11-29-2020

Equity in Program Evaluation: Equity as a Measure in Program Evaluation

Marco S. Thomas
University of Richmond

Equity in Program Evaluation:
Equity as a Measure in Program Evaluation

By

Marco S. Thomas

Masters Capstone

Submitted to:

Nonprofit Studies and Graduate Program
University of Richmond
School of Continuing and Professional Studies

November 29, 2020

Professor: Dr. Andrew Schoeneman

Abstract

Changes to equity and inclusion mean, not only including, but also valuing, and sharing power with, community members and stakeholders of various backgrounds. In addition to race and ethnicity, socioeconomic status, gender identity, physical and mental abilities, as well as where they intersect, should be represented throughout the entire evaluation process. Countless surveys make assumptions about communities without knowing the culture of the community. This study explores where equity does and does not exist, in the process of creating and conducting the evaluations that are used to measure the successful execution of nonprofit programs. The inclusion of program participants and stakeholders has been shown to increase program participation and stakeholder buy-in. The purpose of this study was to examine the evaluation process of nonprofit human service, organizations, in the central Virginia area to explore where equity does and does not exist, in the process of creating and conducting the evaluations. The methodology undertaken by the researcher was a mixed-method research design. The analysis of the data allowed us to understand if the organizations conducted any type of evaluations, who participated in the design process if funding was available specifically for the evaluation process, the organization’s capacity to perform evaluations, specifically equitable and inclusive evaluations, views on stakeholder engagement, and how the collected data was used.

1. INTRODUCTION

CONTEXT

With all that has taken place since the death of George Floyd and the disparities brought to light by Covid-19, the push for a higher level of diversity, equity, and inclusion can be seen almost everywhere. From boardrooms to business advertisements, the world has borne witness to the presentation of minority representation spanning across public and business sectors. In an attempt to promote diversity and inclusion nonprofit and government agencies have brought in consulting firms to create or conduct training, and implement policies, to present a show of solidarity and support for marginalized, and vulnerable, populations but have yet to address internal processes and structures that produce outcomes and promote change.

One particular structured process is how organizations and agencies design and conduct evaluations and assessments. While there is an outward attempt to show that black lives matter within organizations, no internal structural changes to produce equitable outcomes, really take place. There appears to be no serious effort to welcome inclusion outside of the hiring of a few like-minded Black faces. In terms of program evaluations, equity and inclusion are more than an extended invitation to a few token Black employees, and partnering agencies, to check the box of diversity. Changes to equity and inclusion mean, not only including, but also valuing, and sharing power with, community members and stakeholders of various backgrounds. In addition to race and ethnicity, socioeconomic status, gender identity, physical and mental abilities, as well as where they intersect, should be represented throughout the entire evaluation process.

Stakeholder Inclusion and Assumptions

As a member of several intersecting communities, and an employee in the nonprofit sector, I have sat in a survey planning meeting where experts in their fields, collaborate and share ideas on what questions to ask, and the culturally relevant wording to use, to collect data. This has become the solution for addressing the lack of equity in evaluation. We spend time bouncing suggestions off of one another, even correcting and contradicting each other's views. And while this may seem like a noble effort to understand, and engage, a community, assumptions of public safety, brainstorming beautification projects, and addressing community needs, continue to fill survey forms that have no presence of the residents that these changes will affect. And in the off chance that 1 or 2 community members find themselves at the table, their voices are drowned out, disregard, or met with frustration, and offense, if their input is contrary or in opposition to the views of the collective. Again, their presence is just a formality to meet the standard of inclusion or to meet a grant requirement.

Countless surveys make assumptions about communities without knowing the culture of the community. For instance, when it comes to public safety, the expert opinion might gravitate to questions about having more police presence but the presence of police officers as a symbol of safety, may not be the experience of those community members. For example, if we ask 5 questions making police officers the symbols of a safer neighborhood, the survey will reflect 5 answers painting a picture of a neighborhood that welcomes more police. We then, turn around and publish a report expressing the community's desire for more police presence, and present statistics from our survey to show we have engaged the people, and these are the data to support the notion. While the data may be factual, it is not an accurate representation of that community's idea of safety. We end up with data and statistics, however, they are statistics representing our perceptions of what a community thinks when asked about public safety.

Our data would more than likely look different if, instead of assuming, we asked that population how they feel about police presence, instead of feeding them questions to see how they would like our assumptions to be present to their community. In, "Clarifying Concepts and Categories of Assumptions for Use in Evaluation", Nkwake, & Morrow, (2016) explain:

> Assumptions are generally understood as beliefs, expectations or considerations that are taken for granted about how the world works. They may be tacit (unarticulated) or explicit (articulated). They may or may not be valid. Assumptions affect every step of evaluations from assessing relevance and need to program implementation to finally the evaluation of program objective achievement. (p. 97)

They go on to explain that once beliefs are formed, they are upheld, validated, and excused with a collection of intellectual reasons, cogent arguments, and rational explanations. Subconsciously, these beliefs then present themselves as bias that may even be contradictory to our conscious beliefs. Even still, these biases go on to influence individuals' behavior and "can have consequences like social stereotyping, prejudice, and sometimes discrimination" (Nkwake, & Morrow, 2016, p. 98). Whether explicitly or implicitly, we all have our biases and assumptions about certain groups and areas, and without someone present to debunk those assumptions, they can show up, reenforcing themselves in the questions we ask, the way we ask the questions, and how we interpret the answers to those questions (Thomas, Madison, Rockcliffe, DeLaine, & Lowe, 2018).

An Equitable Process to Produce Equitable Outcome

Two days after Breonna Taylor, a 26-year-old African American woman was fatally shot in her apartment by white plainclothes of the Louisville Metro Police Department, Vice President

Joe Biden, and Senator Bernie Sanders of Vermont announced their commitment to selecting a woman as their running mate during the Democratic primary debate (Phillips, 2020; Viser, 2020). After winning the nomination, Biden indicated that he would make his selection based on shared political beliefs and past experiences. He announced that he would run a “vigorous vetting process” to choose a running mate who reflects the racial diversity of the Democratic Party; one “with whom he is similarly ‘simpatico’ on key issues — and for someone who could be entrusted with presidential authority” (Glueck, 2020).

Biden, and members of the Democratic party, had a detailed set of criteria for a running mate, criteria that exceeded biographical or geographical requirements and it was clear that race, ethnicity, and gender were at the top of the list. The second-highest office in the free world was also seeking a display of diversity and inclusion, potentially opening the door for racial equity and gender equality on a national level. This vigorous assessment, and detailed evaluation plan, would begin to shape, what appeared to be, the most diverse, equitable, and inclusive presidential cabinet in this country’s history. This would all mean nothing without an equally inclusive selection process.

If the assessment and evaluation process, itself, lacked diversity, equity, and inclusion, the views, agendas, and practices of the candidates, would reflect the views, beliefs, and biases of those responsible for the selection process. While the candidates would present an appearance diversity, their moral compass, belief system, worldviews, even the laws and policies they support, would be anything but diverse. Montenegro and Jankowski (2020) explained that “an assessment process that is not mindful of equity can risk becoming a tool that promotes inequities” no matter how good the intentions are (p.4).

Chauveron, Samtani, Groner, Urban, & Linver (2021) argued that “despite the importance of including diverse stakeholder interests, needs, and priorities, they are often excluded from the evaluation process, ultimately threatening the quality and opportunity for equity in evaluation” (p.222). They explain that standing with preferred approaches can “privilege certain stakeholder perspectives and silence others” (Chauveron, et al. 2021, p. 222). These practices can inadvertently uphold the “hierarchical power of the evaluator’s voice and may further disenfranchise communities or reinforce narratives of oppression” (Chauveron, et al. 2021, p. 222). Confidence in one cultural worldview blinds the field and its products from the use of culturally diverse questions, definitions of data, and interpretations of findings conceding the credibility and validity of the evaluations. These traditions founded, compensated, and supported by the discipline paved the way for preserving policies and practices that endanger the lives of communities of citizens who do not adhere to a Eurocentric worldview (Caldwell, & Bledsoe, 2019).

From a pragmatic point of view, the omittance of diverse stakeholders can open the door for insensitivities, factual error, and inadequate data for program understanding and enhancement (Chauveron, et al. 2021). Evaluation designs void of stakeholder input can produce data collection instruments that may lack meaningful evaluation questions, implement the least effective methods, or can fail to identify the right participants. Lastly, without connecting stakeholder input to program execution, it can’t be certain its programs are meeting goals or addressing the needs of the population utilizing their services (Chauveron, et al. 2021).

THE PURPOSE OF THIS STUDY

Frye & Hemmer (2012) suggests that “At the most fundamental level, evaluation involves making a value judgment about information that one has available” (p.3). They go on to define evaluation as the "systematic collection and analysis of information related to the design, implementation, and outcomes of a program, for the purpose of monitoring and improving the quality and effectiveness of the program" (Frye & Hemmer, 2012, p.3). In the context of this study, evaluation as a reference to program evaluation and assessments will be referred to as the manner in which organizations measure the value, usefulness, and impact of their programs. This study explores where equity does and does not exist, in the process of creating and conducting the evaluations that are used to measure the successful execution of nonprofit programs.

Defining Equity

United Nations International Children's Emergency Fund (UNICEF) evaluation office interprets equity to mean "that all children have an opportunity to survive, develop, and reach their full potential, without discrimination, bias or favoritism" (Bamberger and Segone, 2011, p. 3). Not to be confused with equality, which removes differences and gives everyone equal distribution of resources, equity-based practices seek to remove the barriers, discriminatory practices, and injustices that favor one group over another (Bamberger and Segone, 2011; Wiggins and Sileo, 2020). The Annie E. Casey Foundation defines equity as, "the state, quality or ideal of being just, impartial and fair" and express that it "needs to be thought of as a structural and systemic concept" while The American Evaluation Association explains it as "the condition of fair and just opportunities for all people to participate and thrive in society regardless of individual or group identity or difference" (Wiggins and Sileo, 2020 p.1). This concept can be seen as an outcome, ensuring that each individual or group has what they require for survival, or a process to which those suffering inequities are meaningfully engaged and included in the process of development, collection, analysis, and dissemination of the evidence that involves, affects, and impact their lives both directly and indirectly (Wiggins and Sileo, 2020).

Oppression and Effects on Equity

In Pedagogy of the Oppressed, Paulo Freire (1972), explains that the oppressed must participate in and develop the pedagogy, or instrument, of their liberation. He expresses that it is impossible for the pedagogy to be truly liberating while remaining distant from those it is supposed to help, by treating them as less than, and unfortunate. He goes on to argue that this method of liberation cannot be developed and practiced by the oppressor; that to be defended and even implemented by those who are responsible for the oppressed condition is in itself, a contradiction. Such is true with assessments and evaluations.

No matter how open we are to understanding the experience of others, we still are prone to our assumption if we have no personal experience with a given phenomenon. This is why different perspectives are key when considering program development and evaluations. Policies, interventions, and programs resulting from data collected through research and evaluations can come out of the best intentions. Still, these policies, interventions, and programs run the risks of upholding, and adding to, inequities that may already exist within their structures when there is no stakeholder input in the development phase (Brown, & Di Lallo, 2020; Chauveron, et al, 2021; Wiggins and Sileo, 2020). In addition, cultural competence in evaluation is understood to be an ethical imperative because safety and security do not mean the same thing to everybody (Brown, & Di Lallo, 2020; Mapitsa, & Ngwato, 2020). In other words, those engaging in these practices must have an everyday understanding of the lives lived by those they seek to assist or assess. It is essential for ensuring the validity of evaluation findings and reducing, and hopefully eliminating, biases and assumptions embedded in the evaluation and its process. The context, inclusive of key systems and power structures, of an intervention has to be considered by researchers, even within seemingly similar conditions because solutions are not always one-size-fits-all (Garcia, & Stevahn, 2020; Stern, Guckenburg, Persson, Petrosino, & Poirier, 2019).

Significance of the Study

Studies have shown that an inclusive, equitable evaluation process does not just consider the perspectives of staff enacting the program, it also values the cares and concerns of program participants and other members of the community who share an interest in, and are impacted by, the program or policy influenced by these evaluations. The inclusion of program participants and stakeholders has been shown to increase program participation and stakeholder buy-in. It also increases understanding and knowledge of the organization and its stakeholders (Chauveron, et al, 2021). This is important in fostering a better understanding of the communities and their members, thus making engagement more organic and welcomed. This aids in debunking stereotypes and dismantling biases. This study is significant in that it seeks to answer the following questions:

At what point are equity and inclusion considered when developing a program evaluation?

How often, if at all, are stakeholders included during the developmental phase of program evaluations?

When developing a program’s evaluation process, who is involved?

How is cultural, and self, awareness a consideration when developing an evaluation plan?

What barriers exist in a stakeholder-inclusive evaluation process?

CHAPTER 2: LITERATURE REVIEW

Over the years research has begun to highlight the work of evaluators who engage in traditional evaluation processes (Public Policy Associates, Inc., 2015; Thomas, Madison, Rockcliffe, DeLaine, & Lowe, 2018). These forms of evaluations have had limited success in assisting organizations and communities make advancements toward racial equality (Public Policy Associates, Inc., 2015). These evaluators, often European Americans, are often ignorant to the everyday lives experienced by their priority populations (often minorities) that social investments serve (Public Policy Associates, Inc., 2015; Thomas, Madison, Rockcliffe, DeLaine, & Lowe, 2018). Many times, these evaluators fail to engage the consumers of services or bring personal biases and misconceptions along if and when they do seek a more inclusive process (Public Policy Associates, Inc., 2015; Thomas, Madison, Rockcliffe, DeLaine, & Lowe, 2018). The evaluations that are designed in the traditional sense, may not examine, or assess diversity, inclusion, and equity; an omission that makes disparities hard to recognize and address (Dean-Coffey, 2018; Public Policy Associates, Inc., 2015).

This chapter reviews existing research on evaluation practices used by Nonprofit organizations. The chapter also examines the use of equitable evaluation processes and the data produced by those evaluation practices. Synthesizing research collected from several peer-reviewed journals, and scholarly databases, this chapter will attempt to discuss alternative methods of conducting evaluations, a different practice that may assist with or hinder inclusivity [e.g., Evaluative Thinking, Equity-based thinking/Equity-based design, and Systems thinking and complexity science (STCS)], specific population to consider when considering equity and inclusion (people with disabilities, LGBTQ+, senior citizen communities, and youth/young adults, etc.), as well as barriers to more equitable evaluations process and reasons why some program may not incorporate stakeholder's voice (finance, staff size, technology, and time constraints).

Practice that Effect Inclusivity in Evaluation

In today's world of evaluation and assessments, the traditional way of doing things, the linear and mechanistic methods using a rational/synoptic process with logic models, predicting outputs and outcomes, are being challenged (Bustamante, Vidueira, & Baker, 2021; Netting & O'Connor, 2013). Communities are constantly changing, and programs have to adjust to those changes, making it difficult to determine and predict outcomes as well as what success may look like. What was originally conceptualized in the program design, and what was originally tracked, looks different later because targeted outcomes and indicators keep changing (Nkwake, & Morrow, 2016).

Advocates are voicing the need for change in the way organizations emphasize the importance of output- “the products or services directly produced through program activities”- and not outcomes – “substantial changes, benefits, or values resulting from the programs and services” (Lee, & Clerkin, 2017, p.111). Existing research not only looks at how Non-profit organizations (NPOs) define and measure Organizational Effectiveness (OE), but it speaks to the need for diverse ways to evaluate programs and more community, and stakeholders, inclusion in the process and planning phase of evaluations, not just in data collection and dissemination (Jones, Wanzer, Skousen, Knight, & Sabarre, 2020; Liket & Maas, 2015; O'Sullivan, 2012). Still, these evaluation processes are often requested, and are important, for adhering to funding and stakeholder requirements, informing organizational learning, and strategic planning (Despard, 2016; Jones, et al, 2020).

The ability to conduct meaningful evaluations could mean the difference between implementing empirically supported interventions and continuing the standard practice in an organization, regardless of how ineffective it may be (Despard, 2016). The use of evaluations and assessments can support the need to make adjustments and changes to daily operations, and an equitable approach to evaluation confronts the structures and practices that have historically, and sometimes intentionally, omitted the voices, knowledge, expertise, capacity, and experiences of community stakeholders, most importantly, Black, and brown people and people living in poverty (Stern, Guckenburg, Persson, Petrosino, & Poirier, 2019). Various research suggests a need for increased capacity for nonprofit evaluation and a change in our attitude and mindset when planning evaluation processes.

Evaluative Thinking

For those who engage in evaluation capacity building (ECB), evaluative thinking (ET) has become more and more familiar (Patton, 2018). Patton (2019) argues, “The most fundamental lesson evaluators have learned is the necessity of matching the evaluation approach to the nature of the situation which includes being clear about who the evaluation is for, what is being evaluated, and how findings will be used” (p.297). Incorporating ET starts with the ability to think critically when considering, planning, and conducting evaluations, and presents new pathways to outcomes that funders and nonprofits are beginning to regard as valuable (Patton, 2018). Not to be confused with the process or instrumentations of evaluation, ET, according to Buckley, Archibald, Hargraves, & Trochim (2015), is “critical thinking applied in the context of evaluation, motivated by an attitude of inquisitiveness and a belief in the value of the evidence, that involves identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective-taking, and informing decisions in preparation for action." (p.378). This mindset synthesizes past events to assist in making future predictions and recommendations (Patton, 2019).

Systems Thinking

Systems change evaluations directs the outcomes of initiative centered around major systems change initiatives (Patton, 2019). While most of the techniques and tools used for evaluations are centered on project thinking, a growing number of policymakers and funders desire a change in systems (Patton, 2019). Systems thinking offers a process and procedures that provide a more holistic evaluation while allowing purposes to be the main focus (Reynolds, 2014). Patton (2019) believes "The foresight offered by systems change evaluation is that program change is different from systems change—and requires different ways of thinking, engaging, evaluating, and, might we add, futuring" (p.300). The project mentality, with its logic models and SMART goals, are treated as closed systems in programs and projects, establishing boundaries and exercising control with predicted desired goals and outcomes (Patton, 2019). This is problematic for systems defined by validity, uncertainty, and the unpredictability of most social and world issues (Patton, 2019). A systems mentality, like real-world issues, requires innovations, adaptability, employing a more responsive, emergent approach (Patton, 2019, Netting & O'Connor, 2013).

There is an understanding that people, communities, organizations, and neighborhoods are open systems -exchange occurs among parts and the environment- constantly evolving, intermingling, and operating together as larger open systems (Frye & Hemmer, 2012). Yet, programs are designed with predictive evaluations processes (logic models, predetermined objectives and outputs, and SMART goals) as if they are addressing closed systems - nothing either enters or leaves- that do not experience change and are linear and static (Frye & Hemmer, 2012). Context cannot be disregarded when thinking about how to evaluate programs that serve communities and neighborhoods that are constantly evolving and experiencing change. Reynolds (2014) believed for those working on human rights and issues with equity, have to have a deeper understanding of the systems they are addressing, and that to make the necessary change that will ultimately change the system itself, they have to move beyond quick fixes and surface solutions. Reynolds (2014) argued in favor of employing a double-loop style of learning -thinking more deeply about one's own assumptions and beliefs- that makes the changes that either embed the solution in the changed system or prevent the problem from ever occurring.

Reynolds (2014), believed that allowing the dominant voices and interests to control the evaluation agenda, there is the risk of the solutions often results in the continuation of what was and is considered to be the right thing to do. Reynolds (2014) continues by acknowledging that, "the challenge here is not only to question the ethical dimension of what's right but to also appreciate the political dimension in acknowledging that what's considered 'right' depends on the context and relations of power circumscribing rightness" (p. 80). In other words, with people and communities in a constant state of evolution, it is key to have an evaluation process that considers relational context, and that may or may not be understood by the dominant voices. In with intentions to do what is right, without diverse input, organizations can run the risk of doing the wrong thing, the right way (Reynolds, 2014).

Evaluative Influence

Mason and Azzam (2019) introduce the concept of placing less attention on the use of evaluations and more on an evaluation's influence which they define as "the capacity or power of persons or things to produce effects on others by intangible or indirect means" (p. 249-250). They proposed that this concept "differs from various types of 'use' that are more direct in nature and include instrumental use (the direct use of evaluation feedback for program modification), conceptual use (changes in how stakeholders think about the program), and process use (where the learning that is generated from engaging in the evaluation process" (Mason & Azzam, 2019 p.250).

Bridging evaluation and a change in attitude mean a shift in a person's disposition towards a program and that programs underlying assumptions. It also gives room for other attitudes and the validity of other individual views (Mason & Azzam, 2019). Mason & Azzam (2019) argued in favor of the connection between evaluation and attitude change as a pathway to program improvements and societal betterment. They explained that a change in an individual's attitude towards a program or its underlying assumptions and the presence, and acknowledgment, of other attitudes and credible diverse opinions, is a starting point for movement along the pathway to social betterment and theoretical evaluation outcomes (Mason & Azzam, 2019).

However, like with most innovated ideas, there are always drawbacks. Because one dimension of evaluations influence happens on an individual level, it has the potential to push one person into the position of the influencer, attempting to bring change to the collective, and can sway individual work practices, and also, change the behavior of the collective (Mason & Azzam, 2019). Also, in contrast, social psychology research contradicts the notion that objective data can influence attitudes towards social issues, especially, in the case of deeply ingrained views and beliefs. Studies on attitude polarization argue that "those who hold strong attitudes assess data in a biased manner, focusing on information that supports their initial view and discounting findings they disagree with" (Mason & Azzam, 2019, p. 251). Mason & Azzam (2019) expresses that the key to combating confirmation bias and strong attitudes and exposing impartial opinions to contradicting sources is understanding if, when, and in what circumstances attitude polarization occurs. A more equitable process can present opposing views and diversify an existing process or introduce inclusion of, the previously mentioned, attitudes and credible diverse opinions, where an individualized process of evaluation exists (Montenegro & Jankowski, 2017).

Deficit Thinking

Lasater, Bengtson, & Albiladi. (2021) states that “Deficit thinking refers to the belief that educational outcome disparities are the result of students’ intellectual, dispositional, moral, or cultural deficiencies” (p.2). Characterized by placing blame on an individual’s assumed deficiencies, this thought process is often directed towards the most marginalized population (Lasater, Bengtson, & Albiladi, 2021). The assumption from deficit thinking gets substantiated when data is used to reinforce biases and stereotypes, and when program-level employees with personal biases get promoted to systems-level positions (Lasater, Bengtson, & Albiladi. 2021; Lichtenstein, 2018). This then has the potential to spill over into the designs of programs and how they are evaluated.

The rise of discussions about social justice being woven within the American Evaluation Association (AEA) is, in itself, the confession of the existence of these injustices, and permeates the profession (Caldwell, & Bledsoe, 2019). It is no secret that research and evaluation have been used to perpetuate stereotypes of intellectual deficiency in Blacks as a way of proving a difference in intelligence. This history and compliance of this behavior remains embedded in evaluation’s theory, methods, practice, and training, nationally and professionally, and is evident in calls for inclusivity and transformation (Caldwell, & Bledsoe, 2019). It is standard practice for social, and political, matter to be defined, constructed, and implemented by research professionals, who do not value or understand the context or life experiences of people of color.

Caldwell, & Bledsoe (2019) expresses concern that for decades, “authors have questioned the adequacy of ‘traditional’ frameworks to evaluate the culturally and socially different” (p.9).

As far back as1976, scholars have “introduced the idea that evaluation was, in fact, a political activity that perpetuated societal structures of power and authority” (Caldwell, & Bledsoe, 2019, p. 8). The practice of evaluation is not neutral when it comes to race or culture, given that its methods, practices, theory of knowledge, validity, and scope are dominated by a Eurocentric ideology which upholds a cultural premise (Caldwell, & Bledsoe, 2019). It has been argued that evaluation, being political, is based upon fields of study that were conceived, perfected, and readily practiced during times of colonialism, slavery, segregation/apartheid, and general mistreatment of people who did not subscribe to a colonial standard. Therefore, the recognizing, and calling out, of structural racism is mandatory to implement social justice, transformative, equitable, and culturally laden approaches to evaluation, even though these approaches are, still to this day, considered abnormal (Caldwell, & Bledsoe, 2019).

Forms of Evaluations

The Interactionalist perspective and symbolic interaction theory is “a theoretical perspective that derives social processes (such as conflict, cooperation, identity formation) from human interaction” (Lawler,2013 p.20). It is the understanding of how “individuals shape society and are shaped by society through meaning that arises in interactions” as well as the meanings placed on symbols (Ritzer, 2018, p. 57). Because none of us exist in social bubbles, our individual experiences and interactions can shape the way we interpret the world around us. These worldviews and personal judgments can knowingly, and unknowingly, show up as biases in our personal and professional lives, including how we structure and conduct program evaluations.

One way of making sure our personal beliefs and biases do not exacerbate historic disadvantages or inequity is to include diverse stakeholders throughout the evaluation process (Chauveron, Samtani, Groner, Urban, & Linver, 2021). Chauveron, et al. (2021) stated that “one critical element of high-quality evaluation is the inclusion of stakeholders at various points in the program evaluation process, including during evaluation planning, implementation, and utilization” (p. 221-222). Bringing in different perspectives from various representative stakeholders can benefit innumerable phases of evaluation, from program planning, evaluation design, and implementation, buy-in, credibility, and use of results while illuminating underutilized voices from areas and classes, that have been historically disadvantaged Chauveron, et al. (2021).

One study in particular, “Including Diverse Stakeholder Voices in Youth Character Program Evaluation”, in which Chauveron, et al. (2021) examined the inclusion of diverse stakeholders in evaluation practices in youth character programs. Three different aspects of diverse stakeholder inclusion were examined, stakeholder inclusion in the evaluation process, perspectives on program performance, and stakeholder input to the evaluation design. What the researchers found was that participants included a larger and more diverse stakeholder perspective in all three areas. In addition to behavior change, the study ultimately discovered “that with so many youth character programs not utilizing youth in their evaluation planning, their key constituency was inadvertently being demoted or even silenced, while funders were largely privileged” (Chauveron, et al. (2021, p.232). Also, what was discovered was stakeholder involvement also helped to improve the evaluation quality by minimizing the factual error and thoughtlessness that prevent the production of program knowledge and improvement. This section looks at several other methods of evaluation that help to elevate stakeholder voice in the evaluation process.

Culturally Responsive Assessments

In Culturally Responsive Assessments rests the notion that evaluators should be mindful of equity and be open-minded about the presence of inequities, as to not feed existing cycles that perpetuate inequity (Montenegro and Jankowski, 2020). Self-reflection helps those who lead and participate, in assessment from allowing biases and stereotypes to influence their perspectives and interpretation of findings. It helps them to not allow their culture to contaminate their views of other cultures as well (Public Policy Associates, Inc., 2015). “In order to minimize the impact that individual experiences, assumptions, and biases have on the evaluation, all members of the evaluation team should: Consider whether any facet of their background makes them ill-suited to properly interpret ideas, wants, and needs of the relevant population” (Public Policy Associates, Inc., 2015, p.8). Doing an honest assessment of one’s own experiences can highlight potential triggers that, for an evaluator, could have negative effects on their ability to be neutral or nonbiased toward a particular group or situation (Thomas, Madison, Rockcliffe, DeLaine, & Lowe, 2018).

Equity-Focused Evaluation

Bamberger and Segone’s (2011), description of Equity-focused evaluation introduces a different perspective. It poses the question, could we go about assessing and evaluating inequities, themselves, as the purpose of the intervention? Their approach, in contrast, is designed to be more responsive to systemic disparities that affect vulnerable populations. In the evaluation process, they discuss an evaluability assessment, “an exercise that helps to identify whether an intervention can be evaluated and whether an evaluation is justified, feasible and likely to provide useful information" (Bamberger and Segone, 2011, p.19). This is more of a summative evaluation to measure program effectiveness and feasibility as opposed to the evaluation design.

Transformative and Transformative Participatory Evaluation

In an interesting twist, Hay (2017) addresses the notion of diversity and inclusion being an intervention in itself. Believing that equity cannot be obtained through the implementation of interventions, transformative evaluation and transformative participatory evaluation take the stance of valuing equity while building communities. Hay (2017) believes that "Doing so necessitates bringing this lens to all that we do, including evaluation" (p. 80). Hay (2017) sees evaluation through two distinct lenses; to either leave inequities and its systems untouched or as a tool "to challenge the status quo and hold a mirror to it" (p.80). This research addresses the inclusion of the disenfranchised through participation in the process, to which there is consideration given in the interpretation of the impact. Be that as it may, this would seem to only be effective when the evaluators have a personal investment in the population suffering inequities. These do nothing to identify and eliminate the implicit biases that may find their way into the evaluation design. This may work for larger systemic inequities that are obvious but could show flaws in smaller more subtle biases that may come about out of pure ignorance or lack of experience with a particular community (Public Policy Associates, Inc., 2015; Szijarto, & Cousins, 2019).

Deconstructed Assessment

There are others who acknowledge this gray area (Montenegro and Jankowski, 2020). When discussing socially just assessment in higher education, it was argued that prejudices can dictate the types of tools being applied, what source of evidence is assigned more validity, and what conclusions are drawn from collected data and the solutions to improve outcomes (Montenegro and Jankowski, 2020). Models used when undertaking evaluations —foregoing intentionality—attempt to influence what decisions are being rendered and what questions are being presented (e.g., asset-based versus deficit-based perspectives toward different initiatives, student populations, sources of data). In favor of socially just assessments, the concept of Deconstructed Assessment was introduced in an effort to not only understand why students are succeeding but to get to the underpinning structures that produce such results from the start (Montenegro and Jankowski, 2020). It goes beyond the results to uncover the structure and practices producing the condition.

Specific Populations

Evaluations processes should always keep in mind, its target population. Evaluations should give considerations to language that is culturally appropriate for anyone, acknowledging differences, as well as similarities during the planning phases. Evaluators should put focus on developing assessment tools that are diversely appropriate and user-friendly for anyone who will be participating. Also, evaluators should be intentional in using findings and results to improve the conditions of the entire population, not just a select few (Montenegro and Jankowski, 2017). In addition to the obvious, consideration for disabilities (physical and mental), sexual orientation, and gender identity is equally important when planning evaluations because “all human beings make assumptions about how people’s bodies and minds work and most of these assumptions normalize the experiences of able-bodiedness and able-mindedness” (Kimball, Abbott & Childs, 2020 p.2). In other words, cultural responsiveness goes beyond racial and ethnic qualities when making the advancement towards equitable evaluation processes. Oppressive systems normalize stereotypes and ideologies that extend beyond class and racial divides. Montenegro and

Jankowski's (2017) notion of fairness, without considering physical accessibility to participation, is an example of good intentions not going far enough (Kimball, Abbott & Childs, 2020). We cannot always consider all possibilities, all the time, this is why diverse inclusion, from start to finish, is important. It is the notion that if you want to go fast, go alone, but if you want to go far, go together.

Related Challenges for Human Service Organizations

According to Despard, (2016) Non-profit Human Service Organizations face challenges when it comes to "understanding the distinction between outputs - data concerning services provided, including the numbers and characteristics of service recipients and hours or other units of service, and outcomes - changes in knowledge, skills, behavior, and circumstances for persons receiving human services" (p.). Often NPHSOs evaluate program process instead of program effectiveness and impact. In addition to needing to shift the way the organization is managed, the transition from outputs to outcomes will require the improvement of data collection and measurement techniques (Despard, 2016). Also, with more voice comes more perspectives on issues, and means of intervening, making a need to deeply explore problem contexts before selecting an evaluation design method (Foote, Ahuriri-Driscoll, Hepi, & Earl-Goulet, 2021).

The most common barrier nonprofits struggle with is inadequate funding to cover evaluation costs, being understaffed, and simply not having the time (Foote et el., 2021). "Compounding these challenges are high rates of staff turnover in human service organizations, especially in certain fields such as child welfare and intellectual and developmental disabilities" (Foote et el., 2021, p.6). Having to constantly replace staff presents challenges of its own, not just in the area of evaluation. Minimizing the deficits caused by having to perform new hires usually take precedence over the evaluation gaps, and because most NPHSOs do not have a dedicated evaluation team, new hires have to be brought up to speed on the process that may be in place (Foote et el., 2021). And lastly, technology has its own place in small budget organizations, but having the best, updated software and electronic to conduct, analyze, and maintain data, has its deficits as well causing NPHSOs to see evaluations as a luxury they cannot afford to undertake (Foote et el., 2021).

Whether Culturally Responsive, Participant-oriented, Transformative, or Emancipatory evaluations, this research understands that there are limitations, barriers, and possible consequences to making evaluation processes inclusive of all stakeholders. With more voices come different points of view, most of whom may not want to get involved (Acero, 2019; Chauveron, Samtani, Groner, Urban, & Linver, 2021). Any change generates opposition and political pushback. People have agendas, and motives that aren't always suited for inclusion and growth, and stakeholders are not always as trusting of research, especially surveys, because of this (Hall, Freeman, & Colomer, 2020; Chauveron, Samtani, Groner, Urban, & Linver, 2021; Odera, 2021). Most importantly, budgets and time constraints will limit almost any evaluation process whether stakeholder inclusive or otherwise (Bopp, Harmon, & Voida, 2017; Mason, & Azzam, 2019). Most community involved, grassroots organizations have limited staff, expertise, and funding, which may mean new hires and possible training. All of these limitations must be considered and accounted for if we are to make program evaluations that will produce the programs that represent the will and needs of those who depend on them.

Conclusion

Equity within, or when considering, an evaluation process is a key to combatting biases and assumptions rooted in racist evaluation structure. This is not to imply that evaluators are intentionally or inherently racist, but our views are shaped by our experiences. Without understanding who has input in the creations of these evaluations then we cannot understand the assumption and biases, and if we do not understand the assumption and biases then we cannot conduct program evaluations in a way that is going to tell us what we need to know. When the data that is needed is not collected, then programs cannot be responsive to, and for, change.

3. Method and Findings

The purpose of this study was to examine the evaluation process of nonprofit human service, organizations, in the central Virginia area to explore where equity does and does not exist, in the process of creating and conducting the evaluations. The methodology undertaken by the researcher was a mixed-method research design. Mixed methods combine qualitative and quantitative research. This type of methodology uses several different measures that include both contextual understanding like surveys, interviews, or observations along with facts or statistics. Using mixed methods helped the researcher investigate the topic to gain a comprehensive look at the subject. Using a survey-style questionnaire for quantitative data, the researcher asked specific, close-ended questions to collect nominal and numerical data from participants. The survey also gave space at the end of each section for participants to expand or give additional comments providing qualitative data as well. The researcher analyzed the data to identify patterns and similarities in responses and organizational characteristics. An interview was conducted to gain personal insight, using prepared structural questions. This means of qualitative research was selected as a way of contextualizing survey responses given in the quantitative portion of the project.

Population

The respondents of this study were staff, and affiliates, nonprofit organizations. For confidentiality purposes, neither the employee nor the organization's identity was recorded during data collection. These participants consisted of human service organizations who identified themselves as being tax-exempt under section 501(c)(3) of the Internal Revenue Code.

Location

The geographical location of this study was the Richmond, VA Metropolitan Statistical Area (MSA). The U.S. Office of Management and Budget (OMB) identifies the area as being inclusive of 17 county-level jurisdictions, including the independent cities of Richmond, Petersburg, Hopewell, and Colonial Heights. Greater Richmond includes the City of Richmond and Chesterfield, Hanover, and Henrico counties. The Richmond MSA encompasses these areas and also includes the cities of Petersburg, Colonial Heights, and Hopewell and the counties of Amelia, Caroline, Charles City, Dinwiddie, Goochland, King William, New Kent, Powhatan, Prince George, and Sussex (Greater Richmond Partnership, Inc., 2016). Due to the confidentiality of this study, the exact locality of each organization was not recorded. This demographic was selected because of the diversity in populations and service providers, as well as the background and history of the region, and accessibility of the researcher to these localities.

The Survey Design

From a review of past research, a selection of survey questions were crafted to gather information about the organization's evaluation process, its use of evaluations, and the participants' perceptions of organizational diversity. Questions were grouped in categories labeled Organizational Demographics, Evaluation Process, Evaluation Capacity, Stakeholder and Community Engagement, and Diversity, Equity, And Inclusion. From these categories data was collected to find out organizational characteristics such as length of time the organization has been operating, staff size, budget size, whether or not the organization conducted program evaluations, the purpose for conducting program evaluations, the instruments used to collect data, whether or not barrier exists that may hinder the process and what they may be, how the process is funded, and the participant's perception of their organization's views on community and stakeholder engagement. These categories and questions were asked under the assumption that the results would produce answers to the research questions.

There was no identifying information such as names, job titles, job description, organization name, location, or the like, asked or collected. After collecting and analyzing the data the researcher hoped to address the following questions, "At what point are equity and inclusion considered when developing a program evaluation?", "How often, if at all, are stakeholders included during the developmental phase of program evaluations?", "When developing a program's evaluation process, who is involved?", "How is cultural, and self, awareness a consideration when developing an evaluation plan?", and "What barriers exist in a stakeholder-inclusive evaluation process?"

Sampling & Sampling Techniques

Ethics approval was obtained through the university's IRB before any recruitment took place. From a contact list of service providers in the central Virginia area, a recruitment email informing participants of the purpose of the study and potential risk of participation, along with a link to an electronic copy of the consent form and survey. In addition to the initial email, follow-up emails were sent out 7 days after the initial email. A link to the electronic survey was emailed out in blind copied group clusters to assure confidentiality and random participation. Any participant that desired to respond was made aware that this was voluntary and at any point, during the survey, they had the right to discontinue the process. Recruitment was also done through several social media platforms.

The sampling was done voluntarily, and participants were informed of their right to opt-out at any time. A survey, consisting of 25 items, with Likert style statements, was presented. One open-end question, for any additional comments, concluded each item. The participants indicated their level of agreement with statements on a 6-point scale ranging from "strongly agree" to "strongly disagree." Questions ranging from the organization's services, demographics, and funding for the organization's evaluations practices, barriers to a more inclusive evaluation process, and the employee's perception of their organization's diversity and inclusivity. There were no identifiable questions such as ethnicity, race, or gender.

Sample

There were 234 emails sent out resulting in 40 recorded consents. Ultimately 25 participants completed the entire survey. The data for this study was collected from these participants through survey analysis. These participants were affiliated with a human service nonprofit organization. Out of the 40 link engagement, all 40 gave consent to continue, with 35 responding to the first question; from those 35 respondents, 32 identified as being tax-exempt under section 501(c)(3) of the Internal Revenue Code. The existence of these organizations ranged from 121 years to less than 12 months.

Organizational Demographics

To get an understanding of the diversity in the service population, the respondent was asked how they would categorize their organization. 5 respondents, or 17.24%, described themselves as providing education services, likewise with housing, and health/wellness, while 10.34% identified themselves as child/family service. The majority, almost 38%, selected others, ranging from recovery programs, support for formerly incarcerated citizens, community organizing/engagement, and voter information, to food justice (grocery delivery) and the faith community.

Table 1.

Demographic Information: Types Service Provide

Q11 - 2. Select the category best describe the service you provide?

#	Field	Choice Count	
1	Education	17.24%	5
2	Child/family service	10.34%	3
3	Health/wellness	17.24%	5
4	Housing	17.24%	5
5	Employment/job-training	0.00%	0
6	Other	37.93%	11
			29

Showing rows 1 - 7 of 7

Table 2.

Demographic Information: Other Types Service Provide

Other
Faith community
Community Organizing
Voter information
Legal
Recovery Program
Community organizing / engagement
support for formerly incarcerated
Food justice (grocery delivery)
support for organizations serving 0-5
community engagement
All of the above

Staff Size and Budget

When asked about the size of their organization, 13 participants, roughly 46.43%, said they have a staff of more than 15 people, with 8 having a staff of 1-5 people working for them, and 25% have between 6-10 assisting with services. When asked about their operating budget, 41.38% said they have a budget of more than $1,000,000, while 3 identified as having a budget ranging between 0-$10,000 to $50,000 to $100,000. 7 organizations operated with a budget range between $100,000-$500,000 and another 7, roughly 24.14% said their budget was between $500,000 and $1,000,000.

Table 3.

Demographic Information: Budget

Q13 - 4. What is the estimated operating budget of this organization?

#	Field		Choice Count
1	$0-$10,000	3.45%	1
2	$10,000-$50,000	3.45%	1
3	$50,000-$100,000	3.45%	1
4	$100,000-$500,000	24.14%	7
5	$500,000-$1,000,000	24.14%	7
6	$1,000,000+	41.38%	12
			29

Staff Diversity

When asked to indicate their level of agreement about staff diversity, 12 participants strongly agreed that they have a very diverse staff, while 4 leaned more to disagree with that statement. 42.86%, 12 participants disagreed with the statement "We have staff who identify themselves as having a disability" and 8 agreed that they have staff who identify as having a disability. When asked about their organization's racial composition 16 participants, 59%, agreed that they have a very diverse racial composition while 6 did not agree with that statement. Most of the organization had employees of various ages as indicated by 27/28 respondents with 1 disagreeing.

Table 4.

Demographic Information: Staff Diversity

Q14 - 5. Indicate your level of agreement with the following statements

Field	We have a very diverse staff at my organization.		We have employees of various ages.		We have a very diverse racial composition.		Our staff is inclusive of all sexual identities		We have staff who identify themselves as having a disability	
Strongly agree	42.86%	12	42.86%	12	37.04%	10	38.46%	10	17.86%	5
Agree	21.43%	6	32.14%	9	22.22%	6	19.23%	5	10.71%	3
Somewhat agree	21.43%	6	21.43%	6	18.52%	5	30.77%	8	3.57%	1
Somewhat disagree	3.57%	1	0.00%	0	0.00%	0	7.69%	2	21.43%	6
Disagree	3.57%	1	3.57%	1	11.11%	3	0.00%	0	42.86%	12
Strongly disagree	7.14%	2	0.00%	0	11.11%	3	3.85%	1	3.57%	1
		28		28		27		26		28

The demographics, with regards to diversity, would inform the assumption being outlined in an evaluation design and the type of instruments used to collect data. Our views and decision are created by our lived experience. Those experiences develop and drive our assumptions concerning different groups of people (Ritzer, 2018; Nkwake, & Morrow, 2016; Lawler,2013). When considering the diversity of the organization as a whole, in the event of evaluations being created and analyzed by staff, most of the respondents felt they worked within a diverse organization. However, when analyzing the data closer for specific populations represented within the organization, the responses suggested strong diversity in ages, agreement with racial diversity declined a little, and the majority expressed a lack of diversity about the disabled population. This is not to suggest that this was the case.

There was also a shift in responses concerning diverse sexual identity which alluding to the thought that there are characteristics that may be unknown without a degree of inquiry, with one respondent saying, “I am unaware of the sexual identities” and another saying, “There may be people with disabilities, however, I am unaware of people who have identified them to staff” As to nudge everyone to consider a response, there was no neutral choice provided in the survey, such as “neither agree nor disagree” to which those with uncertainties or some discomfort may have opted for that response.

Table 5

Demographic Information: Staff Size

Q12 - 3. How many employees work at this organization?

#	Field	Choice Count	
1	1-5	28.57%	8
2	6-10	25.00%	7
3	11-15	0.00%	0
4	15+	46.43%	13
			28

More than half of the organizations represented in the study housed staff sizes of less than 10 people. Half of those reported 5 staff members or less. This was also reflected in the budgets, with more than half of the respondents reporting an operating budget of less than $1,000,000. So, at the very least, there does not appear to be diversity in income. 1 respondent followed up by saying that it was “difficult to assess some types of diversity when you only are 4 people…cannot represent all identity categories even if you tried”. These insufficiencies, staff size, and finances presented themselves as a barrier to an organization that may or may not desire an equitable evaluation process.

More Qualitative Responses for Staff Diversity. One participant explained that they have 68 employees in Haiti and 8 in the US, stating that their “racial profile is different in terms of US-based organizations” going on to say, “We have 72 Black employees and 4 white”. Another explained that it was “difficult to assess some types of diversity when you only are 4 people...cannot represent all identity categories even if you tried.” One respondent admitted that their “diversity has been increasing over the past couple of years but still has a ways to go.”

Table 6.

Demographic Information: Operating Budget

Q13 - 4. What is the estimated operating budget of this organization?

#	Field		Choice Count
1	$0-$10,000	3.45%	1
2	$10,000-$50,000	3.45%	1
3	$50,000-$100,000	3.45%	1
4	$100,000-$500,000	24.14%	7
5	$500,000-$1,000,000	24.14%	7
6	$1,000,000+	41.38%	12
			29

Findings

The following sections are what the researchers discovered regarding Evaluation Process, Evaluation Capacity, Stakeholder and Community Engagement, and Diversity, Equity, And Inclusion. The analysis was about the respondents' level of agreement to whether or not the organization conducted program evaluations, the purpose for conducting program evaluations, the instruments used to collect data, whether or not barrier exists that may hinder the process, what those barriers may be, how the process is funded, and the participant's perception of their organization's views on community and stakeholder engagement. In addition to the quantitative data, qualitative data was also analyzed to bring clarity and context to the numerical information.

Evaluation Process and Purpose of Evaluations

After demographics and organizational composition were gathered, the respondents were provided a variety of statements about their evaluation process. When asked if the organization has a process for evaluating its programs, of the 31 participants who responded, 26, 83.87%, confirmed that they do have a process in place for evaluating its programs. To the statement "Program evaluation is a part of someone's job description in our organization.", 17 of the 24 respondents indicated some level of agreement, with 33.33% strongly agreeing, 20.83% agreeing, and 16.67% somewhat agreeing, leaving 7 who disagreed.

12 out of 24 agree when asked if their evaluations process was designed by an outside consultant. 5 of 23 respondents indicated that their evaluation process was conducted by an outside consultant. Under the category of "Purpose for Evaluations", 22 of the 24 agreed that they use evaluations to measure program performance (outputs) and program effectiveness (outcomes), 18 out of 24 indicated they use evaluation to prevent mission drift, 83%, 20 out of 24 indicated funder requirements as a reason for conducting evaluations, and 19 of the 24 indicated they conduct evaluations to measure program participants' satisfaction.

Table 7.

Number of Organization Whose Program Evaluation Process is Designed by an Outside Consultant

Q21 - 10. Our evaluations process was designed by an outside consultant.

#	Field		Choice Count
1	Strongly agree	21.74%	5
2	Agree	17.39%	4
3	Somewhat agree	13.04%	3
4	Somewhat disagree	4.35%	1
5	Disagree	21.74%	5
6	Strongly Disagree	21.74%	5
			23

Table 8

Number of Organization Whose Program Evaluations are Conducted by an Outside Consultant

Q20 - 11. Our program evaluations are conducted by an outside consultant.

#	Field		Choice Count
1	Strongly agree	12.50%	3
2	Agree	4.17%	1
3	Somewhat agree	8.33%	2
4	Somewhat disagree	8.33%	2
5	Disagree	45.83%	11
6	Strongly Disagree	20.83%	5
			24

Before getting into the mechanics of each organization's data collection methods, as not to assume, data was collected to find out if the organizations even conducted program evaluation. At first glance, the response to the questions seemed to contradict each other. When asked if there was a process for evaluating its programs, the survey generated 31 responses of which 26 answered yes, leaving 5 who said they did not have a process for evaluating their programs. Then when presented with the statement, “My organization has a clear system in place to evaluate our programs”, 2 expressed disagreement with 22 having some level of agreement with the statement. After further analysis I assume that either there is a system in place but not an evaluation process, there was confusion around what the difference was in inquiring about a “process” vs. a “system”, or, and this is because of the decline in responses the second time, maybe there was an assumption that the question was repetitive causing some to skip and move on to the following statement in the be. Either could be an explanation for what appeared to be a discrepancy in the responses of those who said they did not conduct evaluations and those who said they did not have a clear system in place.

Considerations When Planning Community Engagement

When considering the methods of data collection, the researcher’s goal was to find out what was a driving factor when deciding how to engage the community when collecting data. For this reason, the participants were asked to give their level of agreement with what their method of engagement is centered around. When it came to conducting program evaluations, 22 of the 23 participants indicated some level of agreement as it pertains to community needs, while 15 out of 22 respondents indicated that work schedules are considered when engaging in data collection. As for the ages of the population being served, 30.43% indicated that they strongly agree that their organization considers this demographic while a slightly higher 43.48%, or 10 of the 23, agreed with the statement, and 3, 13.04%, said they somewhat agreed that age was a factor when conducting evaluations, leaving 3 who did not agree with the statement concerning their organization and the age of their service population and when considering the level of education a client may have when considering how to approach an evaluation process, 18 out the 22 respondents agree that it is a consideration for their organization.

Community and Stakeholder Engagement in Program Evaluation

When considering an equitable evaluations process, the inclusion of internal and external voices is important. To answer the research questions around stakeholders' inclusion during the developmental phase of program evaluations, and at what point are equity and inclusion considered when developing a program evaluation, the researcher inquired about the input of internal and external participation when designing evaluations. When ask about staff participation in the design model, almost 43% agree that staff gave input into their evaluations process. When asked about community partners, 8 of the 23 represented organizations indicated that they did not agree that partners had input into their evaluation design. While 16 out of 23 respondents agreed that recipients of service were given a voice in the evaluations process, almost a third of the respondent, 7 out of 23, agreed when asked about a community member who did not utilize their service. About board member inclusion into the evaluations process, 7 agree that their board has a say in the design process, while 7 somewhat agreed, leaving 9 respondents who did not agree that board members had input in their evaluation design.

To find out about the inclusion of stakeholders there were several statements posed about the design process used by each organization. Understanding that some organizations bring in outside consultants to implement a method of program measurement, the statement was posed "Our evaluations process was designed by an outside consultant" to which more than half of the

Table 9.

Community and Stakeholders Inclusion in the Design Process

Q31 - 15. Equitable Evaluation Design

Field	Our employees assist with the designing of our evaluation process		Community partners participate in the designing of our evaluation process.		Community members who utilize our services participate in the design of our evaluation process.		Community members who do not utilize our services participate in the design of our evaluation process.		Board members participate in the design of our evaluation process.	
Strongly agree	39.13%	9	17.39%	4	21.74%	5	13.04%	3	21.74%	5
Agree	43.48%	10	26.09%	6	13.04%	3	4.35%	1	8.70%	2
Somewhat agree	8.70%	2	21.74%	5	34.78%	8	13.04%	3	30.43%	7
Somewhat disagree	4.35%	1	17.39%	4	0.00%	0	13.04%	3	8.70%	2
Disagree	4.35%	1	8.70%	2	21.74%	5	43.48%	10	13.04%	3
Strongly disagree	0.00%	0	8.70%	2	8.70%	2	13.04%	3	17.39%	4
		23		23		23		23		23

respondents confirmed that they did solicit the assistance of an outside consultant. This is not to mean that there was no collaboration on the part of the staff and/or community stakeholders partnering in this effort. Having experienced this method of collaboration, a follow-up question should have been asked to gain a better understanding of the process that took place. This notion can be supported by the response to the following statement of “Our program evaluations are conducted by an outside consultant” to which the majority of the respondents indicated that their evaluations were not conducted by outside sources. The data paints the picture that while the creation of the evaluations was outsourced, the actual administering of data collection was not leaving the assumption that there could have been a point of collaboration at least to understand and learn the process.

In addition to outside consultants, we wanted to know who else was involved in the development phase and how often. To this, the respondent was asked about the inclusion of employees, board members, community partners, and community members (who received services and who did not receive services). There were strong indications of employee assistance, further supporting the notion that there was a degree of collaboration between consultants and staff in the designing of these evaluation processes. While there were different degrees of agreement (strongly agree, agree, and somewhat agree. Table. 9) as it pertained to board members and recipients of service, there was stronger representation in the areas ranging from somewhat agree than and somewhat disagree. This suggests to me a level of uncertainty or limited participation. It was clear, however, that there was very little inclusion of community members who did not utilize the service of the organization. There looks to be little engagement with stakeholders who do not have a direct affiliation with the organization and its program. This demographic could be inclusive of family and neighbors for those who receive services, other service providers who also serve the same community members, or their family and friends. When looking at a community as an ecosystem interwoven and connected, these could be perspectives that could shine a light on blind spots or assist in eliminating duplication in services. Overall, 18 express some degree of satisfaction with the organization's evaluation process, and 5 expressed that they were not satisfied with their organization's evaluation process.

Challenges and Barriers

When reviewing past literature on organizations and equitable evaluation process, and as confirmed through participant interviews, the researchers wanted to understand the reasons some organizations may not engage in a more equitable process. Something that was brought to the attention of the researcher was the existence of barriers or hindrances that may prevent an organization from being inclusive of stakeholders. It is important to note that barriers are not only roadblocks preventing some form of program evaluations, they could also be responsible for organizations not being more equitable or the reason some are using the process they elected to use.

Table 10.

Related Challenges to the Evaluations Process

Q41 - 19. Barriers

Field	Barriers prevent my organization from conducting program evaluations.		Barriers prevent my organization from conducting evaluations that are more equitable and inclusive.		Adequate funding is a barrier to program evaluations.		Staffing is a barrier to conducting program evaluation.		Technology is a barrier to conducting program evaluations.		Proper training on conducting program evaluations is needed at my organization.	
Strongly agree	8.00%	2	12.50%	3	16.00%	4	16.00%	4	20.00%	5	20.00%	5
Agree	12.00%	3	4.17%	1	16.00%	4	12.00%	3	12.00%	3	32.00%	8
Somewhat agree	44.00%	11	29.17%	7	24.00%	6	36.00%	9	8.00%	2	16.00%	4
Somewhat disagree	4.00%	1	16.67%	4	4.00%	1	4.00%	1	16.00%	4	12.00%	3
Disagree	16.00%	4	16.67%	4	24.00%	6	20.00%	5	20.00%	5	4.00%	1
Strongly disagree	16.00%	4	20.83%	5	16.00%	4	12.00%	3	24.00%	6	16.00%	4
		25		24		25		25		25		25

In addition to the responses collected from the respondents concerning staff size, a large percentage of the respondents indicate training as an additional barrier to conducting program evaluation, stating “We need more internal commitment to it, difficult skill set to apply”. I gave attention to the number of respondents who agreed that barriers prevented them from conducting program evaluations along with those who felt barriers did not prevent them from conducting program evaluation versus the number who felt that barriers prevented and did not prevent them from conducting evaluations that are more equitable and inclusive. This, to me, speaks to the assumption of respondents think it would not require too much more effort to incorporate an evaluation model that was equity-focused. Perhaps the desire for more training suggests that it is simply a matter of being trained on inclusive methods and having the staff implement them. I think this shows a greater need in understanding the benefits of equity in evaluations and not so much, not having the ability to incorporate an equitable model. And while funding was not the biggest barrier, it was a barrier, nonetheless.

The data seem to indicate staff, training, and technology to be more of a barrier than funding with a large percentage (68%) indicating that their budget allocates money specifically for evaluations/assessments. On the other hand, there were strong indications that funding was not provided or sought specifically for program evaluations leaving me to assume, and the data supports, that the main purpose is administrative in nature. The high number in support of “outcomes”, “outputs”, “participants satisfaction”, and “funding requirements” suggests the use of logic models, which as previously stated in the literature, is often a requirement of funders (Jones, et al, 2020). Hard to answer these questions because there are different kinds of evaluations.

Table 11.

Funding for Program Evaluation

Field	Our operating Budget allocates money specifically for evaluations/assessments.		We seek funding specifically for program evaluations.		We receive funding specifically for program evaluations.	
Strongly agree	12.00%	3	0.00%	0	0.00%	0
Agree	24.00%	6	8.33%	2	8.00%	2
Somewhat agree	32.00%	8	25.00%	6	8.00%	2
Somewhat disagree	4.00%	1	16.67%	4	16.00%	4
Disagree	24.00%	6	41.67%	10	56.00%	14
Strongly disagree	4.00%	1	8.33%	2	12.00%	3
		25		24		25

Qualitative Responses to Challenges and Barriers. About hindrances when considering program evaluation, respondents gave their perspectives on what may be needed or could be improved to increase evaluation capacity. One participant stated that they “do not have a good documentation system so it impacts the efficacy and effectiveness of our evaluation process" while another said, "We need more internal commitment to it, difficult skill set to apply". One respondent expressed the need for "adequate staff to provide these services" with 7 responding that they agree and 9 stating that they somewhat agree. The remaining 9 did not agree that staffing was a barrier to conducting program evaluation. With technology, funding, and staff being a few barriers, a high percentage of respondents agreed that proper training on conducting program evaluations was needed with one survey participant saying, "I would love to participate in a training for program development and evaluation".

Additional Qualitative Analysis

Whether or not organizations conduct performance evaluations to stay on track with their mission or measure the impact a program has on the community it serves, most of the representatives express that their organization conducts some form of evaluation. Regardless of if it is a participant stating that "We have been successful in conducting evaluations of programming and services and modify our services according to results received" or one wondering "Are evaluations this important?", they all express some level of agreement that their organization believes it is important to consider the expressed needs of community members and they believes it is important to consider the concerns of community partners. Having made changes in their organization as a result of feedback received from stakeholder and community engagement, and having a belief that the relationships they hold with stakeholders "plays a pivotal role in how the organization moves forward and the direction in which it goes" it is not clear if organizations see the incorporations of an inclusive evaluation process as an avenue to achieving their goal or if they are "working on building this into our work so that it's not 'one more thing' that we have to do".

4. Discussion

Evidence has determined that evaluations can contribute greatly to program interventions by placing more focus on outcomes measurement. The validity of data, and their analysis, is key to a credible evaluation process. Outcome measurements that are influenced by bias-driven assumptions and beliefs, can misinform results that influence change (Nkwake, & Morrow, 2016; Sussman & Wilson, 2019). As it is, communities are being provided services, from programs, based on data, collected from data collections instruments that reflect these assumptions and biases of the creators. In turn, we address problems that are not always the priority of the community and provide interventions that do not always address the root causes. If we do not understand who has input in the creations of these evaluations then we cannot understand the assumption and biases, and if we do not understand the assumption and biases then we cannot conduct program evaluations in a way that is going to tell us what we need to know. When the data that is needed is not collected, then programs cannot be responsive to, and for, change. However, when we know who is at the table providing input then we can better identify where the blind spots are.

This study explores where equity – defined as being "the state, quality or ideal of being just, impartial and fair"- does and does not exist, in the process of creating and conducting the evaluations that are used to measure the successful execution of nonprofit programs (Wiggins and Sileo, 2020 p.1). In other words, this study seeks to pinpoint where stakeholder and community input get introduced in the program evaluation process. The research examined the demographics and evaluation practice of nonprofit human service organizations in the Richmond Tri-city area. Through a series of questions, we surveyed respondent of each organization to collect data on the organization's size, diversity, and methods in which each organization create and conduct program evaluations. The analysis of the data allowed us to understand if the organizations conducted any type of evaluations, who participated in the design process if funding was available specifically for the evaluation process, the organization's capacity to perform evaluations, specifically equitable and inclusive evaluations, views on stakeholder engagement, and how the collected data was used.

Major Discoveries

After careful examination of the data, the research did not produce any major theme or consistencies, however, there was a tone that presented itself a few times. Overall, each respondent that used the comment sections to expand, expressed their belief that their organization either employed a diverse staff, was working towards a more diverse staff, or expressed reason for the composition of their staff. There was also, in my opinion, a feeling of offense or a need to defend the practices of the organization. This is often a response that has been observed when highlighting possible practices that may promote inequality and inequity. What we might want to consider is, without this awareness, change cannot take place. Also, No one is calling us bad people because we need to make some adjustments in our views and practices. Before we can begin to shift our practices, we have to be ok with admitting the current ones are not effective.

Implications for Nonprofits Organizations

Before the introduction of Community-Base Participatory Research (CBPR), which involved the start-to-finish inclusion of those being impacted by the intervention resulting from data, community engagement started and stopped at the data collection point in research and evaluations (O'Sullivan, 2012; Lucero, Boursaw, Eder, Greene-Moton, Wallerstein, & Oetzel, 2020). CBPR argued that communities needed to be included throughout the entire research process, this is also inclusive of the program and other types of evaluations processes (Horowitz, Robinson, & Seifer, 2009). In addition to understanding, if the organizations participating in this study, had any type of process to measure the success of their programs, we sought to find out at what point recipients of the program's service, and other stakeholders, were included in the evaluation process.

Although there was some indication of stakeholder inclusion, there was also evidence to suggest more effort could be made to include the experiences of community partners, community members who do not receive services, and board members. Each of these groups has a unique perspective that could help inform decisions. There was also a low representation of those who identify as having a disability when asking about staff diversity. As expressed in the qualitative data, there is some identifier that is not visible or readily known without inquiring, and therefore, evaluators may not be aware of those to include. From a policy standpoint, there is a matter of intrusion in addition to privacy laws concerning a person's medical status, sexual identity, and sexual orientation. There are Human Resources and other policies are in place for the protection of certain vulnerable populations. A manner of including these voices while being sure to protect privacy would be to partner with agencies that employ and/or serve these populations. There is also a means of recruitment that will produce a voluntary presence of such populations. This can also, lead to the snowball effect in which one member of the desired community will tell another about an opportunity or needs, creating a selection through word of mouth.

Implications for Funders

When the topic of program evaluations came up in my conversations, the topic of funding was often the second, if not the main resource that was said to be limited. This was the reason I found it odd that it was not directly implied to be a barrier for organizations to conduct equitable evaluations. I say directly because, training, staff, and technology were top, and they are also major overhead expenses for nonprofit organizations. There were quite a few implications regarding funding, with the first being the lack of income diversity within the represented organizations. When looking at the staff size to budget ratio, with the understanding that pay-roll with probably an organization's biggest expense, a $1,000,000 budget is not much with a staff of 5 or more people. Likewise, a $10,000 budget probably means a small organization is most than likely run by volunteers. Where then would a lengthy evaluation process come in, where there is, compensation for community members' time and effort. And then there is the need for training and technology. Equitable evaluations are not just diverse people who bring diverse perspectives, it is also a diverse model to cater to diverse needs. This has to be taught and train and requires time and compensation.

The was also an indication that the main reason evaluation of any kind was being conducted was due to funder and donor requests and requirements. This was an indication of the power and influence donors, and funders have to push for more equitable evaluations, with the resource and connections to assist with making the process easier. Most foundations are already using consultant firms. Community Foundations already have access to training. They are already doing outcome evaluations because they probably have no use for the output of process data collection. A funder and simply change the requirement of a grantee, requesting more diversity in their reports. Or request proof of inclusion of stakeholders in the process. As reported in the literature, the logic model is still in use mostly because of funding and donor requirements.

From an organizational level, while some identified that there are employees with the responsibility of evaluation, and budget line items specific for evaluations, few sought funding specifically for program evaluation, and almost none received funding specifically for program evaluation. This lack and need is a shared responsibility. Sure, we can argue that funders are not funding program evaluations, but we cannot overlook the fact that organizations are not seeking and making these demands. Lastly, one reason evaluations maintain their roots in structural racism is that there remains no reward, incentive, or real reason to confront the status quo and make that shift to equity and inclusion of community stakeholders (Caldwell, & Bledsoe, 2019). Evaluators run more of a risk of a reprimand than a reward. Funders can change this behavior as well. The effort to pursue and advance equity needs to be celebrated and rewarded and at this point, it is a lonely place for an evaluator to find themselves.

Limitations

Looking back, a few of the responses and lack thereof, we could have provided more clarity regarding some questions asked and concepts used in the survey. I addition to some skipped responses, there was some respondent who expressed a lack of understanding of some terminology. One needed more clarity around the types of evaluations in question stating it was "Hard to answer these questions because there are different kinds of evaluations. Each grantor is evaluating us in different ways, and we send evals to our clients and volunteers". There were also assumptions made on the part of the research. For instance, there was the assumption that there was no collaboration with staff or stakeholders and outside consultants who participated in the designing of the evaluation process. There could have been follow-up questions to gain more clarity around what that process looked like. Lastly, working in a sector that conducts a lot of surveys, I understand survey fatigue. People just are not fans of surveys, with that said, I would suggest having several methods of data collection to help fill in gaps.

Conclusion

It is extremely important to understand how our personal biases and assumptions show up in the work we do. To assume that we have none and that there is no need for self-reflection can, at the very least, cause us to spend time and resources with very little change taking place. In addition to organizations, and their staff, understanding the importance of having diversity in perspectives seated at the same table, funders and policymakers need to understand the importance as well as not to assume to be experts of someone else's experiences. A lack of representation of a population does not just risk providing insufficient services, it could mean not considering the accessibility of an office or the ability to understand and complete our surveys. It could look like not considering the literacy rate of a community because it has never been something we had to consider when navigating our lives.

An equitable evaluation process would have minimized some of the limitations that presented themselves in this study. The IRB process combined with Covid-19 and other semester constraints did not make for an equitable process, which lead to assumptions and ultimately, confusion for the survey respondents. Equity within an evaluations process, as suggested in the literature, also means understanding our limitations, the limitations of the community than providing an evaluation model, or options of several models, that will not only best serve the needs of the community, but also best serve the needs of the staff. One interview respondent, when asked about their process, said that with the composition of the staff, they conduct an evaluations process that best works for them, even though it had not been revised in 6 years.

Even though funding did not seem to be a barrier for most of the participants of this study, they also did not appear to think seeking funding specifically for program evaluations was important, or maybe it was not considered or even worse, they may simply not know it was an option. Outside of donor requirements, there is usually nothing that requires an organization to prove they are making a difference in the communities they reside and serve. We need to understand the difference between community outreach, community service, and community engagement. Because we do toy drives (community service) and provide job training (community outreach), does not mean the community expressed a need for either (community engagement). And when these needs are expressed, interventions should also be planned, designed, and implemented in partnership with that community not imposed on that community. Moving forward, let us stop walking around as experts and allow ourselves to also become students.

Reference

Acero, N. (2019). Equitable evaluation: Elevating participant voices in youth development program evaluation. California State University – Northridge. http://hdl.handle.net/10211.3/214662

Bamberger, M., & Segone, M. (2011). How to design and manage equity-focused evaluations. UNICEF Evaluation Office.

Bellwether Education Partners. (2020). Approaches for diverse, equitable, and inclusive evaluation.

Bopp, C., Harmon, E., & Voida, A. (2017). Disempowered by data: Nonprofits, social enterprises, and the consequences of data-driven work. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 3608-3619). http://dx.doi.org/10.1145/3025453.3025694

Brown, M. A., & Di Lallo, S. (2020). Talking circles: a culturally responsive evaluation practice. *American Journal of Evaluation*, 41(3), 367–383. https://doi.org/10.1177/1098214019899164

Bustamante, M., Vidueira, P., & Baker, L. (2021). Systems thinking and complexity science informed evaluation frameworks: Assessment of The Economics of Ecosystems and Biodiversity for Agriculture and Food. *New Directions for Evaluation*, 2021(170), 81â€ "100. doi:10.1002/ev.20455

Caldwell, L. D., & Bledsoe, K. L. (2019). Can Social Justice Live in a House of Structural Racism? A Question for the Field of Evaluation. *American Journal of Evaluation*, 40(1), 6–18. https://doi.org/10.1177/1098214018815772

Chauveron, L. M., Samtani, S., Groner, M. G., Urban, J. B., & Linver, M. R. (2021). Including diverse stakeholder voices in youth character program evaluation. *American Journal of Evaluation*, 42(2), 221–236. https://doi.org/10.1177/1098214020917218

Dean-Coffey, J. (2018). *What's race got to do with it? equity and philanthropic evaluation practice. American journal of evaluation*, 39(4), 527–542. https://doi.org/10.1177/1098214018778533

Despard, Mathieu R. (2016) Strengthening evaluation in nonprofit human service organizations: results of a capacity-building experiment, human service organizations: management, leadership & governance, 40:4, 352-368, DOI: 10.1080/23303131.2016.1140101

Donnelly, J. (2010). Maximizing participation in international community-level project evaluation: a strength-based approach. *Evaluation Journal of Australasia*, 10(2), 43–50. https://doi.org/10.1177/1035719X1001000206

Foote, J., Midgley, G., Ahuriri-Driscoll, A., Hepi, M., & Earl-Goulet, J. (2021). Systemic evaluation of community environmental management programmes. *European Journal of Operational Research*, *288*(1), 207-224. https://doi.org/10.1016/j.ejor.2020.05.019

Freire, P. (1972). *Pedagogy of the oppressed*. New York: Herder and Herder

Frye, A. W., & Hemmer, P. A. (2012). Program evaluation models and related theories: AMEE guide no. 67. *Medical teacher*, *34*(5), e288-e299. https://doi.org/10.3109/0142159X.2012.668637

Hall, J. N., Freeman, M., & Colomer, S. E. (2020). Being culturally responsive in a formative evaluation of a professional development school: Successes and missed opportunities of an educative, values-engaged evaluation. *American journal of evaluation*, 41(3), 384–403. https://doi.org/10.1177/1098214019885632

Glueck, Katie (2020). "Joe Biden says he would choose a woman as his running mate. but who?". *The New York Times*.

Garcia, G. L., & Stevahn, L. (2020). Situational awareness and interpersonal competence as evaluator competencies. *American journal of evaluation*, 41(1), 107–124. https://doi.org/10.1177/1098214018814941

Greater Richmond Partnership, Inc. (2016). Demographics. Retrieved from Greater Richmond Partnership, Inc: http://www.grpva.com/data-and-downloads/demographics/

Hay, K. (2017). Equity is not an intervention: Implications of evaluation—reflections from India. In S. Sridharan, K. Zhao, &A. Nakaima (Eds.). *Building capacities to evaluate health inequities: some lessons learned from evaluation experiments in china, india and chile. New directions for evaluation*. *154*, 79–89

Horowitz, C. R., Robinson, M., & Seifer, S. (2009). Community-based participatory research from the margin to the mainstream: are researchers prepared?. *Circulation*, *119*(19), 2633-2642. https://doi.org/10.1161/CIRCULATIONAHA.107.729863

Jones, N. D., Azzam, T., Wanzer, D. L., Skousen, D., Knight, C., & Sabarre, N. (2020). Enhancing the Effectiveness of Logic Models. *American journal of ev*aluation, 41(3), 452–470. https://doi.org/10.1177/1098214018824417

Kimball, E., Abbott, J., & Childs, J. (2020). Cripping equity and assessment: disability as identity and culture in the context of culturally responsive (Equity Response).

Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment

Lasater, K., Bengtson, E., & Albiladi, W. S. (2021). Data use for equity?: How data practices incite deficit thinking in schools. *Studies in Educational Evaluation, 69*, 100845. https://doi.org/10.1016/j.stueduc.2020.100845

Lawler, S. (2013). Identity: sociological perspectives, 1st Edition. Polity Press

Lee, C., & Clerkin, R. M. (2017). The adoption of outcome measurement in human service nonprofits. *Journal of public and nonprofit affairs, 3*(2), 111-134. https://doi.org/10.20899/jpna.3.2.111-134

Lichtenstein, Melanie (2018) Conceptual model of educational planning and program evaluation: addressing deficit thinking at the top. *The William & Mary Educational Review*: Vol. 5: Iss. 1, Article 7. Available at: https://scholarworks.wm.edu/wmer/vol5/iss1/7

Liket, K. C., & Maas, K. (2015). Nonprofit Organizational Effectiveness: Analysis of Best Practices. *Nonprofit and voluntary sector quarterly*, 44(2), 268–296. https://doi.org/10.1177/0899764013510064

Lucero, J. E., Boursaw, B., Eder, M. “Mickey,” Greene-Moton, E., Wallerstein, N., & Oetzel, J. G. (2020). Engage for Equity: The Role of Trust and Synergy in Community-Based Participatory Research. Health Education & Behavior, 47(3), 372–379. https://doi.org/10.1177/1090198120918838

Mapitsa, C. B., & Ngwato, T. P. (2020). Rooting evaluation guidelines in relational ethics: lessons from Africa. American journal of evaluation, 41(3), 404–419. doi:10.1177/1098214019859652

Mason, S., & Azzam, T. (2019). In need of an attitude adjustment? the role of data visualization in attitude change and evaluation influence. *American journal of evaluation*, 40(2), 249–267. https://doi.org/10.1177/1098214018778808

Montenegro, E., & Jankowski, N. A. (2017). *Equity and assessment: moving towards culturally responsive assessment.* (Occasional Paper No. 29). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Montenegro, E., & Jankowski, N. A. (2020). *A new decade for assessment: embedding equity into assessment praxis* (Occasional Paper No. 42). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Netting, F., & O'Connor, M. (2013). Program planning and implementation: designing responses to address community needs. In M. WeilM. Reisch, & M. L. Ohmer *The handbook of community practice* (pp. 757-772). SAGE Publications, Inc., https://www.doi.org/10.4135/9781412976640.n36

Nkwake, A. M., & Morrow, N. (2016). Clarifying concepts and categories of assumptions for use in evaluation. *Evaluation and program planning*, *59*, 97-101. https://doi.org/10.1016/j.evalprogplan.2016.05.014

Odera, E. L. (2021). Capturing the added value of participatory evaluation. *American journal of evaluation*, 42(2), 201–220. https://doi.org/10.1177/1098214020910265

O'Sullivan, R. G. (2012). Collaborative evaluation within a framework of stakeholder-oriented evaluation approaches. *Evaluation and program planning*, *35*(4), 518-522. doi: 10.1016/j.evalprogplan.2011.12.005. Epub 2012 Feb 2. PMID: 22364849.

Patton, M. Q. (2018). A historical perspective on the evolution of evaluative thinking. *New directions for evaluation*, *2018*(158), 11-28. https://doi.org/10.1002/ev.20325

Patton, M. Q. (2019). Expanding futuring foresight through evaluative thinking. World Futures Review, 11(4), 296–307. https://doi.org/10.1177/1946756719862116

Phillips, Amber (2020). Who are the women probably on Joe Biden's shortlist for vice president? *The Washington Post*.

Public Policy Associates, Inc. (2015). Considerations for conducting evaluation using a culturally responsive and racial equity lens. Lansing.

Reynolds, M. (2014). Equity-focused developmental evaluation using critical systems thinking. *Evaluation*, *20*(1), 75-95.

Ritzer, G. (2018). *Sociological theory, 10 Edition*. Thousand Oaks, Cali. Sage Publications

Stern, A., Guckenburg, S., Persson, H., Petrosino, A., & Poirier, J. (2019). Reflections on applying principles of equitable evaluation.

Sussman, J., & Wilson, M. R. (2019). The use and validity of standardized achievement tests for evaluating new curricular interventions in mathematics and science. *American Journal of Evaluation*, 40(2), 190–213. https://doi.org/10.1177/1098214018767313

Szijarto, B., & Cousins, J. B. (2019). Making space for adaptive learning. american journal of evaluation, 40(2), 160–176. https://doi.org/10.1177/1098214018781506

Thomas, V. G., Madison, A., Rockcliffe, F., DeLaine, K., & Lowe, S. M. (2018). Racism, social programming, and evaluation: where do we go from here? *American journal of evaluation*, 39(4), 514–526. https://doi.org/10.1177/1098214018772910

Viser, Matt; Johnson, Jenna; Scherer, Matt (2020). Biden pledges he'll pick a female running mate, sanders says he likely will in debate under the shadow of coronavirus. *The Washington Post.*

Wiggins, M., & Selio, A. (2020). What's the role of equity in evaluation policy? The Forum for Youth Investment.