Evaluating Youth Development Nonprofits: How Performance Measurement Leads to Effective Performance Management

Ryan Paquette

Advisor Beth Gazley

Abstract

Successful nonprofits regularly engage in performance measurement and data collection. These organizations have an obligation to fulfill their mission, and collecting and using data in the most effective way can improve their ability to do that. Currently, there is a lack of research on the best practices in performance measurement specifically concerning youth development organizations. Interviews with five youth development organizations in Monroe County provide examples of patterns in the most effective performance measurement systems. Specifically, this study points to the importance of measuring a range of data, including outputs and outcomes for both managerial and program measures. Among the organizations, there was a positive relationship between investment in training staff in data analysis and the use of data in organizational decision making. Additionally, the organizations reported a consistent struggle to engage frontline staff in performance measurement, no matter the amount of top-down investment. These lessons from Monroe County organizations are likely applicable nationally to similar organizations with similar missions. The study gives youth development organizations insight into how they can improve performance measurement, which ultimately leads to improving children's lives.

I. Introduction

An increasing call for accountability in the nonprofit sector has created a focus on organizational effectiveness over the past few decades (Worth, 2021). Nonprofits must find ways to demonstrate their success to stakeholders, such as grant-makers, current donors, and clients, but it can be complicated to do so as mission driven organizations. Nonprofits' success is determined by a double bottom line. Of course, to run successful programs they must rely on their financial bottom line, ensuring they are profitable enough to provide their services. However, their success is also measured by their ability to achieve a public-benefit mission

(Worth, 2021). Measuring social good and achievement of a mission can be complicated. A variety of indicators and outcomes demonstrate nonprofit programs are working. This evaluation is known as performance measurement. It is a crucial tool nonprofits use to demonstrate accountability, improve their programs, and ultimately create public benefit.

It is now commonplace for nonprofit organizations to collect data, whether due to contractual obligations with donors, or to create accountability to the public. However, not all organizations revisit this data, many leaving it as untouched computer files. When organizations use their data as part of a performance management plan, they analyze it, draw conclusions, and change their programs in response. The data informs them on ways they can better serve their community.

There is growing empirical research assessing how nonprofits are using performance measurement to improve their effectiveness. More research in this area is needed, as improving performance measurement systems will lead to more effective programs, better use of organizational finances, and a better understanding of how to achieve specific outcomes. The following literature review summarizes some of the common conclusions from prior performance measurement research in youth development programming.

Five youth development nonprofits in Monroe County were interviewed in order to determine the characteristics of effective performance measurement systems in this subsector of nonprofit organizations. While a limited sample, the data from these interviews provided valuable information on what types of measures youth development organizations think are important, how mandated data collection affects organizational decision-making, and how job structure can create a top down emphasis on evaluation and data analysis. The organizations that were more likely to use performance measurement in their strategic thinking showed that evaluation means more than collecting data. Nonprofits must show a commitment to continuously improving their performance measurement systems in order to better serve the youth in their community.

II. Literature Review

In order to improve programs and demonstrate success, nonprofits must collect data. These objective, data driven, measures of an organization's impact are known as performance measurement. The separate effort to manage and make use of the information generated from performance measurement activities is known as performance management (Positer et al., 2014). The distinction between these two concepts is important in understanding organizational evaluation, as it can be difficult for nonprofits to effectively utilize data collected through performance measurement in performance management systems (Lee & Nowell, 2015; de Lancer Julnes & Holzer, 2001). The difficulty lies in moving from adoption of these systems to actual implementation. Organizations appear to have an easier time adopting performance measurement systems, that is defining measures and goals, but then struggle to implement these systems in a usable manner (Positer et al., 2014; de Lancer Julnes & Holzer, 2001).

Performance measurement systems involve people, tools, and indicators used to measure success. Taking all these factors into account, there are best practices organizations can use to promote the use of performance measurement in their decision making. There is a significant body of theoretical research on the best performance measurement models, with nearly all scholars agreeing that nonprofits' approach to measurement must be composed of several dimensions (Lee & Nowell, 2015; LeRoux & Wright, 2010; Sowa et al., 2004). Involvement of frontline staff and an organizational culture which values evaluation can also positively impact the use of data in performance management (Bryan et al., 2021; Lee, 2020; Perez Jolles et al., 2017). Naturally, nonprofits face challenges in their evaluation processes as well. Performance measurement in the youth development sector is uniquely characterized by lack of accepted outcomes and indicators of success (Wells & Johnson, 2001).

Performance Measurement Models

In 2004, Sowa, Selden, and Sandfort addressed the lack of consensus on how to measure organizational effectiveness, specifically the “lack of distinction between levels and units of analysis” (Sowa et al., 2004, p. 712). They created a multidimensional, integrated model of nonprofit organizational effectiveness (MIMNOE). This model divides performance measurement into two primary dimensions, the quality of management and the amount of program effectiveness. These dimensions reflect the argument that an organization with a successful program, but unhappy employees and poor management, is not an effective organization (Herman & Renz, 2008; Sowa et al., 2004). These dimensions are further divided into the categories of capacity (the processes and structures) and outcomes. These categories must be assessed in an objective and continuous manner to give a full picture of organizational effectiveness (Sowa et al., 2004).

MIMNOE serves as an example of how models try to encompass the units of measurement that demonstrate effectiveness, while being flexible enough to be applicable across the sector. Through this model, the authors anticipated organizations could move beyond the idea of cause and effect and began to analyze the complicated interrelationships that determine an organization’s impact (Sowa et al., 2004). Subsequently, in their nine theses on performance measurement, Herman and Renz elaborated on these relationships by discussing interorganizational networks. They stress how organizations can be perceived as effective based on the cumulative work of their peers (Herman & Renz, 2008). These authors also stress Sowa's argument that organizations must use multidimensional measures and look beyond program evaluation. They do, however, argue that Sowa's MIMNOE is still a program-driven model, and must take managerial effectiveness and organizational networks into greater account (Herman & Renz, 2008).

Sowa's MINMOE and Herman and Renz's nine theses are just two of many performance measurement models described in the existing literature. For that reason, authors Chongmyoung Lee and Branda Nowell categorized the different performance measurement perspectives into a framework. They define seven dimensions of performance measurement: inputs; organizational capacity; outputs; outcomes: behavioral and environmental changes; outcomes: client/customer satisfaction; public value accomplishment; and network/institutional legitimacy (Lee & Nowell, 2015). MINMOE, for example, is a model that emphasizes the organization capacity element. Organizational capacity models stress "the human and structural features" that enable organizations to run successful programs (Lee & Nowell, 2015, p. 302). Lee and Nowell's scholarship can be used to recommend a principle of performance measurement to an organization based on their "(1) funding type, (2) task programmability and observability, and (3) environmental turbulence" (Lee & Nowell, 2015, p. 309). Most importantly, Lee and Nowell highlight how modeling and recommending performance measurement systems is only half the battle. "Performance measurement is not an end within itself;" it is useless if organizations cannot link them to performance *management* – i.e., use these measures to effectively change their programs (LeRoux & Wright, 2010, p. 572; Bryan et al., 2021; Lee & Nowell, 2015). Finding the connection between doing performance measurement and using evaluation results will be the key to improving organizations and the lives of people they serve (Bryan et al., 2021; Lee, 2020; Sowa et al., 2004).

Performance Measurement in Use

Due to accountability requirements, organizations have been found to view performance measurement as a chore (Lee, 2020; LeRoux & Wright, 2010; Sawhill & Williamson, 2001). It can be seen as a waste of time and resources, only to collect “a lot of data, but very little information” (Sawhill & Williamson, 2001, p. 371). For other nonprofits, the collection of data is seen as an obligation, sometimes contractually, to donors or other stakeholders. Because of this feeling, it’s possible that only a small number of organizations use the data they collect in their decision making (Lee & Nowell, 2015).

There is a growing body of research on the characteristics of nonprofits that connect their performance data and theory of change to improve their strategic thinking. There is empirical evidence showing nonprofits that use a large number and wide range of performance measures are also more likely to consider the results of this data in performance management – i.e, in their organizational decision-making (Lee, 2020; LeRoux & Wright, 2010). This finding relates to Sowa’s theory that using multidimensional performance measures is crucial. Based on a survey of several hundred American nonprofits, LeRoux and Wright find that there is a direct positive relationship between organizations with a wide range of performance measures and those that report they are effective in their strategic thinking. In fact, among a list of many organizational practices (such as board governance and funder relations), collecting a wide range of performance measures had the second largest effect on an organization’s strategic thinking (LeRoux & Wright, 2010). When organizations have a range of data to choose from, they are more likely to find some of these measures useful and will then utilize the results in their decision making (Lee, 2020; LeRoux & Wright, 2010). Additionally, a greater number of measures satisfies the varying interests of stakeholders, as foundations, the government, and clients all have different goals and priorities which indicate organizational success (Ross and Thakur, 2014). LeRoux and Wright also found that specific measures were more useful than others. Their survey indicated that cost efficiency measures, followed by outcome measures, made the biggest impact on managers' thinking. Client satisfaction and industry benchmarks were not statistically significant in terms of impacting an organization's decision making (LeRoux & Wright, 2010).

Beyond the specific data an organization collects, managerial practices can improve how a nonprofit organization uses performance measurement effectively. First, staff at all levels must understand why data is being collected. If frontline workers do not understand the data collection process, they will be unable to accurately report information on program indicators (Perez Jolles et al., 2017). If staff do not understand collection procedures, organizations will eventually stop utilizing program measures altogether. Without organization wide education on data collection, it will remind a contractual chore (Bryan et al., 2021). The best way to educate staff on performance measurement is to create a culture of learning in the nonprofit (Bryan et al., 2021; Lee, 2020; Perez Jolles et al., 2017). Perez Jolles conducted a survey of 460 child welfare organizations and found that a culture emphasizing growth and learning had one of the highest impacts on frontline staff's understanding of performance measures. Lee highlights several studies where organizations with a culture that rewards risk-taking, development, and innovation are more likely to use their performance measures in decision making. Managerial choices, such as educating all staff on data collection and creating a culture of growth and learning, can shift performance measurement from a task to a useful tool.

Focus on Youth Development Organizations

Youth development nonprofits are a substantial sub-sector of the largest category of public charities, human services organizations (Mckeever et. al, 2016). Youth development is defined as organizations which have a mission focused on improving children's social-emotional status, reducing the likelihood to engage in risky behaviors, and ensuring children have an easy transition into adulthood (Horne et al., 2021). There are over 7,000 youth development nonprofits in the country, and, like all other human services organizations, they have growing public support (Mckeever et. al, 2016, p. 170 & 180). These organizations, human services in general and youth development nonprofits specifically, face their own unique challenges regarding program evaluation.

In 2011, authors Lynch-Cerullo and Cooney authored a general review of performance measurement in the human service sector, primarily discussing a dilemma practitioners at these organizations face. Human service organizations rely heavily on grant funding. The commercialization of the nonprofit sector over the past few decades has forced organization to comply with data requirements in order to compete with other nonprofits for grants (Lynch-Cerullo & Cooney, 2011). Funders require nonprofits prove their "impact," but for human service organizations, measuring impact is difficult (Lynch-Cerullo & Cooney, 2011).

Lynch-Cerullo and Cooney provide four points to explain why measuring impact is so challenging. First, they highlight how human service organizations work in tandem with other nonprofits and welfare services. Additionally, the economy and political opinions can greatly affect people's lives. When a client sees an improvement in outcomes, it is almost impossible to attribute that to one program (Lynch-Cerullo and Cooney, 2011). Of course, this holds true for youth development organizations. Children's well-being is particularly impacted by their education and family stability, factors far outside of nonprofits' control (Wells & Johnson, 2001). Secondly, because human service organizations are funded by multiple sources with different missions and expectations, they are required to prove impact to meet these various goals (Lynch-Cerullo and Cooney, 2011). Third, "social change can often occur very slowly" and after only 2-3 years of measurement a program may "look like a failure" but in twenty years data could show it had positive outcomes (Lynch-Cerullo and Cooney, 2011, p. 377). For example, youth development organizations may have goals such as educational and job attainment, which are not measurable for years into the future.

Finally, the greatest challenge youth development organizations face is developing measures that show they succeeded at the abstract concept of improving a human life (Lynch-Cerullo and Cooney, 2011; Mosley, 2018; Wells & Johnson, 2001). For youth development specifically, Wells and Johnson characterize these measures as trying to quantify the "elusive goals of child safety" and "supporting child well-being" (Wells & Johnson, 2001, p. 171). This ambiguity has attributed to a lack of accepted performance measures, which ultimately has limited the growth and development of this sector (OJJDP, 2014). A study of youth serving organizations in DC asked them to rate the quality of their evaluation systems. On average, the organizations said they struggled with identifying and focusing on youth specific outcomes (Ross & Thakur, 2014). Mosley highlights how the vagueness of measuring impact on youth is "part of its charm" to funders, because they can demand organizations "do better" without specifying what that means (Mosley, 2018, sec. 3). Yet youth development organizations remain in a position where they need funding. This need creates a necessity for developing sector-wide acceptable outcome measures.

Mosley argues organizations would have a better understanding of universal outcome measures if the scholarship and research on evaluation was unified across sectors. She highlights large discrepancies in the research on evaluation done by nonprofit journals and that done by social work and health journals (Mosley, 2018). Further, scholars argue performance measurement for youth development programs could be improved by including practitioners in the research on evaluations (Martinek, 2017; Larson, 2015). However, youth development evaluation research can be inaccessible for nonprofit managers. Nonprofit youth organizations, like the Girl Scouts of America and Big Brothers Big Sisters, have published research, but it has been "invisible" on scholarly sites (Bialeshki & Conn, 2011, p. 1). This research is more likely to be presented by organizations and journals studying educational research or social work, not nonprofit journals (Bialeshki & Conn, 2011). The unification of research across sectors, and the inclusion of practitioners in this research, should be a priority for scholars.

Addressing the Gap in Research

The findings of this thesis will answer the research question- what are the characteristics of effective performance measurement systems in youth development nonprofits? By interviewing practitioners from five organizations in Monroe County, it is addressing the need to include practitioners' perspectives and opinions in research. As this thesis exclusively focuses on youth development from a nonprofit lens, it will allow for the challenges and best practices of performance measurement systems in this specific subsector be fully explored.

Research on this topic is crucial, because there is value in assuring our youth development programs are working and we are providing children the easiest path to success. We know youth development programs can succeed. Evidence shows participation in youth development services decreases children's likelihood to be involved in drug use or gang violence and to experience mental health issues (OJJDP, 2014). Additionally, these programs improve academic achievements and “life skills such as job responsibility” (OJJDP, 2014, p. 4). When nonprofits effectively make this impact, children enter adulthood with the skills they need to thrive. They will ultimately need less investment and support from public and private social services as adults, which produces a financial and societal benefit for communities. However, nonprofits must have the tools to engage in effective performance measurement to show their programs work.

III. Methods.

Identifying the Population

I conducted semi structured interviews with five youth development nonprofits that serve Monroe County, Indiana to discuss the achievements and challenges of their performance measurement systems. First, I determined my population by defining youth development organizations and writing inclusion and exclusion criteria. I defined youth development nonprofits as those whose primary goal and mission consists of improving children’s educational outcomes, behavioral outcomes, citizenship, leadership, or advocacy skills. I included organizations that served any population of children from ages 0-18. Finally, the organizations had to have an office in Monroe County to offer some within-county comparisons.

I excluded organizations where the youth development component was a secondary mission of a different project (for example, an organization that has a mission to end homelessnes but also has a youth development program for the children experiencing homelessness they serve). I also did not include schools or foundations. While these organizations may have a primary mission that fit the youth development definition, I anticipated the structure of schools and foundations is such that their performance measurement systems would not be easily comparable to other nonprofits. Obviously, I also excluded organizations that were no longer running.

I used the Guidestar Pro database to determine the population of youth development organizations in Monroe County. I filtered organizations in Monroe County, IN and then added other filters individually. First, I filtered organization by their subject area, picking youth development. Next, I filtered them just by population served, using the children and youth category. There are 17 organizations in both these categories, and 96 organizations that are categorized as at least one of these. I looked at the 96 organizations which satisfied at least one of these, examining their mission or website if necessary. I filtered the organizations that were repeated more than once on Guidestar, not actually in Monroe County, did not have any online contact information, or did not fit my inclusion and exclusion criteria. Ultimately, 15 organizations fit my criteria. I then used these organizations' websites to find an executive staff member to interview. If possible, I contacted a staff member whose job title involved performance measurement, evaluation, or program improvement.

I contacted organizations the third week of December via email. Of the 15 organizations I emailed, eight responded. Three of those eight did not respond to my second email, prompting me to send another follow up. Ultimately, five organizations agreed to do interviews. These were the Boys and Girls Club of Bloomington, Big Brothers Big Sisters of South Central Indiana, Boy Scouts Hoosier Trails Council, Girl Scouts of Central Indiana, and the Villages of Indiana. These interviews were conducted between January and March 2022, over Zoom, with the exception of one that was in person. These conversations were semi structured in the sense that I followed a script, with questions addressing each of my hypotheses, and I aimed to collect the same information from each organization. However, I also asked follow-up questions or would skip questions if they were answered in a different part of the conversation. These conversations were all recorded and later transcribed.

Forming Hypotheses

After completing a literature review, several common themes of effective, and ineffective, performance measurement systems were evident. I divided these themes into “who, what, when, why, and how” concepts of performance measurement. This provided a logical structure of different characteristics of performance measurement systems I could focus on. From these categories, I came up with 22 hypotheses. For example, as a “who” hypothesis, I wrote, “if an organization has trained staff whose primary job description deals with performance measurement, people at different levels are more likely to have an understanding of performance measurement and its importance.” With the help of my advisor, I narrowed these down to the 12 strongest hypotheses; these were the ones supported which were most easily measurable and supported by my literature review. Then I did a “gut check” which ones were truly important to my research and of genuine interest to me within the youth development sector. Finally, I had eight hypotheses for which I developed interview questions. I categorized these nine into those that addressed the range of performance measures, frontline staff, organizational investments and mandated data collection.

Why Monroe County

Monroe County was chosen as a convenience sample due to time and financial constraints. I also anticipated interviewing organizations in Monroe County would increase my response rate, as I have a few personal connections in the local youth development sector. Staying in my own community also creates greater credibility for myself as an Indiana

University student, as many of these organizations have partnerships with the university, or at least are highly familiar with it. Interviewing organizations solely in Monroe County also provides consistency in the organizations' experiences. Because they are in the same community, they may share similar values, which may be reflected in their performance measurement systems. They also serve similar populations. In some ways this is a limitation, but conversely it can act almost as a control variable.

IV. Findings

Category	*Hypothesis*	*# of orgs where support was found*	*No Support*	*Indeterminate*	*N*
Range of Measures	If organizations collect a range of performance measures, specifically data on both managerial and program outputs and outcomes, they also have more avenues to improve their programming.	5			5
Frontline Staff	In organizations with trained staff whose primary job description deals with performance measurement, people at different levels are more likely to have an understanding of performance measurement and its importance.	2 (1)*	1	1	5

Frontline Staff	Organizations where employees at all levels, specifically frontline staff, recognize the importance of data collection are also more likely to utilize performance measurement in their decision making.		5		5
Frontline Staff	Frontline staff have a better understanding of the importance of performance measurement in organizations where all the collected data is accessible to them.	(2)*	3	1	5
Investments	Organizations with an office or staff member dedicated to performance measurement will be more likely to use data in their decision making.	3 (2)*			5
Investments	Organizations that invest significant time into data collection also have a stronger motivation to reflect on it and use it			5	5
Mandates	Organizations that collect data for a government, national affiliate, or grant mandate may be more consistent in their collection practices.	3		2	5
Mandates	In cases of mandatory data collection, organizations may not follow through on examining and responding to the data as they see it as an obligation/chore.		1	4	5

**Parentheses indicate an interview supported the hypothesis but in the opposite direction of the relationship. For example, if a hypothesis states more investment relates to more frequent use of data, the interviews represented by the number in parentheses were those which had less investment and less frequently used data.*

Range of Measures: Hypothesis 1

If organizations collect a range of performance measures, specifically data on both managerial and program outputs and outcomes, they also have more avenues to improve their programming.

This hypothesis was supported by data from interviews with all five organizations. Every organization collected at least two of the following types of measures: program outputs, program outcomes, managerial outputs, and managerial outcomes. Each type of measure provided unique information, which subsequently gave the organization insight into how to improve their planning and programming. When an organization did not collect all four types of measures, they would omit certain performance management decisions in their discussion (e.g., increasing staff training in response to a certain managerial output measure) that were included by other respondents who measured a greater range of data. Specific examples of each category of data are further explained in the discussion section.

Frontline Staff: Hypothesis 2

In organizations with trained staff whose primary job description deals with performance measurement, people at different levels are more likely to have an understanding of performance measurement and its importance.

This hypothesis was supported by conversations with three organizations, it was not supported by the conversation with one organization, and the last interview had indeterminate results. No matter the amount of staff training, none of the organizations claimed that their frontline staff understood the importance of performance measurement. However, two of the organizations that did have a staff member dedicated to performance measurement were the only ones whose board had an understanding of data and its importance.

The Boys and Girls Club of Bloomington has two trained staff members in data collection and shared that their board "[loves] data!... It really matters to them, not only so they can tell our story when they're in the community, but they are a part of a lot of our strategic planning groups that we have… so they use data in those groups to help make decisions our next strategic plan… they use that in a lot of their decision making." A second organization also has staff who are trained in performance measurement, in addition to a large quality assurance leadership team. That organization shared a similar sentiment that their board values performance measurement and is "always using" national research and the data their organization collects.

Another organization supported this hypothesis in the opposite direction. They do not have staff trained in performance measurement at the local level and have less understanding of data across levels in the organization. They claimed that the board cares about data but has a lack of understanding about what components of the data are important. According to the respondent the board views numbers without context and do not "tend to look at the why behind those measurements and outcomes, because covid has been a factor and staff limitations."

The interview with a representative from The Girl Scouts of Central Indiana did not support this hypothesis because their organization indicated a negative relationship between the two variables. GSCI has trained staff, including the interview respondent. Her role is helping the "staff and board know what data [they] have access to and how to interpret that data from a more holistic perspective." She reported that while their board is moving towards understanding she thinks they do not "really get it yet. I think they want programs to be amazing and to achieve outcomes, but I don't know that they really understand the depth of that yet." The respondent from the fifth and final organization provided insufficient information on the board's understanding of data.

Frontline Staff: Hypothesis 3

Organizations where employees at all levels, specifically frontline staff, recognize the importance of data collection are also more likely to utilize performance measurement in their decision making.

No data was collected that was able to support this hypothesis, as none of the organizations met the condition of the independent variable. All five organizations reported that their frontline staff struggled to understand the importance of performance measurement and data collection. Yet, all the organizations reported many ways they conduct performance management, despite the lack of understanding from frontline staff. Most of these examples were top-down measures, such as adjusting strategic plans or directors changing programs in response to satisfaction surveys.

Hypothesis 4

Frontline staff have a better understanding of the importance of performance measurement in organizations where all the collected data is accessible to them.

None of the interviews provided information that supported this hypothesis. The respondents claimed that frontline staff had varying levels of access to data, but regardless of access, no one said their frontline staff have a sound understanding of performance measurement. The Boys and Girls Club of Bloomington had one of the highest levels of frontline staff accessibility, yet still claimed, "it's not really about access, everyone has access, even our frontline staff can access all of [the collected data], but we haven't done a great job of making sure people understand that they have access… I don't think it's in the forefront of everybody's mind, the frontline staff, but we attempt, in our staff meetings [to discuss the importance of various measures]."

Investments: Hypothesis 5

Organizations with an office or staff member dedicated to performance measurement are more likely to use data in their decision making.

This hypothesis was supported by all five interviews. Three of the organizations do have local staff whose job is centered on performance measurement and two organizations do not. The two organizations that do not have local staff dedicated to performance measurement were less likely to use collected data in their decision making. While they viewed performance measurement as important, they were more likely to focus on anecdotal experiences of the staff, youth, and families. One organization said, "families want to come and experience [the program] not necessarily talk about it, so getting them to participate [in data collection surveys] is a challenge." The other echoed a similar sentiment that board leadership gets too caught up in "how many kids are we serving, how many matches do we have, and how many matches are we closing. They don't tend to look at the ‘why’ behind those measurements and outcomes.” Again, these organizations highly value data but placed an equal or higher emphasis on the day-to-day experiences of everyone at the organization.

The three organizations which did have staff or offices dedicated to performance measurement showed a higher interest in using data collection in decision making. Specifically, they expressed themes of wanting to improve their performance measurement systems, by collecting higher quality data and doing it in a more streamlined way. One representative explained, “I hear a lot in the nonprofit world ‘ugh I have to collect data.’ If you want to do well at your job, you have to be evaluated. You can't just expect good stories to really reflect this is how we do a great program… You don't want to spin your wheels doing something that’s not effective.” It was important to these organizations, or at least the interview respondents, to have evidence programs were succeeding. If they did not have this evidence, the organizations change programs accordingly.

Investments: Hypothesis 6

Organizations that invest significant time into data collection also have a stronger motivation to reflect on it and use it.

This hypothesis has indeterminate results. It was difficult to reliably assess the amount of time an organization spends on data collection and performance measurement. Some respondents said people across the organization engage in data collection multiple times a day, others said once per month. These numbers did not accurately reflect the rest of the interview, in which most of the organization seemed to have similar procedures, and all seemed to collect data more than once a month.

<u>Mandates: Hypothesis 7</u>

Organizations that collect data for a government, national affiliate, or grant mandate may be more consistent in their collection practices.

Interviews with three organizations supported this hypothesis, as they have extensive mandated data requirements and reported being quite consistent in their collection practices. For example, one of these organizations has a grant that requires they have an external evaluator whose sole responsibility is collecting reliable data. This contributes to more consistent and objective collection. A second organization, The Villages, must report to the Department of Child Services due to the nature of their programs. Their respondent explained that they are “extremely accountable to our funding sources" and that she “counted 40 metrics we have to track for the Department of Child Services. 40 metrics, and then qualitative data in an interview format that we have to produce for them on an annual basis.” These mandates result in their collection practice being more routine and require that their data is logged in a specific method through their electronic health record.

The other two interviews provided indeterminate results. These two organizations had the least rigorous mandated data collection out of the five and seemed the least consistent in their collection practices. However, they still reported being routine in some ways. For example, one of these nonprofits did have a few mandated grants, for the Eli Lilly Foundation and United Way, but stated that right now “establishing a routine is probably still where we are with that…

everybody is asking me, will you write a survey for this group of people… the progress is that they're asking me to do that, whereas in the past they would have thrown something together and our same volunteer would have gotten five surveys from people asking similar questions." They are moving towards more organized collection practices but are still somewhat chaotic. The other organization has minimal mandates and no required data collection from their national affiliate organization, but their respondent still claimed their collection practices are relatively routine.

Mandates: Hypothesis 8

a. ***In cases of mandatory data collection, organizations may not follow through on examining and responding to the data as they see it as an obligation/chore.***
b. ***Conversely, if the mandate requires them to collect information they otherwise would not, they may be more likely to use it in organizational decision making .***

Feelings about mandated data collection varied by respondent, several of them expressing opinions that supported both the hypothesis and counter hypothesis. Several of the organizations currently receive United Way grants. For these grants, the United Way requires specific outcome reports. One respondent primarily supported the original hypothesis. His organization is just starting the process of writing entry and exit surveys for the United Ways of Jefferson and Morgan County. He stated that their organization is "lean and mean and every time you add something you have to take away from something else." Writing and conducting these entry surveys "takes a lot of resources on [their] part… a lot of people hours to get out there and administer surveys."

A second organization has a nationwide grant funded by the Department of Justice which requires a specific pre- and post-test to be given every year. Their representative said the required collection was simultaneously an annoyance and gave their organization important information. “At times we have been so annoyed with it. Like when it has to be collected, how it has to be collected, all these rules and boxes we have to fit ourselves in when we're just trying to keep our heads above water most times and get our jobs done. But then also we’ll be in a meeting and be like you know we’re kinda thankful for the surveys because without them we wouldn’t know this about a child.” The Villages had a similar opinion, both recognizing the importance of these mandates, but also acknowledging the requirements took up a lot of time and were specifically frustrating for frontline staff.

The Girl Scouts of Central Indiana takes a different approach to mandated data, aligning more with the counter hypothesis but not fully supporting it. Their representative stated, “Whatever data that we get we're gonna look at and try and use it to improve.” However, receiving a grant with required data collection is not the catalyst for the Girl Scouts when it comes to measuring specific outcomes. Rather, they pay significant attention to what a grant is requiring of them before they apply to “make sure it fits where we want to go. If it requires a lot of crazy external data collection it may not be feasible.” In this case, the grant is neither incentivizing them or disincentivizing the GSCI to look at data. Rather their interest in certain outcome collection incentivizes them to apply for specific grants.

V. Discussion and Analysis

Limitations

There are several limitations to this study. First, the findings are not generalizable to the county or beyond. With only five interviews, it is a narrow scope of the data collection practices of youth development nonprofits. Moreover, there was a lack of diversity in the structure of the organizations interviewed. Except for one, they were all local chapters or independent franchises of large national member organizations. The patterns discussed below are most likely applicable to other organizations with this structure. It is also impossible to determine causal relationships within the findings. Because this is a cross sectional study, it cannot be established that the independent variables preceded the dependent variables. Other influential variables cannot be eliminated because there were no control groups. Additionally, this study is affected by single source bias because only one staff member was interviewed from each organization. Obviously, each individual at an organization has a different perspective on data, performance measurement, and the organization as a whole. Without interviewing other people at the organization, there is no way to verify what the subjects said was accurate.

Monroe County is used as a convenience sample. Interviewing organizations from just one community cannot give an accurate picture of the national youth development sector. One of the most striking differences between Monroe County and the country on average is the racial makeup. Monroe County is overwhelmingly white (US Census Bureau, 2020). This lack of diversity in both the staff and youth served likely impacts what organizations view as important to measure and their organizational culture. All being from the same community, these five nonprofits may emphasize specific measures that are impacting the children of Monroe County that other communities would focus less on. With this said, it was decided that Monroe County was still a useful sample as its other demographic measures were not complete outliers in comparison to other US counties. The median income is less than that of the United States on average and there is a higher percentage of people in poverty. The community also serves both rural and urban populations (US Census Bureau, 2020). These characteristics make Monroe County similar to other communities that likely have a high level of nonprofit involvement.

Patterns in the Findings

A. Range of Measures

There is a significant amount of scholarship stressing the importance of using multiple measures to create a full picture of organizational effectiveness. I assessed how organizations use indicators for both managerial and program success. This hypothesis builds off MINMOE, specifically the argument that programs cannot succeed without a well-run organization (Sowa et al., 2004). As previously stated, all five organizations supported the hypothesis that when nonprofits collect a greater number of measures, they also have more avenues to improve their organization. Examples of each type of measure are given below with insight into how the nonprofit used that measure to improve programming, management or both.

I. *Program Outputs*

Program outputs may be the most intuitive type of measure for an organization to collect. All five organizations gave examples of specific output measures and how they use them to improve programs. It should not be assumed that program outputs are any less useful to measure because they are generally the most straightforward to collect. Throughout the interview one respondent continued to stress the importance of measuring an output, average daily attendance, saying:

> “I definitely want to use data points as a driving factor in our decision making. If we don't have kids attending, if they don't want to come, there's a problem right there, something we’re doing that needs to change. That doesn’t mean we just need free time all day, that's

> not actually what kids want. Using data to find out how to help kids achieve what they want to achieve, and then to help them get the tools to make the decisions to create academic success, healthy lifestyle, and leadership skills. Like that's our jobs as adults is to help them to become better adults than we were."

II. *Program Outcomes.*

Program outcomes, especially long-term outcomes, can be more difficult to measure. Even so, information from these interviews suggests that this type of data is crucial for determining if an organization is creating the desired impact. A powerful program outcome measure from one of the organizations was their alumni study. This study showed their program achieved outcomes such as the children being more likely to vote and having higher educational attainment and job prospects. If organizations can figure out what part of their programming is influencing these long-term outcomes, then they have an incredibly powerful tool for improving children's lives.

III. *Managerial Outputs.*

One of the organizations spoke extensively about a specific managerial output measure, the data they collect on staff trainings. They have an online program that notifies staff when they have a new training to complete. Additionally, supervisors are notified when staff are overdue on trainings. Knowing this data is crucial for the organization because their programs deal with highly sensitive issues. The organization can become incompliant with state requirements if staff fall behind on trainings. A unique way this managerial output works to improve the nonprofit is that staff must do trauma informed training. While this is important for working with the youth they serve, the respondent stressed how it helps things run smoother at the office, because many coworkers have experienced trauma through the work they do. Assuring all staff are up to date on training means they interact better not just with clients, but each other.

IV. *Managerial Outcomes*

The Boys and Girls Club of Bloomington has an online system called the YPQ, which is the youth program quality assessment. It is an external tool developed by an outside group that does not measure program outcomes, but rather measures the effectiveness of how their programs run. According to the interview respondent this tool is “big in understanding if we’re doing our job.”

Again, this was one of the more intuitive hypotheses. One would expect most nonprofit managers and boards to agree collecting a range of measures is important. However, that does not mean they do it. Staff may not understand the distinction between managerial and program outcomes. They also may not have the capacity to collect all these measures. For example, the Boys and Girls Clubs of Bloomington collect an impressive range, but they are lucky to have systems like the YPQ to capture the more complex managerial outcome measures. Like much of performance measurement systems, it goes beyond knowing just knowing best practices, organizations must have the resources, financial and human, to execute them.

B. Frontline Staff

All five organizations were prioritizing involving their frontline staff in data collection, but none had succeeded. There were minimal claims that frontline staff and volunteers understood or cared about performance measurement, and to everyone I interviewed, that was a problem. They described getting frontline staff involved in different ways, one saying his role was to be a “cheerleader” to educate volunteers and get them to care. Another described how they are currently doing “pulse checks” with the kids, where they ask how they are feeling, then share that data with the staff. Despite efforts, challenges have remained for all the organizations. One respondent said:

> "I think times have changed so much, there used to be a time in the beginning when we didn't have so much oversight and so many demands around compliance. I was able to go out to every office yearly and share outcome measurements with the frontline staff, it was wonderful, they loved it. We would have conversations. Times have changed. I have many more jobs besides what I am describing to you, this is like one piece of what I do. There are so many more demands within the frontline staff and myself, we just can't do that anymore. I mean we do share information, we share outcome information with their supervisors, like the regional directors and clinical directors, and they take it to staff."

Though it varied within each organization, all five were using data in their decision making. They were collecting lots of measures and sending those measures to their national affiliate organizations. On the national level, these nonprofits have immense capacity. Big Brothers Big Sisters, for instance, is well recognized for experimental trials they have done on community-based mentoring. The Girl Scouts has an entire research institute where they conduct alumni studies. The data collected at these local chapters is contributing to compelling evidence on the outcomes these nonprofits produce. Yet, this information is not being relayed in a top-down approach to the frontline staff. They do not seem to realize why it is important to collect the data, and how powerful of a tool these measures can be.

The literature stresses the importance of frontline staff involvement in data collection and analysis. All five organizations are spending time getting frontline staff up to speed, with limited success. Yet, without frontline staff involvement, they continue to actively use data in their decision making. This suggests it may not be necessary to have frontline staff with a full understanding of data as long as they understand how to run programs. These organizations are “lean and mean,” and spending time educating frontline staff takes away from other duties. It is possible that attempting to involve staff and volunteers may not be worth sacrificing time which could be spent improving the organization in other ways.

C. Investments

Ultimately it was difficult to measure an organization’s time and financial investments into their performance measurement system. Research on the impact of each type of investment would likely be better done through an expansive, quantitative study. These interviews did reveal some interesting patterns on the effect emotional investment and job structure has on performance measurement systems.

The data indicated a correlation between having a staff member or team dedicated to performance measurement and the use of it in organizational decision making. As previously mentioned, this may be a result of single source bias. The three organizations where the respondent’s job title dealt with data and performance measurement appeared to use these systems more in organizational decision making. However, it is possible that these results were reflecting the interviewees’ personal opinions on data and how they use it in their individual decision making. They may have a biased opinion on the importance of performance measurement that general staff, like the ones interviewed at the other two organizations, do not share.

With that said, there may be some truth to the fact that these organizations do use performance measurement more in their strategic thinking. After all, the point of having an employee who is dedicated to data analysis is that the organization improves at evaluating itself. Through my interviews it was clear that organizational motivation to use data comes from a top-down approach. As one respondent said, “I think [our organization] always, from the top down, had that culture of valuing data and outcome information.” This respondent gave a lot of credit to the Vice President of Programs for creating that culture. While the results of my interview may reflect personal bias, they may also reflect the success these individuals have had in instituting a culture of performance measurement in their organization. Having an individual dedicated to that task likely makes a large impact on the effectiveness of an organization’s performance measurement system.

D. Mandates

All organizations collected some mandated data. Of course, collecting data for philanthropic or governmental grants is a necessary part of managing a nonprofit. Few would argue accountability in the nonprofit sector is a negative. It is valuable to see evidence that the money invested in a program is making an impact. Even though some respondents expressed annoyance at the additional hoops to jump through, many still explicitly expressed support for this accountability. One said, “all money comes with strings, and rightfully so. All money needs oversight. I think we need to make sure data is visible, and we are being held accountable for what we're saying we're going to do with money.” In addition to creating accountability, mandated collection also produced the positive externality of making organizations more consistent in their collection practices. Specifically, the nonprofits who received government funding described being more routine in their collection.

A negative consequence of mandated data is that three of the organizations complained about the extra work needed to do. They clearly viewed it as a distraction from their mission. This was with the exception of the Girl Scouts of Central Indiana. They had a unique perspective where they only applied to funding with data collection requirements that were a close fit to the outcomes they seek to collect. For them, the mandated data collection was mission fulfillment.

They were additionally motivated to apply for funding based on the fit of the data, rather than being motivated to collect data because they were solely interested in the funding. Of course, the Girl Scouts benefit from being a large national organization. A smaller nonprofit may not have the option of seeking only funding that aligns perfectly with their mission. In that case it may be worth jumping through the data collection hoops for the grant. With that said, using a mission fulfillment mindset like GSCI could help nonprofits eliminate feelings of annoyance with mandated collection, and ultimately cause them to use that data more in their decision making.

VI. Conclusions

The five organizations interviewed in this study all demonstrated characteristics of having effective performance measurement systems. They all have staff who value data and want a way to demonstrate to the public, their supervisors, and their frontline staff that the work they are doing is impactful. The organizations benefited from having a wide range of measures and having a staff member dedicated to performance measurement. They struggled engaging frontline staff, but in their attempts to help them understand data, they showed the benefits of a top-down culture emphasizing evaluation. More research on the need for staff understanding of performance measurement could help organizations avoid wasting resources trying to educate frontline staff. Mandated data collection was a reality for all the organizations, but only the nonprofit that viewed this collection as mission fulfillment did not see it as a chore. Creating structures that help outside funding requirements and nonprofits missions align could also solve this problem.

Going forward, scholars in the evaluation field, nonprofit management field, and youth development field should work to unify their research into more accessible spheres. Measuring outcomes for children can be complicated. Yet, creating programs that support the youth in our communities is so important. Children are a vulnerable group, and when they receive data driven programming that is proven to create positive outcomes, that can make a world of difference. This difference not only manifests in children’s lives, but in society as a whole. Youth development nonprofits advocate for kids who will one day be adults, and if they have been a part of meaningful programming, they will likely grow up to be better citizens.

Works Cited

Bialeschki, M. D., & Conn, M. (2011). Welcome to Our World: Bridging Youth Development Research in Nonprofit and Academic Communities. Journal of Research on Adolescence, 21(1), 300–306.

Bryan, T. K., Robichau, R. W., & L'Esperance, G. E. (2021). Conducting and utilizing evaluation for multiple accountabilities: A study of nonprofit evaluation capacities. Nonprofit Management and Leadership, 31(3), 547–569. https://doi.org/10.1002/nml.21437

Carnochan, S., et al. (2014). Performance Measurement Challenges in Nonprofit Human Service Organizations. Nonprofit and Voluntary Sector Quarterly, 43(6) 1014–1032. Retrieved from https://journals.sagepub.com/doi/pdf/10.1177/0899764013508009

Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2004). Positive Youth Development in the United States: Research Findings on Evaluations of Positive Youth Development Programs. The ANNALS of the American Academy of Political and Social Science, 591(1), 98–124. https://doi.org/10.1177/0002716203260102

de Lancer Julnes, P. & Holzer, M. (2001). Promoting the Utilization of Performance Measures in Public Organizations: An Empirical Study ofFactors Affecting Adoption and Implementation. Public Administration Review. Retrieved from https://onlinelibrary.wiley.com/doi/epdf/10.1111/0033-3352.00140

Herman, R. D., & Renz, D. O. (2008). Advancing nonprofit organizational effectiveness research and theory: Nine theses. Nonprofit Management and Leadership, 18(4), 399–415. https://doi.org/10.1002/nml.195

Horne, C. et al. (2021). Conceptualizing and Measuring the Promotion of Nonprofit Organizations' Evidence Use by U.S. Social Service Funding Programs. Journal of Public and Nonprofit Affairs. 7(2). Retrieved from https://www.jpna.org/index.php/jpna/article/view/482

Larson, R. (2015). Understanding Youth Development From the Practitioner's Point of View: A Call for Research on Effective Practice, Applied Developmental Science, 19:2, 74-86, Retrieved from Larson, R. (2015). Understanding Youth Development From the Practitioner's Point of View: A Call for Research on Effective Practice, Applied Developmental Science, 19:2, 74-86, DOI: 10.1080/10888691.2014.972558

Lee, C. (2020). Understanding the Diverse Purposes of Performance Information use in Nonprofits: An Empirical Study of Factors Influencing the Use of Performance Measures. Public Performance & Management Review, 43(1), 81–108. https://doi.org/10.1080/15309576.2019.1596136

Lee, C., & Nowell, B. (2015). A Framework for Assessing the Performance of Nonprofit Organizations. American Journal of Evaluation, 36(3), 299–319. https://doi.org/10.1177/1098214014545828

LeRoux, K., & Wright, N. S. (2010). Does Performance Measurement Improve Strategic Decision Making? Findings From a National Survey of Nonprofit Social Service Agencies. Nonprofit and Voluntary Sector Quarterly, 39(4), 571–587. https://doi.org/10.1177/0899764009359942

Lynch-Cerullo, K. & Cooney, K. (2011). Moving from Outputs to Outcomes: A Review of the Evolution of Performance Measurement in the Human Service Nonprofit Sector. Adminstrationin Social Work. Retrieved from file:///C:/Users/rypaquet/Downloads/pee%20pee%20poo%20poo.pdf

Martinek, T. (2017). Enhancing youth development programs through logic model assessment. Revista Internacional de Ciencias del Deporte, (42), 302-316. Retrieved from file:///C:/Users/ryann/Downloads/1180-6518-6-PB.pdf

Mckeever, B. et al. (2016). The Nonprofit Almanac. Rowman & Littlefield. 170-180.

Poister, T., et al. (2014). Managing and Measuring Performance in Public and Nonprofit Organizations. Jossey-Bass.

Office of Juvenile Justice and Delinquency Prevention. (2014). Positive Youth Development. A Product Model Programs Guide. Retrieved from https://permanent.fdlp.gov/gpo122628/PositiveYouthDevelopment.pdf

Perez Jolles, M., Collins-Camargo, C., McBeath, B., Bunger, A. C., & Chuang, E. (2017). Managerial Strategies
to Influence Frontline Worker Understanding of Performance Measures in Nonprofit Child Welfare Agencies. Nonprofit and Voluntary Sector Quarterly, 46(6), 1166–1188. https://doi.org/10.1177/0899764017728366

Ross, M & Thakur, M. (2014). Improving Youth Programs and Outcomes in Washington, D.C. Brookings Institute. Retrieved from https://www.brookings.edu/wp-content/uploads/2016/06/BMPP_PEPNet_Dec5.pdf

Sawhill, J. C., & Williamson, D. (2001). Mission Impossible? Nonprofit Management & Leadership, 11(3), 371. https://doi.org/10.1002/nml.11309

Sowa, J. E., Selden, S. C., & Sandfort, J. R. (2004). No Longer Unmeasurable? A Multidimensional Integrated Model of Nonprofit Organizational Effectiveness. Nonprofit and Voluntary Sector Quarterly, 33(4), 711–728. https://doi.org/10.1177/0899764004269146

United States Census Bureau. (2020) Quick Facts Monroe County. Retrieved from https://www.census.gov/quickfacts/fact/dashboard/monroecountyindiana/INC110220

Wells, S. & Johnson, M. (2001). Selecting Outcome Measures for Child Welfare Settings: Lessons for Use in Performance Management. Children and Youth Services Review. Retrieved from https://reader.elsevier.com/reader/sd/pii/S0190740901001323?token=9002B4DDD4FA5139F88AF825FB94AD71EB93DF1B8A81DC4FBB89E822BCF4060765E1E81DD2351FBA0DCD4FFAFC1987C8&originRegion=us-east-1&originCreation=20220503115414

Worth, M. (2021). Nonprofit Management: principles and practice. Sage Publishing.