill, S.H., Baizerman, M. & Compton, D.W. (2002). Toward a Definition of the ECB Process: A Conversation with the ECB Literature. *New Directions for Evaluation*, 93, 7-26.

Stufflebeam, D. L. (2002). Institutionalizing evaluation checklist. Retrieved from https://wmich.edu/evaluation/checklists. Accessed June 29, 2023.

Taylor, David and Balloch, Susan. (2022). Introduction - The Politics of Evaluation: An Overview. In Taylor, David and Balloch, Susan, eds., *The Politics of Evaluation: Participation and Policy Implementation*, pp. 1 – 18. Bristol University Press DOI: https://doi.org/10.46692/9781847421210.002

Taylor-Ritzler, T., Suarez-Balcazar, Y., Garcia-Irirarte, E., Henry, D.B., and Balcazar, F.E. (2013). Understanding and Measuring Evaluation Capacity: A Model and Instrument Validation Study. *American Journal of Evaluation*, 34 (2), 190-206.

Torres, R.T. and Preskill, H. (2001). Evaluation and Organizational Learning: Past, Present and Future. *American Journal of Evaluation*, 22 (3), 387-395.

Toulemonde, Jacques. Incentives, Constraints and Culture-building as Instruments for Development of Evaluation Demand. In R. Boyle & D. Lemaire (Eds.), *Building effective evaluation capacity: Lessons from practice*. New Brunswick, NJ: Transaction, 153-168.

Trevisan, M.S. (2007). Evaluability Assessment From 1986 to 2006. *American Journal of Evaluation* 28(3), 290-303.

Volkov, B & King, J. (2007). A Checklist for Building Organizational Evaluation Capacity. The Evaluation Center: Western Michigan University. Retrieved from https://wmich.edu/sites/default/files/attachments/u350/2018/eval-cap-bldg-volkov%26king.pdf . Accessed June 4, 2002.

Wade, Jay and Kallemeyn, Leanne. (2020). Evaluation capacity building (ECB) interventions and the development of sustainable evaluation practice: An exploratory study. *Evaluation and Program Planning* 79, 1-9.

Wholey, Joseph S. (2015). Exploratory Evaluation. In Newcomer, Kathryn E., Hatry, Harry P., and Wholey, Joseph S. eds., *Handbook of Practical Program Evaluation, 4th edition*. Hoboken, NJ: John Wiley & Sons, 88-107.

Williams, D.D. & Hawkes, M.L. (2003). Issues and Practices Related to Mainstreaming Evaluation: Where Do We Flow from Here? *New Directions for Evaluation*, 99, 63-83

Author Biographies

Salvatore P. Alaimo is Professor at Grand Valley State University, in Grand Rapids, Michigan, USA where he has taught courses in philanthropy and nonprofit administration for fifteen years. He received the Fulbright-Masaryk Scholarship in NGO Management and taught and conducted research in the Czech Republic 2022-2023. Some of his publications include book chapters in *The Jossey-Bass Reader on Nonprofit and Public Leadership, Leadership in Nonprofit Organizations: A Reference Handbook*, and the *Handbook of Research on Nonprofit Economics and Management* as well as journal articles in *International Journal of Community Well Being, New Directions for Evaluation,* and *Journal of Nonprofit Education and Leadership.* He has produced two documentaries, a feature length film *What is Philanthropy?* and a short film entitled *Sew What?.* Professor Alaimo earned his Ph.D. in Philanthropic Studies at Indiana University.

Jakub Pejcal serves as a lecturer at the Faculty of Economics and Administration at Masaryk University, in Brno, Czech Republic where he also plays an active role in the Centre for Nonprofit Sector Research. His teaching and research span a variety of topics, with a focus on the economics, accounting, and management of non-governmental nonprofit organizations. Currently, his research delves into key issues facing nonprofit sector, including financial health, transparency and accountability, evaluation, and capacity building.

Simona Smrčková is a Phd student in the Department of Public and Social Policy, within the Faculty of Social Sciences, at Charles University, in Prague, Czech Republic. Her research interests are in evaluation and evaluation capacity building.

Health and the Work Absence Gap Across Employment Sectors in the United States

Michael E. Bednarczuk - Austin Peay State University

Those in the public sector in the United States are historically known to have more work absences than those in the private sector. While long attributed to various individual-level or organizational-level characteristics, there has not been an examination of the role that physical and mental health may play in impacting that difference. Using data gathered from the National Health Interview Study, descriptive statistics found that those in the public sector tend to be in worse physical health but better mental health than those in the private sector. While Poisson models found that health did not impact the likelihood of those in the public sector having more absences than those in the private sector, a Blinder-Oaxaca decompositional analysis found that the majority of the gap in absences across sectors can be attributed to the distribution of characteristics across the private and public sectors. These findings suggest that demographic and organizational differences across the public and private sectors, not simply health, are the main determinants of the gap in work absences across sectors.

Keywords: Work absences, public sector, private sector, health, Blinder-Oaxaca decomposition

Introduction

There are a bevy of differences between the public and private sectors. These range from the more abstract, such as levels of altruism (Dur & Zoutenbier, 2015), to the more tangible, such as civic engagement (Brewer, 2003; Taylor 2010) and even frequency of blood donations (Houston, 2006). One consistent difference pertains to work absences: those in the public sector take more time off from work than those in the private sector.

There are numerous individual and organizational sources that may account for this gap. Among individual-level factors, the primary causal mechanism is attributed to laziness (Delfgaauw & Dur, 2008; Dur & Zoutenbier, 2015). At the organizational level, factors such as differences in employment protection across sectors, differences in compensation during absences, and decreased wage dispersion within the public sector are commonly cited as potential reasons for the gap (Bossaert, 2005; Kearney & Mareschal, 2014; Depalo, Giordano, & Papapetrou, 2015). However, one factor that has been under-examined in the American context is physical and mental

health. While other studies have modeled health status (Pfeifer, 2013; Riphahn, 2004), those analyses have been in countries with health care systems that are much more socialized than in the United States.

Aside from organizational differences in health care, it is also reasonable to believe that there may be differences in the individual health of those in separate sectors of the economy within the United States. For example, work has shown that extrinsic motivations can be important to those in the public sector (Park & Word, 2012; Stazyk, 2013; Van de Walle et al., 2015). Specific to this context, the benefits of the public sector can be particularly attractive to employees, and one of these benefits is generally more affordable health insurance (French & Emerson 2014). Furthermore, the public sector workplace tends to have a higher proportion of both women and the elderly, which are two groups that utilize higher amounts of health care (Congressional Research Service [CRS], 2014).

While previous studies have noted the potential importance of health on absences, they have either modeled covariates of health but not direct measures of health or have focused on countries other than the United States (Mastekaasa, 2020; Pfeifer, 2013; Riphahn, 2004). However, for this analysis, two direct measures of health can be modeled: one that captures the respondent's mental health, and another that measures the respondent's physical health, both of which are taken from a lengthy survey battery.

To examine the role of health on work absences across sectors, several years' worth of data gathered from the National Health Interview Study are analyzed. Descriptive statistics found that mental health was better among those in the public sector ($t(54,680)=5.76$, $p<0.001$). However, health was poorer among those in the public sector across the three physical health categories of upper body, lower body, and social ($t(54,680)=-2.19$, $p=0.01$; $t(54,680)=-1.81$, $p=0.04$; $t(54,680)=-1.65$, $p=0.05$). That said, Poisson models found that including all four of the aforementioned measures of health did not impact the gap in work absences across sectors. Furthermore, a Blinder-Oaxaca decompositional analysis found that about 93% of the work-absence gap across sectors could be explained by differences in characteristics across the two groups.

This article proceeds as follows. The literature is reviewed, then the hypotheses are proposed. The data are described, and then the results are presented. The article closes with a discussion and the conclusion.

Literature Review and Hypotheses

Public sector employees are known to have more work absences than their private sector counterparts. In the United States, public employees take off around 40% more time than those in the private sector (D'Amuri, 2017). This gap is constant across the globe; there is evidence for it in other countries such as Canada, Norway, and Germany (Mastekaasa, 2020; Pfeifer 2013; Uppal & LaRochelle-Cote, 2013).

Understanding the causes of absenteeism are critical for several reasons, but particularly for its effect on governance. Work absences add to the direct costs of administration, as it can lead to overtime or understaffing. Performance is also impacted by absenteeism; in a meta-analysis, Viswesvaran (2002) highlighted the negative relationship between absenteeism and effort. Given the negative consequences of absenteeism, deepening our understanding of the causes of this problem is a critical challenge.

Several explanations have been put forth to explain the source of this difference. Some scholars have focused on institutional reasons, such as a lack of incentives within the public sector to encourage attendance. Due to features such as increased employment protections or seniority systems, public employees may not believe that higher attendance would increase the likelihood of promotion, nor would absences lead to an increase in the likelihood of dismissal (Bossaert, 2005; Depalo, Giordano, & Papapetrou, 2015). When these incentives change, behavior changes as well; for example, Italian public sector workers were less likely to be absent following the passage of a law reducing sick leave compensation (De Paola, Scoppa, & Pupo, 2014). Given their current incentives, however, United States public sector employees may be more apt to take time off from work.

There are other institutional forces that may shape attendance. For example, work has demonstrated that managerial turnover can impact absenteeism; public sector employees are more apt to not be at work following the hiring of a new boss (Lokke & Sorensen, 2021). Related, relationships with management tend to reduce absenteeism (Pihl-Thingvad et al., 2022). The political environment may also indirectly shape attendance; a study of Brazilian public sector employees found that negative political attacks on the public sector had a negative effect on emotional and physical health (Lotta, Tavares, & Story, 2023). Larger organizations tend to have more absences (Barmby & Stephen, 2000). Those with paid sick leave are likely to have more absences than those who do not; furthermore, absences increase with tenure, and there can be differences in absences across occupations (Callison & Pesko, 2022; Hackett, 1990; Mastekaasa, 2020).

Other scholars have focused on differences in the distribution of underlying personal attributes to explain work attendance behaviors. Some have suggested that those in the public sector may be lazier than those in the private sector (Delfgaauw & Dur, 2008; Dur & Zoutenbier, 2015). This laziness may increase the likelihood of absence. In contrast, others have pointed to public service motivation as an attribute that may increase attendance, but evidence of its effect is either nonexistent or mixed (Gross, Thaler, & Winter, 2019; Jensen, Andersen, & Holten, 2019; Koumenta, 2015; Wright, Hassan, & Christensen, 2017).

However, what has been lacking from studies analyzing the difference in work absences across sectors among employees in the United States is a focus on health. While scholars have suggested that health differences may impact this difference (Mastekaasa, 2020), and there have been some attempts to include it in studies of other countries (Pfeifer, 2013; Riphahn, 2004), it has not been analyzed in the American context. Pfeifer (2013) found that those in the public sector had more absences than the private sector, even when modeling health, while Riphahn (2004) found that employment protections increased work absences even when modeling health.

It is worthwhile to examine the impact of health on attendance in the United States for several reasons. First, while health has been included in studies of other countries, the institutional context in the United States is unique enough that additional analysis is likely warranted. For example, in their studies, both Pfeifer (2013) and Riphahn (2004) analyze Germany. Germany has stronger employment protections for those in both the public and private sector as well as a health care system where insurance is not as tethered to employment as it commonly is in the United States; both differences could influence the distribution of employees across sectors or alter their decision-making.

Second, the health of those in the United States in the public sector merits detailed examination. There may be selection reasons for health differences across sectors. There is evidence that many are drawn to government employment in the public sector because they possess "public sector

motivation"; that is, they are motivated by more self-interested factors such as "job security, wages/salary, and fringe benefits" (French & Emerson, 2014). This is distinct from the more widely-studied "public service motivation," which analyzes the intrinsic motivations of those in the public sector (Perry & Wise, 1990; Perry 1996). Public employment may be an attractive incentive for those in poor health who may fear loss of employment in the private sector and subsequent difficulties in obtaining insurance without an employer. Related, there is also evidence that more absence-prone individuals seek out employment in the public sector (Mastekaasa, 2020).

Additionally, the demographics of the public sector suggest that there may be differences in health across sectors. The public sector typically has a higher proportion of both women and the elderly than the private sector (CRS, 2014); both groups usually have increased absences, sometimes due to health-related concerns (Martocchio, 1989; Mastekaasa & Olsen, 1998).

There may also be socialization differences that have a disparate effect on the health of those in the public sector. For example, red tape has a negative effect on both the attitudes and behaviors of government employees; furthermore, public service motivation increases the magnitude of this relationship such that red tape has a larger negative effect on those with higher public service motivation (Quratulain & Khan, 2015). Such negative attitudes and behaviors could contribute to increased work absences.

Prior studies have found that the areas of poor health of those in the public sector tend to cluster around two categories: physical and mental. In a literature review of studies on the health of public sector employees, mental and behavioral disorders (particularly depression) as well as musculoskeletal system diseases were the largest causes of absenteeism among government workers (Sampaio & Baptista, 2019).

Therefore, this study measures the effects of physical and mental health on the sector differences in absences. While the first two hypotheses test well-established findings in the literature, the third hypothesis proposes that differences in health have a demonstrable effect in minimizing the absence gap across sectors.

H1: As mental health worsens, work absences increase
H2: As physical health worsens, work absences increase
H3: When health is modeled, the gap in work absences across sectors is reduced

Data

The data used to address the effect of physical and mental health on the gap in absences across sectors comes from the National Health Interview Survey. This annual survey is administered by the National Center for Health Statistics, which is a part of the Department of Health and Human Services. It is a cross-sectional household interview survey that uses a multistage area probability design to create a sample that is representative of the United States civilian noninstitutionalized population. The survey is conducted via face-to-face interviews performed by the Census Bureau. The data used in the analysis includes those who were employed in the adult sample from the years 2014 through 2018; the descriptives are available in the appendix. Several years were analyzed to increase the sample size, but the results were robust across individual years.

The dependent variable is the number of days in the past year that the respondent was absent from work. The question used asked, "During the past 12 months, that is, since [12-month ref. date], about how many days did you miss work at a job or business because of illness or injury (do not

include maternity leave)?" Over half of the sample did not record an absence, while ninety percent who did record an absence had five absences or fewer.

The key independent variables measure the sector of employment and health. For the former, a dummy variable was created from a measure using 2012 North American Industry Classification System codes obtained from the Census. For this variable, those who noted that they were in "Public Administration Industries" were coded as 1, all else 0.

For the latter, both physical and mental health are measured. The questions were grouped together in the survey to make it clear that the relevant questions pertained to either physical or mental health. Physical health is a three-factor measure that identifies respondents who reported any difficulty with any one of twelve different functional activities. The questions used asked the following:

"By yourself, and without using any special equipment, how difficult is it for you to...

1. Walk a quarter of a mile - about 3 city blocks?
2. Walk up 10 steps without resting?
3. Stand or be on your feet for about 2 hours?
4. Sit for about 2 hours?
5. Stoop, bend, or kneel?
6. Reach up over your head?
7. Use your fingers to grasp or handle objects?
8. Lift or carry something as heavy as 10 pounds such as a full bag of groceries?
9. Push or pull large objects like a living room chair?
10. Go out to things like shopping, movies, or sporting events?
11. Participate in social activities such as visiting friends, attending clubs and meetings, going to parties?
12. Do things to relax at home or for leisure (reading, watching TV, sewing, listening to music)?"

Each question used the same Likert response scale of "Not at all difficult," "Only a little difficult," "Somewhat difficult," "Very difficult," and "Can't do at all." Each response was given an integer value with higher values corresponding to increased difficulty. The items have a high internal consistency (Cronbach's alpha = 0.87). Factor analysis revealed three underlying factors in the data: these factors were formed primarily through questions 1-5, 6-9, and 10-12. These factors were subsequently labeled "lower body health," "upper body health," and "social health," respectively. See the Appendix for more detail about the factor analysis.

Mental health is a summary measure composed of six questions pertaining to the respondent's current mental and emotional health and the extent to which those feelings interfere with her or his life or daily activities. The questions used asked the following:

"During the past 30 days, how often did you feel...

- So sad that nothing could cheer you up?
- Nervous?
- Restless or fidgety?

- Hopeless?
- That everything was an effort?
- Worthless?"

Each question used the same Likert response scale of "A lot," "Some," "A little," and "Not at all." Each response was given an integer value with higher values corresponding to increased agreement. The items have a high internal consistency (Cronbach's alpha = 0.83). Factor analysis revealed a single underlying factor, which was labeled "mental health."

Additional covariates are included to model individual and organizational characteristics. Age, race, sex, and personal income were the individual-level covariates. The presence of paid sick leave, organizational size, tenure, and occupation were modeled as well. Age was coded as a continuous variable and a squared variable (to account for the non-linear effect of aging). The remaining demographic variables were dummy coded, while income was coded as a continuous variable and was measured as the respondent's income in the previous year. Sick leave was dummy coded, while tenure was coded as a continuous variable. Organization size was a categorical variable of the number of people who worked with the respondent with the following groups of employees: 1, 2-9, 10-24, 25-49, 50-99, 100-249, 250-499, 500-999, 1000 and up. See the Appendix for a list of occupations.

Results

The data are briefly described here, with additional information available in the appendix. The average number of work absences a year was slightly more than three days. About 16% of the sample works for the public sector (N(private)=45,493, N(public)=9,289). The average age is around 42, and a majority of the sample is white, while the most common gender is male. Over half work in organizations that offer paid sick leave. The average tenure in a job is a little more than eight years, while the average income is close to $50,000.

Pairwise t-tests show that there are significant differences for each variable (see table 1 below). There are several possible explanations for these differences. It could be a statistical artifact, as larger sample sizes can lend themselves to an increased likelihood of detecting differences. However, given the magnitude and multitude of these differences, there is a possibility that some of them may also contribute to the work absence gap.

Looking at the descriptives by sector shows many similarities and differences across the sectors. Demographically, those in the public sector are, on average, just a few years older than those in the private sector. The public sector is slightly less white as well. However, there are stark differences in areas such as gender and paid sick leave: women make up a much larger percentage of the workforce in the public sector, and over 86% in the public sector have paid sick leave, compared to only 56% in the private sector.

Descriptive data suggests differences in health across sectors, but these differences are not in the same direction. Those in the public sector score lower than the private sector on the mental health measure, suggesting that those working in the government have better mental health than others. However, public sector workers averaged considerably higher scores on all three measures of physical health. These results suggest that public sector employees may have worse physical health than those in other sectors of the economy.

To model the data, since the number of absences is a count variable with a variance larger than its mean, a Poisson model with year fixed effects and standard errors clustered at the region-level was used. This type of model estimates conditional means that are robust to overdispersion and impose

more general and less strict assumptions than negative binomial estimates. Year fixed-effects account for year-specific shocks to absenteeism, such as a bad flu season, while the standard errors are clustered at the level at which the data forms a panel. The results are weighted to include post-stratification adjustments using Census Bureau population control totals. Both models are estimated on the same analytic sample. The results from the first model exclude the health measures, while they are included in the second model; this is to highlight the effect of these covariates on work absences.

Table 1. Demographic Descriptive Statistics by Sector With T-Test Results

	Public (n=9,289)		Private (45,493)			
Variable	M	SD	M	SD	t	p
Work Absences	3.760	13.310	3.196	14.189	5.760	<0.001
Mental Health	-0.056	0.884	-0.030	0.955	-3.720	<0.001
Lower Body Health	0.071	1.028	0.013	0.943	8.110	<0.001
Upper Body Health	-0.001	0.019	-0.031	0.004	80.050	<0.001
Social Health	-0.010	0.982	-0.048	0.953	5.300	<0.001
Age	44.533	13.035	41.858	13.868	26.700	<0.001
Square of Age	2153.056	1181.904	1944.383	1221.099	23.700	<0.001
White	0.779	0.415	0.799	0.400	-6.670	<0.001
Male	0.441	0.496	0.556	0.497	-31.820	<0.001
Paid Sick Leave	0.861	0.346	0.555	0.497	95.810	<0.001
Tenure	10.111	9.131	7.931	8.868	32.580	<0.001
Personal Income	51071.750	33392.250	47849.010	37184.350	11.940	<0.001

Note. M = Mean; SD = Standard Deviation. T-tests were conducted with unequal variances assumed due to the large difference in sample sizes.

It is worthwhile to describe how to interpret the incidence rate ratios (IRR) in table 2 below. A coefficient greater than one shows that the variable has a positive effect on work absences; that is, it increases work absences. Likewise, a coefficient less than one shows that the variable has a negative effect on work absences.

Looking first at the demographic variables in the more fully specified model, both measures of age were significant, showing that absences increase with age, but that this effect slows over time. This could be due to selection effects, as those who are elderly and in poor health may be more apt to retire. Race, sex, and income had no effect. The organizational variables suggest that employees at organizations that offer paid sick leave are more likely to have more absences than those that do not, as are those with longer tenures. As organization size increases, absences increase among

those in the largest organizations, while various occupations had more absences than others.[1]

Most relevant to this manuscript, there is no work absence gap across either model. In the first model, which does not include either health measure, those working in the public sector take about 20% more days off of work than those in the private sector (3.19 days versus 3.83 days). However, when this difference is plotted, there is a considerable overlap of the confidence intervals; as was discussed in Belia et al. (2005), if the confidence intervals overlap by more than a quarter of their total length, the difference between the groups is not statistically significant.[2]

Additionally, there is little change when the predicted probabilities of employment sector on work absences are plotted when controlling for health status (3.20 days versus 3.76 days); furthermore, the degree of overlap of the confidence intervals suggests that this difference is not statistically significant, either. That said, all four measures of health are significant in the second model; as either physical or mental health declines, absences increase, which supports the first two hypotheses.

To better understand the relationship between health and work absences across sectors, the predicted probabilities of work absences at different levels of health (e.g., poor, average, good) were reviewed. Across all four health factors, there was no statistically significant effect of different levels of health on the work absence gap across sectors. In other words, health does not appear to be shaping the work absence gap at any demonstrable level.

Table 2. Poisson Regression Models of United States Work Absences by Health, 2014-2018

Variable	IRR	95% CI	P>z	IRR	95% CI	P>z
Mental Health				1.152*	[1.105,1.202]	<0.001
Lower Body Health				1.258*	[1.233,1.283]	<0.001
Upper Body Health				1.162*	[1.101,1.225]	<0.001
Social Health				1.157*	[1.103,1.214]	<0.001
Public Employment	1.200*	[1.142, 1.261]	<0.001	1.175*	[1.101,1.254]	<0.001
Age	1.062*	[1.044, 1.079]	<0.001	1.051*	[1.033,1.07]	<0.001
Square of Age	0.999*	[0.999, 1]	<0.001	0.999*	[0.999,1]	<0.001
White	1.019	[0.839, 1.239]	0.842	1.052	[0.89,1.245]	0.549
Male	0.833*	[0.695, 0.998]	0.048	0.945	[0.79,1.13]	0.538
Paid Sick Leave	1.052	[0.914, 1.21]	0.484	1.149*	[1.033,1.277]	0.01
Tenure	1.005*	[1.002, 1.008]	0.002	1.006*	[1.004,1.009]	<0.001
Personal Income	1.000*	[1, 1]	<0.001	1.000	[1,1]	0.069
Organization Size						
2-9 Employees	0.972	[0.78, 1.211]	0.797	0.978	[0.809,1.182]	0.815
10-24 Employees	1.085	[0.83, 1.42]	0.549	1.069	[0.753,1.517]	0.708
25-49 Employees	1.106	[0.852, 1.435]	0.45	1.054	[0.827,1.345]	0.672
50-99 Employees	1.040	[0.794,	0.796	0.998	[0.74,1.346]	0.98

		1.362]				

100-249 Employees	1.080	[0.743, 1.569]	0.687	1.028	[0.663,1.596]	0.901
250-499 Employees	1.335	[0.858, 2.078]	0.202	1.294	[0.877,1.908]	0.197
500-999 Employees	1.308*	[1.005, 1.701]	0.046	1.215	[0.909,1.625]	0.191
1000+ Employees	1.313*	[1.105, 1.559]	0.002	1.238*	[1.001,1.532]	0.053
Occupation Type						
Business and Financial Operations Occupations	0.954	[0.732, 1.245]	0.742	0.969	[0.758,1.239]	0.814
Computer and Mathematical Occupations	1.167	[0.807, 1.687]	0.411	1.218	[0.831,1.786]	0.31
Architecture and Engineering Occupations	0.810	[0.618, 1.061]	0.13	0.839	[0.653,1.076]	0.171
Life, Physical, and Social Sciences Occupations	0.975	[0.689, 1.379]	0.893	0.923	[0.606,1.405]	0.712
Community and Social Service Occupations	1.153	[0.801, 1.659]	0.439	1.188	[0.82,1.72]	0.359
Legal Occupations	1.015	[0.757, 1.361]	0.911	1.012	[0.807,1.268]	0.909

Education, Training, and Library Occupations	0.799	[0.619, 1.031]	0.088	0.842	[0.652,1.087]	0.191
Arts, Design, Entertainment, Sports, and Media Occupations	1.014	[0.756, 1.359]	0.921	1.046	[0.778,1.406]	0.759
Healthcare Practitioners and Technical Occupations	1.027	[0.751, 1.405]	0.859	1.034	[0.724,1.476]	0.848
Healthcare Support Occupations	1.175	[0.91, 1.517]	0.205	1.150	[0.888,1.49]	0.28
Protective Service Occupations	1.060*	[1.02, 1.102]	0.002	0.963	[0.817,1.135]	0.657
Food Preparation and Serving Related Occupations	1.127	[0.984, 1.29]	0.082	1.148	[0.98,1.344]	0.086
Building and Grounds Cleaning and Maintenance Occupations	1.347	[0.881, 2.059]	0.166	1.233	[0.847,1.796]	0.271
Personal Care and Service Occupations	1.035	[0.819, 1.307]	0.754	0.953	[0.711,1.279]	0.766
Sales and Related Occupations	1.030	[0.928, 1.144]	0.565	1.022	[0.895,1.167]	0.737

Office and Administrative Support Occupations	1.109	[0.912, 1.348]	0.311	1.083	[0.855,1.371]	0.521
Farming, Fishing, and Forestry Occupations	0.834	[0.65, 1.071]	0.156	0.902	[0.762,1.067]	0.236
Construction and Extraction Occupations	1.695*	[1.572, 1.828]	<0.001	1.651*	[1.571,1.734]	<0.001
Installation, Maintenance, and Repair Occupations	1.726*	[1.411, 2.111]	<0.001	1.700*	[1.342,2.153]	<0.001
Production Occupations	1.615	[1.23, 2.122]	<0.001	1.589*	[1.266,1.993]	<0.001
Transportation and Material Moving Occupations	1.742*	[1.372, 2.212]	<0.001	1.640*	[1.316,2.045]	<0.001
Military Specific Occupations	0.688*	[0.587, 0.807]	<0.001	0.753*	[0.64,0.886]	0.001
Year						
2015	0.949	[0.808, 1.113]	0.511	0.861	[0.731,1.014]	0.072
2016	0.942	[0.788, 1.126]	0.519	0.888	[0.726,1.086]	0.25
2017	1.021	[0.916, 1.138]	0.676	0.996	[0.928,1.069]	0.919
2018	1.132	[0.896, 1.431]	0.3	1.032	[0.805,1.323]	0.807
Constant	0.697	[0.48, 1.013]	0.058	0.803	[0.585,1.102]	0.177

* = p<0.05
Note: IRR = Incidence Rate Ratio; 95% CI = 95% Confidence Interval. The excluded category for "Organization Size" is "1 Employee" while the excluded category for "Occupation Type" is "Management Occupations"

Additionally, a Blinder-Oaxaca decomposition analysis was performed. This statistical technique is useful for examining the difference in the means of a dependent variable between two groups: in this case, it examined the difference in absences between the private and public sectors (Blinder, 1973; Oaxaca, 1973). It can separate the differences in means into a part that is "explained" by differences in group characteristics and another part that cannot be explained by those characteristics. The analysis showed that around 93% of the gap could be explained by differences in characteristics across the two groups. That is, if public and private sector members had identical characteristics, the gap in absences would be largely diminished. That said, there is still a roughly 7% part of the gap that is not attributed to those differences in characteristics. The Blinder-Oaxaca decomposition analysis therefore suggests that a primary driver of the absence gap lies in the different distribution of characteristics across the public and private sectors. See the appendix for a table of these results.

Given these unexpected results, an additional model was examined that excluded those in the private sector; the results of this model can be seen in table 3 below. There are interesting contrasts that are found in this data. For example, among the demographic characteristics, age is not a significant factor, while men are more likely to be absent than women. This latter factor is the opposite of earlier findings in the literature. Turning to additional factors, while paid sick leave increased the likelihood of being absent, tenure had no effect. There are also differences in the occupations; those in business and financial operations occupations are more likely to be absent, while those in education, training, and library occupations are less likely to be absent.
This suggests that the unexpected results may be driven by factors such as the decreased likelihood of women and educators to be absent in the public sector.

Table 3. Poisson Regression Models of United States Public Sector Work Absences by Health, 2014-2018

Variable	IRR	95% CI	P>z
Mental Health	1.138*	[1.048,1.236]	0.002
Lower Body Health	1.319*	[1.298,1.342]	<0.001
Upper Body Health	1.143*	[1.078,1.212]	<0.001
Social Health	1.106*	[1.016,1.204]	0.019
Age	1.056	[0.976,1.142]	0.177
Square of Age	0.999	[0.999,1]	0.19
White	0.980	[0.722,1.331]	0.898
Male	1.047*	[1.033,1.061]	<0.001
Paid Sick Leave	1.490*	[1.242,1.789]	<0.001
Tenure	1.002	[0.991,1.013]	0.685
Personal Income	1.000	[1,1]	0.664
Organization Size			
2-9 Employees	1.606	[0.731,3.525]	0.238

10-24 Employees	2.095	[0.838,5.243]	0.114
25-49 Employees	2.909	[0.961,8.807]	0.059
50-99 Employees	1.755	[0.679,4.535]	0.245
100-249 Employees	1.933	[0.884,4.227]	0.099
250-499 Employees	2.056	[0.906,4.667]	0.085
500-999 Employees	2.622	[0.984,6.985]	0.054
1000+ Employees	1.861	[0.853,4.058]	0.119
Occupation Type			
Business and Financial Operations Occupations	1.373*	[1.135,1.659]	0.001
Computer and Mathematical Occupations	0.919	[0.608,1.389]	0.689
Architecture and Engineering Occupations	0.932	[0.625,1.39]	0.731
Life, Physical, and Social Sciences Occupations	0.679	[0.34,1.358]	0.274
Community and Social Service Occupations	1.414	[0.78,2.564]	0.254
Legal Occupations	0.930	[0.761,1.137]	0.479
Education, Training, and Library Occupations	0.725*	[0.559,0.94]	0.015
Arts, Design, Entertainment, Sports, and Media Occupations	0.732	[0.508,1.055]	0.095
Healthcare Practitioners and Technical Occupations	0.852	[0.693,1.048]	0.129
Healthcare Support Occupations	2.562	[0.781,8.405]	0.121
Protective Service Occupations	1.047	[0.903,1.216]	0.542
Food Preparation and Serving Related Occupations	1.658*	[1.094,2.514]	0.017
Building and Grounds Cleaning and Maintenance Occupations	1.198*	[1.029,1.395]	0.02
Personal Care and Service Occupations	1.226	[0.699,2.149]	0.477
Sales and Related Occupations	0.561	[0.18,1.751]	0.319
Office and Administrative Support Occupations	1.151	[0.814,1.626]	0.426

Farming, Fishing, and Forestry Occupations	0.458	[0.186,1.13]	0.09
Construction and Extraction Occupations	1.347	[0.733,2.479]	0.337
Installation, Maintenance, and Repair Occupations	0.954	[0.51,1.784]	0.882
Production Occupations	1.098	[0.44,2.74]	0.841
Transportation and Material Moving Occupations	1.319	[0.634,2.745]	0.459
Military Specific Occupations	0.663*	[0.564,0.78]	<0.001
Year			
2015	0.727	[0.594,0.89]	0.002
2016	0.779	[0.653,0.929]	0.006
2017	0.771	[0.703,0.845]	<0.001
2018	0.948	[0.726,1.237]	0.693
Constant	0.392	[0.037,4.173]	0.438

* = $p < 0.05$
Note: IRR = Incidence Rate Ratio; 95% CI = 95% Confidence Interval. The excluded category for "Organization Size" is "1 Employee" while the excluded category for "Occupation Type" is "Management Occupations"

In summary, several conclusions can be made about the role of health on work absences across sectors. Those in the public sector appear to be in worse physical health, but better mental health, than those in the private sector. There is no difference in the work absences across sectors, regardless of health status. Finally, public sector employees of the same health status are predicted to have just as many work absences as their private sector counterparts.

Discussion

It is important to explore why there is no work absence gap across sectors. This stands in contrast to the typical findings both from the United States and around the world. Given the time frame under analysis, a likely source could be longitudinal changes to the composition of the sectors and to the institutions within those sectors. Other studies relied on data from earlier points in time; perhaps the types of people across sectors, as well as the rules governing them, have changed over time. If so, then studies relying on older data may reach a different conclusion. For example, the Affordable Care Act, passed in 2010, sought to make private insurance more affordable, thus weakening the link between employment and insurance. Perhaps this change impacted people's preference for work in the private sector. Additional studies should look for any potential changes in the types of people in those sectors, as well as the structures governing them.

Another factor potentially influencing the findings may lie in the lack of variance in the dependent variable. Regardless of sector, most individuals do not miss much work; as was referenced earlier, over half of the sample did not miss any days of work at all in the past year. The point estimates from the predicted probabilities suggest a potential difference between sectors that only amounts

to a fraction of a day. Future studies may wish to examine this gap in different countries, or perhaps during times when the workforce may have more absences than usual (pandemics, for example).

Also, there may be additional variations within the United States that may impact these relationships. Unionization rates are usually greater among public sector employees than private sector employees, which may lead to more favorable leave policies among the former group.States also have different regulatory environments which could shape patterns of leave. Also, there may be a seasonality to work intensity in the public sector which impacts the timing of absences. Other work has also been able to include variables for factors such as education (Mastekaasa, 2020). Later work could also examine these potential differences.

Turning now to other findings from the data, why may those in the public sector be in worse physical health than those in the private sector? It is possible that the work itself in the public sector has a negative effect on the physical health of those within it. Given that the model accounted for occupation type, however, this conclusion does seem unlikely. That said, work has shown that several occupational stressors can contribute to poor health; a meta-analysis concluded that organizational constraints and interpersonal conflict have the largest effects on health, with the primary symptoms relating to sleep disturbances and gastrointestinal problems (Nixon et al., 2011). While prior studies have attempted to analyze the selection of absence-prone individuals to the public sector (Bohm & Riedel, 2013; Riphahn, 2004), they have not included measures of health in their models. Future work should explore the attraction and socialization consequences of this finding.

One potential shortcoming of this analysis is the reliance on recall data for the number of absences. People may misremember how often they were not at work due to their health. Official data from the respondent's place of work would overcome this obstacle, and future studies may wish to obtain this information. However, the use of the variable in this analysis may only be a complication to the extent that workers in different sectors have different tendencies to misremember their absences; if there are no systematic differences in recall, then this may be a minimal problem.

The measures of health used may also be incomplete. For example, the mental health variable only asked how the respondent felt over the past 30 days, while absences covered the past year; recent problems may not have had time to impact absences, while recently improved health may mask poorer health from earlier in the year that contributed to absences. That said, such a bias may potentially make it harder to find the hypothesized effect. Ensuring that the timelines for both measures are the same would add an extra measure of confidence in the results of future studies.

Relatedly, the dependent variable in the model captures absences from both illness and injury. Decomposing this into its constituent parts would allow for a more finely-grained analysis concerning work absences: for example, are specific health conditions related to different types of absences? Additional studies would benefit from the inclusion of such measures.

Finally, workers may be absent for other reasons. Government employees may have time off for federal holidays that are not observed by private sector employees; for example, Juneteenth was initially recognized as a federal holiday in 2021, yet many businesses are still open on that day. Public sector workers may also be awarded more "personal days" than those in the private sector. Maternity leave is also excluded by the measure used in this study, and the public sector has a much larger percentage of women than the private sector. A measure that takes a broader account of the employment patterns within and across sectors may better capture these sources of variation.

Conclusion

Those in the public sector in the United States had been known to have more work absences than those in the private sector. This manuscript set out to examine the role of physical and mental health in impacting that difference. Data gathered from years of surveys in the United States was used to analyze this problem. Among the findings, this study found that those in the public sector generally had worse physical health but better mental health than those in the private sector. However, regardless of health status, there was no difference in work absences across sectors.

This article does demonstrate some support for the arguments put forth by Goodsell (2014) in defense of bureaucrats. Goodsell (2014) argues that "critics have overblown the faults and misdeeds of our bureaucrats" and that those in the public sector should instead be "recognized for the extent to which they embody principles essential to effective governance" (pp. 118-119).

Most relevant, Goodsell (2014) states that bureaucrats "are willing to make a personal commitment to that career" (p. 119). The findings in this manuscript suggest that, even with an increased likelihood of having a job with sick leave, those in the public sector are no more likely to miss work than those in the private sector. Perhaps this means that such benefits are not as prone to abuse as their skeptics may argue.

Furthermore, women and teachers in the public sector are among those who are the least likely to miss work, which is a direct challenge to popular conceptions. In a time when public education faces increasing skepticism, this article suggests that detractors may wish to direct the ire away from narratives surrounding professional absences.

Future studies of absences in the public sector should expand beyond including measures of health. There is still much to learn about different types of absences, perhaps due to maternity leave or due to federal holidays. Relatedly, additional studies could focus on differences within particular occupational sectors to gain a deeper understanding of the nuances of work absences.

More attention could also be paid to the consequences of the work absences. Studies could focus on the impact to healthcare spending, or the effect on governance. For example, if the public sector shared the same demographics as the private sector, how would health care spending on public employees change? How do bureaucracies ensure that they are meeting all of their obligations when absences vary? As the public sector faces challenges with respect to recruitment, and with an aging population, the importance of such issues continues to grow.

Notes

1. Those occupations that were more physical in nature tended to see increased absences, such as construction and extraction occupations, or installation, maintenance, and repair occupations. Notably, these sorts of careers are concentrated more heavily in the private sector than the public sector. For example, production occupations make up over six percent of the private sector sample but just around one percent of the public sector sample. The skewed distribution of more absence-prone occupations shows the importance of modeling the effects of this variable.

2. See the Appendix for figure 1 and figure 2 of the predicted probabilities.

Disclosures

The authors declare that there are no conflicts of interest that relate to the research, authorship, or publication of this article.

References

Barmby, T., & Stephen, G. (2000). Worker absenteeism: why firm size may matter. The Manchester School, 68(5), 568-577.

Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. Psychological methods, 10(4), 389.

Blinder, A. S. (1973). Wage discrimination: reduced form and structural estimates. Journal of Human resources, 436-455.

Bohm, T, & Riedel, N. 2013. On selection into public civil service. CESifo Economic Studies 59:34–71.

Bossaert, D. 2005. The flexibilisation of the employment status of civil servants: From life tenure to more flexible employment relations? Luxembourg: European Institute of Public Administration.

Brewer, G. A. (2003). Building social capital: Civic attitudes and behavior of public servants. Journal of Public Administration Research and Theory, 13(1), 5-26.

Callison, K., & Pesko, M. F. (2022). The effect of paid sick leave mandates on coverage, work absences, and presenteeism. Journal of Human Resources, 57(4), 1178-1208.

Congressional Research Service. 2014. Selected Characteristics of Private and Public Sector Workers. https://sgp.fas.org/crs/misc/R41897.pdf

D'Amuri, F. 2017. Monitoring and disincentives in containing paid sick leave. Labour Economics 49:74–83.

Delfgaauw, J., & Dur, R. 2008. Incentives and workers' motivation in the public sector. Economic Journal 118:171–91.

De Paola, M., Scoppa, V., & Pupo, V. (2014). Absenteeism in the Italian public sector: The effects of changes in sick leave policy. Journal of Labor Economics, 32(2), 337-360.

Depalo, D., Giordano, R., & Papapetrou, E. 2015. Public–private wage differentials in euro-area countries: Evidence from quantile decomposition analysis. Empirical Economics 49:985–1015.

Dur, R., & Zoutenbier, R. 2015. Intrinsic motivations of public sector employees: Evidence for Germany. German Economic Review 16:343–66

French, P.E., & Emerson, M.C. (2014) Assessing the Variations in Reward Preference for Local Government Employees in Terms of Position, Public Service Motivation, and Public Sector Motivation, Public Performance & Management Review, 37:4, 552-576, page 571, DOI: 10.2753/PMR1530-9576370402

Goodsell, C. T. (2014). The new case for bureaucracy. CQ press.

Gross, H. P., Thaler, J., & Winter, V. (2019). Integrating public service motivation in the job-demands-resources model: An empirical analysis to explain employees' performance, absenteeism, and presenteeism. International Public Management Journal, 22(1), 176-206.

Hackett, R. D. (1990). Age, tenure, and employee absenteeism. Human Relations, 43(7), 601-619.

Houston, D. J. (2006). "Walking the walk" of public service motivation: Public employees and charitable gifts of time, blood, and money. Journal of public administration research and theory, 16(1), 67-86.

Jensen, U. T., Andersen, L. B., & Holten, A. L. 2019. Explaining a dark side: Public service motivation, presenteeism, and absenteeism. Review of Public Personnel Administration. doi:10.1177/0734371X17744865

Kearney, R. C., & Mareschal, P. M. 2014. Labor Relations in the Public Sector, Fifth Edition. London: CRC Press.

Koumenta, M. 2015. Public service motivation and organizational citizenship. Public Money & Management 35:341–8.

Lokke, A. K., & Sorensen, K. L. (2021). Top management turnover and its effect on employee absenteeism: Understanding the process of change. Review of Public Personnel Administration, 41(4), 723-746.

Lotta, G., Tavares, G. M., & Story, J. (2023). Political attacks and the undermining of the bureaucracy: The impact on civil servants' well-being. Governance.

Martocchio, J. J. (1989). Age-related differences in employee absenteeism: a meta-analysis. Psychology and aging, 4(4), 409.

Mastekaasa A. (2020). Absenteeism in the public and private sector: Does the public sector attract high absence employees? Journal of Public Administration Research and Theory. 2020; 30(1): 60-76. https://doi.org/10.1093/jopart/muz003

Mastekaasa, A., & Olsen, K. M. (1998). Gender, absenteeism, and job characteristics: a fixed effects approach. Work and Occupations, 25(2), 195-228.

Nixon, A. E., Mazzola, J. J., Bauer, J., Krueger, J. R., & Spector, P. E. (2011). Can work make you sick? A meta-analysis of the relationships between job stressors and physical symptoms. Work & Stress, 25(1), 1-22.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. International economic review, 693-709.

Park S. M., Word J. (2012). Driven to service: Intrinsic and extrinsic motivation for public and nonprofit managers. Public Personnel Management, 41, 705-734

Perry, J. L., & Wise, L. R. (1990). The motivational bases of public service. Public administration review, 367-373.

Perry, J. L. (1996). Measuring public service motivation: An assessment of construct reliability and validity. Journal of public administration research and theory, 6(1), 5-22.

Pfeifer, C. 2013. Cyclical absenteeism among private sector, public sector and self-employed workers. Health Economics 22:366–70.

Pihl-Thingvad, S., Winter, V., Schelde Hansen, M., & Willems, J. (2022). Relationships matter: how workplace social capital affects absenteeism of public sector employees. Public Management Review, 1-28.

Quratulain S, & Khan AK. (2015). Red Tape, Resigned Satisfaction, Public Service Motivation, and Negative Employee Attitudes and Behaviors: Testing a Model of Moderated Mediation. Review of Public Personnel Administration. 2015;35(4):307-332. doi:10.1177/0734371X13511646

Riphahn, R. T. 2004. Employment protection and effort among German employees. Economics Letters 85:353–7.

Sampaio, E., & Baptista, J. S. (2019). Absenteeism of Public Workers—Short Review. In Occupational and Environmental Safety and Health (pp. 345-353). Springer, Cham.

Stazyk E. C. (2013). Crowding out public service motivation? Comparing theoretical expectations with empirical findings on the influence of performance-related pay. Review of Public Personnel Administration, 33, 252-274.

Taylor, J. (2010). Public service motivation, civic attitudes and actions of public, nonprofit and private sector employees. Public administration, 88(4), 1083-1098.

Uppal, S., & LaRochelle-Côté, S. 2013. Understanding public–private sector differences in work absences. Insights on Canadian Society. Ottawa: Statistics Canada, Catalogue no. 75-006-X.
https://www150.statcan.gc.ca/n1/pub/75-006-x/2013001/article/11862-eng.htm

Van de Walle S., Steijn B., & Jilke S. (2015). Extrinsic motivation, PSM and labour market characteristics: A multilevel model of public sector employment preference in 26 countries. International Review of Administrative Sciences, 81, 833-855.

Viswesvaran, C. (2002). Absenteeism and measures of job performance: A meta-analysis. International Journal of Selection and Assessment, 10(1-2), 12-17.

Wright, B. E., Hassan, S., & Christensen, R. K. 2017. Job choice and performance: Revisiting core assumptions about public service motivation. International Public Management Journal 20:108–31

Author Biography

Dr. Michael E. Bednarczuk is an assistant professor of political science and public management in the Department of Political Science & Public Management at Austin Peay State University. He studies political behavior and public service.

Appendix 1. Descriptive Statistics

Variable	M	SD	Min	Max
Work Absences	3.290	14.066	0	365
Mental Health	-0.034	0.944	-0.631	6.932
Lower Body Health	0.022	0.957	-6.413	9.859
Upper Body Health	-0.042	0.958	-6.672	15.190
Social Health	-0.023	0.959	-3.993	19.700
Public Employment	0.160	0.367	0	1
Age	42.284	13.776	18	84
Square of Age	1977.711	1217.479	324	7056
White	0.797	0.403	0	1
Male	0.538	0.499	0	1
Paid Sick Leave	0.604	0.489	0	1
Tenure	8.279	8.949	0	35
Personal Income	48352.570	36613.650	1	149000
Organization Size				
2-9 Employees	0.168	0.374	0	1
10-24 Employees	0.146	0.353	0	1
25-49 Employees	0.117	0.322	0	1
50-99 Employees	0.111	0.314	0	1
100-249 Employees	0.116	0.321	0	1
250-499 Employees	0.067	0.250	0	1
500-999 Employees	0.067	0.250	0	1
1000+ Employees	0.112	0.316	0	1
Occupations				
Business and Financial Operations Occupations	0.057	0.233	0	1
Computer and Mathematical Occupations	0.040	0.195	0	1
Architecture and Engineering Occupations	0.025	0.156	0	1

Life, Physical, and Social Sciences Occupations	0.012	0.108	0	1
Community and Social Service Occupations	0.020	0.139	0	1
Legal Occupations	0.013	0.112	0	1
Education, Training, and Library Occupations	0.065	0.246	0	1
Arts, Design, Entertainment, Sports, and Media Occupations	0.023	0.151	0	1
Healthcare Practitioners and Technical Occupations	0.063	0.243	0	1
Healthcare Support Occupations	0.022	0.145	0	1
Protective Service Occupations	0.022	0.146	0	1
Food Preparation and Serving Related Occupations	0.046	0.210	0	1
Building and Grounds Cleaning and Maintenance Occupations	0.036	0.186	0	1
Personal Care and Service Occupations	0.032	0.175	0	1
Sales and Related Occupations	0.096	0.294	0	1
Office and Administrative Support Occupations	0.119	0.324	0	1
Farming, Fishing, and Forestry Occupations	0.006	0.079	0	1
Construction and Extraction Occupations	0.050	0.218	0	1
Installation, Maintenance, and Repair Occupations	0.033	0.179	0	1
Production Occupations	0.056	0.230	0	1
Transportation and Material Moving Occupations	0.056	0.230	0	1
Military Specific Occupations	0.001	0.037	0	1

Note. M = Mean; SD = Standard Deviation; Min = Minimum; Max = Maximum

**Appendix 2.** Blinder-Oaxaca Decomposition of the Work Absence Gap

	Coef.	Std. Err.	Percent
Endowments	0.568	0.018	92.71
Coefficients	0.045	0.027	7.29
Interaction	0.612	0.022	

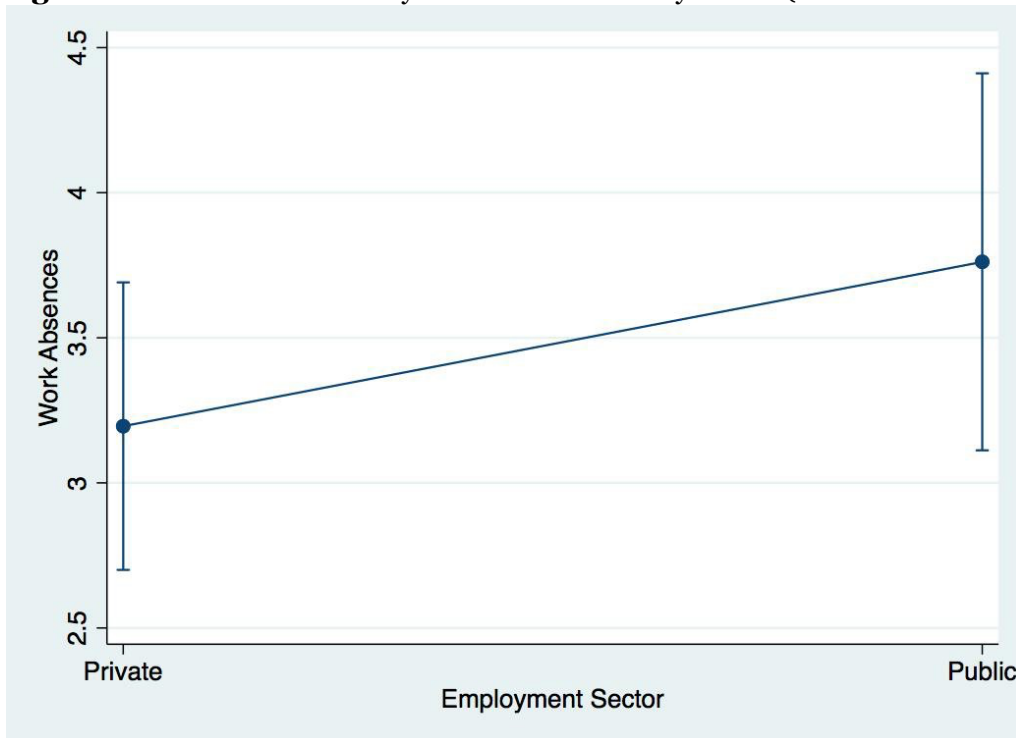**Appendix 3.** *Factor Analysis of Physical and Mental Health Components*

Variable	Mental	Lower	Upper	Social
sad	0.2407	0.01568	-0.03026	-0.03846
nervous	0.22507	-0.03291	0.01268	-0.02885
restless	0.21325	-0.02652	0.04259	-0.04819
hopeless	0.25356	-0.01295	-0.04121	-0.00797
effort	0.22473	0.01534	-0.02936	-0.0196
worthless	0.23392	-0.03465	-0.03633	0.01931
walk	-0.02445	0.35005	-0.13575	0.02234
climb	-0.02578	0.33501	-0.12965	0.02081
stand	0.01522	0.43703	-0.17593	-0.21847
sit	0.00426	0.09323	0.14456	-0.02065
stoop	0.00263	0.30468	0.06578	-0.13263
reach	-0.01812	-0.10563	0.48715	-0.10214
grasp	-0.00919	-0.13683	0.53304	-0.1315
carry	-0.0371	0.01259	0.2597	0.03683
push	-0.02789	0.10501	0.20802	-0.00587
shop	-0.04041	-0.00255	-0.15487	0.40434
social	-0.01906	-0.09813	-0.18072	0.48571
relax	-0.01439	-0.28456	0.03144	0.42189

Note: These variables are referenced in the manuscript when the questions used to measure physical and mental health are described; they are abbreviated here for space concerns. The table displays the factor loadings for the analysis.

Figure 1: Predicted Probability of Work Absences By Sector (Without Health Variables)
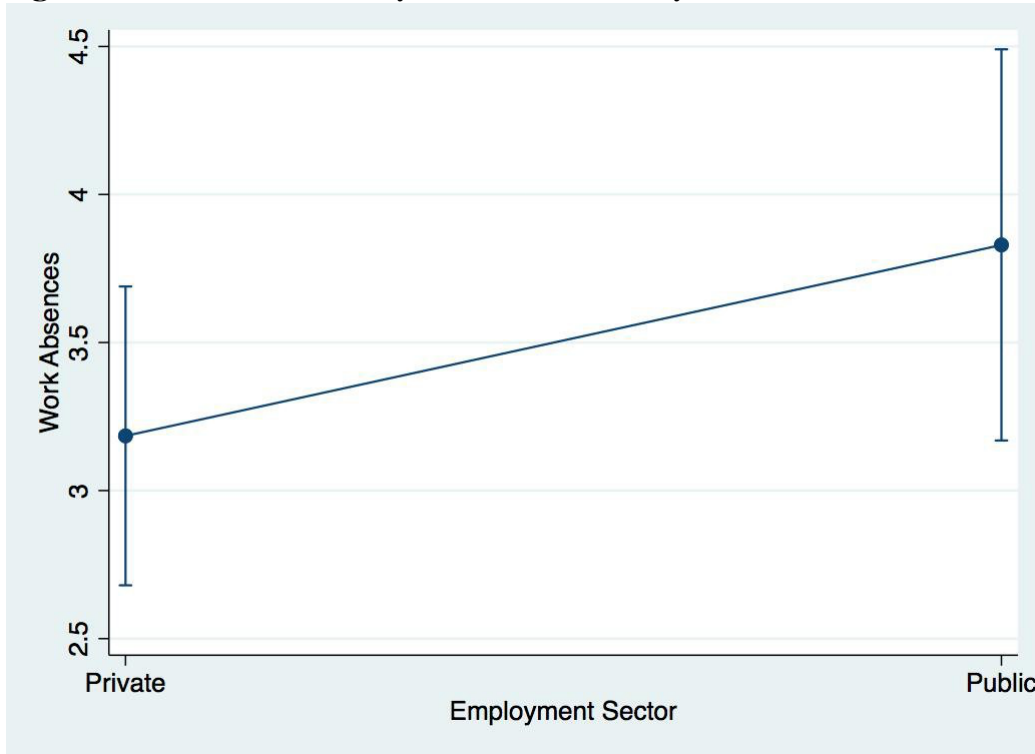
Figure 2: Predicted Probability of Work Absences by Sector (With Health Variables)