Quality Translated? Comparing Nonprofit Evaluation Practices in Germany and the United States

Karl Urban, University of British Columbia

ABSTRACT

This article presents findings from a study of management practice in youth career assistance non-profits in Germany and the United States, focusing on the area of evaluation. It was hypothesized that the institutional frameworks of the welfare regime, public administration, and the nonprofit sectors' origins play an essential role in shaping evaluation practices at the level of operative management. Interviews with managers in both countries were conducted utilizing the World Management Survey in a mixed methods design. Data were evaluated using statistical methods and qualitative content analysis. The findings indicate significant quantitative and qualitative differences between nonprofit evaluation practices in both countries. These results are discussed within the institutional framework used for hypothesis formulation, concluding with suggestions of future research avenues for internationally comparative nonprofit scholarship.

RÉSUMÉ

Cet article présentera les résultats d'une étude sur les pratiques de gestion des organisations à but non lucratif d'aide à la carrière des jeunes, en Allemagne et aux États-Unis. Le thème central sera sur l'évaluation. L'hypothèse est que les cadres institutionnels du régime de protection sociale, l'administration publique et les origines de secteur non lucratifs jouent un rôle essentiel dans le façonnement des pratiques d'évaluation au niveau de la gestion opérationnelle. Des entretiens ont été menés avec des directeurs des deux pays avec la World Management Survey, mais en utilisant un conception de méthodes mixtes. Les données ont été évaluées à l'aide de méthodes statistiques et d'une analyse qualitative. Les constatations indiquent des différences quantitatives et qualitatives entre les pratiques d'évaluation dans les deux pays. Ces résultats sont discutés dans le cadre institutionnel utilisé pour formuler l'hypothèse, et se terminent par des suggestions de pistes de recherche potentielles pour une bourse d'étude comparative internationale du secteur non lucratif.

Keywords / Mots clés : nonprofit management, quality assurance, youth career services, international management research / gestion à but non lucrative, assurance qualité, services d'orientation professionnelle pour les jeunes, recherche en gestion internationale

Managers in human service nonprofit organizations (HSNs) conduct their work in contexts informed by locally situated and globally informed institutional paradigms, which are frequently ambiguous and sometimes contradictory (Anheier, 2010; Sanders, 2012). Neoliberal social policy practice, including increased contracting of government services to nonprofit, for-profit, and hybrid organizations, has led to wholesale and subtle import of business-like management practice, blurring the lines between sectors (Bromley & Meyer, 2017; Skelcher & Smith, 2015). The institutionalization of hybrid and for-profit organizational forms, such as B-Corps and social enterprises, has hastened the discursive and material shift of HSNs toward business-like management (Kopaneva & Cheney, 2019; Mendell, 2017). However, social services management, particularly in HSNs, does not consist solely of copying public sector austerity programs or private sector emphases on efficiency and effectiveness (Hasenfeld, 2015). Current scholarship shows a wide array of nonprofit management behaviour, including outright rejection of external pressures, tacit and explicit forms of resistance, hybrid forms of organizing, and orderly acceptance (Maier, Meyer, & Steinbereithner, 2016). These varieties of nonprofit management are informed by contextual demands (Baines, Charlesworth, & Cunningham, 2014) and the boundedly rational responses to these demands (Hasenfeld, 2015; Hasenfeld & Garrow, 2012; Simon, 1979). Cross-national quantitative studies of management practice also point to significant differences between countries and organizational types (Bloom, Lemos, Sadun, & Van Reenen, 2015; Bloom & Van Reenen, 2010; Delfgaauw, Dur, Propper, & Smith, 2011; Keller, 2011).

This article focuses on the management area of evaluation as an example of topical and contested nonprofit organizing practice (Arvidson & Lyon, 2014; Carnochan, Samples, Myers, & Austin, 2014; Moxham, 2010) in Germany and the United States. Contrasting the contours of evaluation practices and their relationships with different contexts or "accountability regimes" (Ebrahim, 2009) allows for a nuanced understanding of the unique conditions of nonprofit management and the third sector more broadly. According to Benjamin and Campbell (2020), "evaluation involves the systematic gathering of information about an entity, called an evaluand, to determine its merit or worth, inform decision making and improve policy" (p. 198). The example of HSN evaluation practice provides an intriguing entry point to wider debates on the efficacy of social interventions (Ebrahim, 2019) and widely varying notions of accountability (Perrin, 2018). Both issues are of special relevance to current nonprofit scholarship given the role of evaluation in the (neoliberal) marketization and professionalization of the nonprofit sector (Lee & Clerkin, 2017; Post & Dodge, 2019; Roy, Eikenberry, & Teasdale, 2021). This article makes a novel contribution to the comparative, context-situated study of nonprofit management practices.

In sketching out a coherent theoretical background for this study, applications of internationally comparative scholarship on for-profit management practice prove to be limited. After briefly summarizing this field, attention turns to the article's analytic approach, organizational institutionalism. Evaluation practice in HSNs will be contextualized within the historically, socially, and culturally located institutions. Comparative theories of the welfare state (Esping-Andersen, 1990), public administration reform (Pollitt & Bouckaert, 2017), and the social origins of the nonprofit sector (Salamon & Anheier, 1998; Salamon, Sokolowski, & Haddock, 2017) will be explicated and applied to the topic of evaluation practice. These comparative theories will inform later data analysis and discussion.

Data was collected as part of a larger comparative study of German and US nonprofit management practices in the field of youth career assistance services (YCAS). Youth career assistance nonprofits present a compelling case for comparative study because YCAS reflects the particular historical, social, and institutional growth of social services and the nonprofit sector in Germany and the United States (Ryan, 2001). Youth career assistance, known in Germany as *Jugendberufshilfe*, offers holistic support to youth and young adults before, during, and after the transition from school to higher education, apprenticeships, or work, with the secondary goal of social participation (Oehme, 2016). Whereas youth social work and career assistance is legally, institutionally, and financially anchored in German society (Sozialgesetzbuch, 1990), similar services in the United States are located in an ambiguous, localized, and piecemeal institutional framework oriented around the public school system and a fragmented post-secondary education landscape (Savickas & Savickas, 2019).

Data were generated using a standardized interview guide, the World Management Survey (Delfgaauw et al., 2011; Van Reenen & Bloom, 2007). Analysis of the data utilized a mixed methods approach to illuminate specific differences and similarities between evaluation practices, and to contrast reasoning for these practices within the specific institutional contexts. The primary research questions pursued in this study include

a What differences exist between quality assurance practices in YCAS nonprofits in Germany and the United States?
b How do differences and similarities in nonprofit quality assurance practices relate to their respective institutional environments?

THEORETICAL BACKGROUND
Institutionally informed comparative nonprofit management
Much of internationally comparative management research has sought to survey or create a normative basis for management practice across borders (Farmer & Richman, 1964; Hofstede, 2001; House, Dorfman, Javidan, Hanges, & Sully de Luque, 2014; Van Reenen & Bloom, 2007). In contrast to these institutionally blind approaches, this article offers a theoretical foundation based in organizational institutionalism. Organizational institutionalism is a school of organizational theory that emphasizes the importance of myth, ritual, and script in creating and sustaining organizations (Scott, 2017). Tacit meaning, identity, and storytelling take priority over nominally definitive categories such as efficiency and effectiveness in organizational life (Greenwood, Oliver, Lawrence, & Meyer, 2017; Meyer & Rowan, 1977). In contrast to actor-centred approaches, the organizational institutionalist perspective views managerial practice as fundamentally shaped by legitimacy and appropriateness, rather than efficiency and effectiveness. Institutional fields require context-specific practices of actors. Managers "translate" and "edit" practices to recontextualize and redramatize them in the specific organizational location in which they operate. Wedlin and Sahlin (2017, p. 7) characterize translation as "simultaneous processes of movement and transformation."

The institutional environment for HSNs consists of a complex system of intersections between public, private, and civil society actors and structures (Anheier & Toepler, 2020; Casey, 2016b). These structures, and the logics associated with them, shape and are shaped by practices, including by those of managers in nonprofit organizations. This relationship is one of mediation between structure and ac-

tion, fostered to an extent by managers' translations of ideas and practices. Thereby, actors' agency within organizations is located temporally, spatially, culturally, socially, and institutionally, which allows and disallows action (Emirbayer & Mische, 1998; Hargrave & Van De Ven, 2009).

Nonprofit context: Welfare regimes

Welfare regimes are arguably the most important contextual factor that shapes German and US HSN management, providing not only funding and regulation, but also a broad set of operational logics that socialize and otherwise shape an organization's mission and work (Casey, 2016a; Salamon et al., 2017). Esping-Andersen (1990) theorized three welfare regime types: liberal, conservative, and social democratic. Although not conceptually exhaustive, this analysis is an effective way to organize discussion of basic differences vis-à-vis ideal types in comparative nonprofit studies (Casey, 2016b; Salamon & Anheier, 1998). It seeks to answer the question of, on an institutional and path-dependent level, "how countries can either use the market or diverge from it by . . . conservatively upholding social stability" (Schröder, 2013, p. 14). Its primary typification and conclusions regarding Germany and the United States have been echoed by more recent scholarship, for example, by Starke, Obinger, & Castles, (2008) on policy convergence and Schröder (2013) on interinstitutional complementarity.

Germany represents a prototypical "conservative" welfare regime. This type is characterized as corporatist, in that workers, companies, and other societal interest groups engage in organized power sharing, while remaining subordinate to the state. Rooted in feudal relations, the conservative regime is patriarchal in the sense of protection and privilege bestowed by the state to certain social groups, precluding a liberal "obsession with market efficiency and commodification" (Esping-Andersen, 1990, p. 27). Conservative welfare regimes serve to preserve social order and hierarchy. Stratification is a key aspect of the conservative paradigm insofar as it contextualizes the individual by fossilizing social orders (Van Kersbergen & Kremer, 2008). In line with the (neo-)corporatist philosophy of the German welfare regime, many services are provided by private HSNs associated with major societal interest groups such as the Catholic and Protestant churches, and organized labour (Esping-Andersen, 1990). Accountability is thereby diffused throughout the major players in the welfare regime (Anheier, Hass, & Beller, 2013).

Conversely, the United States' welfare regime is representative of the "liberal" type according to Esping-Andersen (1990). Liberal welfare regimes are market facing, seeing the invisible hand of the free market as an equalizer of opportunity among atomized individuals (Lynch, 2014). Modest and strictly regulated (i.e., means tested) support and services are provided only to the lowest-income and most disadvantaged citizens. This leads to a significant stigmatization of service users and incentivizes benefit refusal, as Esping-Andersen (1990, p. 26) writes: "the limits of welfare equal the marginal propensity for welfare instead of work." The liberal welfare regime actively supports marketization of HSN services. This is achieved through subsidization or contracting, or by reducing benefits with the intention of passively encouraging private initiative to fill the gap created by the state's absence (Esping-Andersen, 1990), a trend that has increased in the past decades in both countries (Henriksen, Smith, & Zimmer, 2012). Notions of accountability are heavily influenced by these marketizing arrangements (Abramovitz & Zelnick, 2018; Sandberg, Elliott, & Petchel, 2019).

Nonprofit context: Public administration reform

The politics, policy, and political culture of public administration is closely intertwined with HSN management practice (Pandey & Johnson, 2019). The United States and Germany both have decentralized federal structures, which gives states and local governments extensive leeway in enacting reform programs. Both countries' public reforms have been particularly influenced by the New Public Management paradigm, which is characterized by an emphasis on business-like management, efficiency, decentralization, and input-output control (Pollitt & Bouckaert, 2017).

Germany's public sector administration is influenced by the characteristics of the *Rechtsstaat* in which the state is viewed as a "central integrating force within society, and its focal concerns are with the preparation, promulgation, and enforcement of laws." (Pollitt & Bouckaert, 2017, p. 61) Welfare policy is enshrined in extensive legal forms. Recently, neo-Weberian reforms to public administration have incorporated both elements of classic Weberian bureaucracy, building on a distinct public administration culture, and democratic reform aspects, such as increased citizen consultation. Though the state's role has been strengthened somewhat as a provider of solutions to social problems, there is also greater emphasis placed on public satisfaction with services. A results-driven culture, as opposed to traditional concerns about procedure, has become more salient (Pollitt & Bouckaert, 2017). Overall, the German trajectory is one towards piecemeal implementation of increased, though often disjointed, accountability and measurement systems in the public sector (Jantz, Christensen, & Lægreid, 2015).

In the United States, the Public Interest model of public administration is posited by Pollitt and Bouckaert (2017) to be the dominant paradigm. The role of public administration is perceived as that of a "fair and trusted referee," ensuring "independence of the play of sectional interests" (Pollitt & Bouckaert, 2017, p. 61). In the United States, public administration is considered accountable to a wide range of political and economic interests including businesses and the public at-large. There is an emphasis on market-like values, such as efficiency, competition, and innovation. Some reforms have been implemented, especially those aimed at greater transparency and short-term savings (Pollitt & Bouckaert, 2017). The public sector's relationship with nonprofits has emphasized these aspects for some time (Salamon, 1999; Suárez, 2010). However, similarly to Germany, a generalized and coherent reform paradigm has been lacking. This may in part be due to the federal division of power present in both countries (Pollitt & Bouckaert, 2017).

Nonprofit context: Origins of civil society

The specific comparative institutional origins of the nonprofit sector is of primary concern in ascertaining differences in evaluation practices in HSNs. Social Origins Theory (e.g. Salamon & Anheier, 1998; Salamon et al., 2017) offers an explicitly institutional comparative framework for the nonprofit sector. It formulates an empirically based response to theories of government and market failure (Weisbrod, 1977), supply side (James, 1987) trust based (Krashinsky, 1986), welfare state (Hicks & Swank, 1992), and interdependence theories (Salamon, 1995). According to Salamon and colleagues (1998; 2017), institutional patterns in the social economy are located within complex, historically contingent backdrops that explain differences in size, funding type, and other key sectoral characteristics. Casey's (2016b) Cultural Frames typology offers a similar basis of comparison for both countries.

Germany's nonprofit sector is characterized by Salamon et al. (2017) as a "welfare partnership" (or "corporatist" by Salamon and Anheier [1998]). Based on the nonprofit sector's high levels of state funding and national gross domestic product (GDP), this typification relates to traditions of cooperation and negotiation among different groups in German (civil) society (e.g., organized labour, church, state). These institutionalized practices were and still are often coordinated and directly funded by the state. The result is a focus on the provision of services, rather than advocacy in opposition to state policy (Salamon et al., 2017). This type of close relationship between the welfare regime and third sector may imply less formal and systemic evaluation practices.

The US third sector is described by Salamon et al. (2017), similarly to Esping-Andersen (1990), as liberal with some welfare partnership elements. The nonprofit economy comprises a relatively large share of GDP but receives relatively low levels of state funding. The liberal typology is characterized historically by a lack of direct social conflicts between economic classes leading to societal conditions that remain largely unchallenged at an institutional level. Human service nonprofit organizations are viewed as a private solution to collective societal problems, set in the frame of marketized service providers that nonetheless receive the majority of their funding from public sources (Salamon et al., 2017). The sector has a stronger reliance on the historic and present largesse of philanthropically minded upper-class and middle-class social groups, coupled with the ethos of personal responsibility (Eikenberry, 2015). These factors could potentially result in increased transparency and quantitative impact measurement.

Nonprofit evaluation regimes?

Although there is a paucity of (comparative) empirical research on internal evaluation practice in HSNs in Germany and the United States, limited applications to this management area can be found in the comparative theories described in the previous subsections. Frequent use of quantitative data measurement and extensive quality dialogue are congruent with Public Value concepts in US public administration and liberal, market-economic orientations in the welfare regime (Esping-Andersen, 1990; Pollitt & Bouckaert, 2017). In these contexts, managers may see a need to legitimize nonprofit practices through clearly recognizable scripts that are easily understood by all potential stakeholders, especially funders. An emphasis on individualism and entrepreneurialism, coupled with transparency, is a strong thread in comparative theories of the US nonprofit institutional environment (Casey, 2016b; Salamon et al., 2017). Broader cultural values likely undergird this paradigm. United States management is thought to be more performance oriented and less motivated by collectivism (House, Hanges, Javidan, Dorfman, & Gupta, 2004).

German management is seen to be focused on (post-)corporatist and sometimes cooperative welfare state-nonprofit relations (Bode, 2011). This may preclude rigorous quantitative evaluation practices since accountability is diffused throughout the entire social policy apparatus (Anheier et al., 2013; Esping-Andersen, 1990). Evaluation, particularly in the nonprofit sector, is driven primarily by political concerns (Stockmann & Meyer, 2020). Indeed, this diffusion of responsibility may lead to ritualized uses of monitoring and evaluation in HSNs (Bode, 2019; Bode & Turba, 2020). While neo-Weberian reforms in Germany may have increased the pressure to implement public facing quality measures, the path dependent, traditionalist tendencies of the nonprofit sector likely outweigh these changes

when compared with US nonprofits (Pollitt & Bouckaert, 2017; Salamon et al., 2017). When the German nonprofit sector partakes in a welfare partnership with the state, trust may be higher, and the pressure to evaluate impact stringently, rationally, and transparently may be lower.

METHODS

This study employed a mixed methods design, using both quantitative and qualitative data to triangulate findings on differences between nonprofit evaluation practice in Germany and the United States, as well as associations between these practices and the institutional environments. Comparative studies can relationally illuminate new aspects about two or more research subjects (Bazeley, 2013). Initially, the quantitative data analysis served the purpose of identifying statistically significant differences in evaluation practice between German and US HSN management. Quantitative and qualitative data were analyzed separately, then later integrated holistically to produce rich, triangulated findings, which imply forms of agential, institutional (as opposed to mechanical) causality (Howe, 2012). In this sense, the qualitative data analysis contextualizes what is indicated by the quantitative analysis (i.e., adding further substance to the cut-and-dry statistical analysis (Schoonenboom & Johnson, 2017). In turn, question-level quantitative analysis provides additional depth to the discussion of findings, since certain evaluation practices may see larger or smaller differences.

RESEARCH INSTRUMENT

A modified version of the World Management Survey (WMS), specifically the section on "monitoring," was employed as the primary research instrument. The WMS has been employed extensively in cross-national comparisons of management practices in a variety of relevant settings, including education (Bloom et al., 2015) and healthcare (McConnell, Hoffman, Quanbeck, & McCarty, 2009; Tsai, Jha, Gawande, Huckman, Bloom, & Sadun, 2015). The instrument has also found applications in comparative nonprofit management (Delfgaauw et al., 2011; Keller, 2011). For the purposes of this study, the author further modified Delfgaauw et al.'s (2011) nonprofit WMS interview guide, which was based on the work of Bloom and Van Reenen (2007). As Delfgaauw and colleagues' work related to the field of foster care placement, changes were made to accommodate the study of the YCAS nonprofit field. Additional cosmetic changes were made to increase accessibility of the survey language for managers in the United States and Germany.

The "monitoring" section of the WMS instrument puts forward a range of questions about the collection and use of outcome data, and how this data is used to improve organizational and individual performance. This can usefully be equated to evaluation as a management practice in nonprofit organizations (see appendix). However, it should be noted for the sake of conceptual clarity that monitoring and evaluation differ in important ways. Monitoring (i.e., the measurement of predefined outcomes) is a component of evaluation (i.e., an appraisal of whether desired impact has been achieved (Liket, 2017). In the results and discussion sections, the term evaluation will refer to both practices, while monitoring will be used when referring specifically to the more limited aspect of data collection for output control.

The WMS instrument operationalizes management on a normative scale (Van Reenen & Bloom, 2007). At first, this approach may seem to undercut the stated perspective of the study. World Management Survey criteria for "good" management stand for a variety of modern business management practices, especially those associated with Lean Management (Bloom & Van Reenen, 2010). Higher scores can thus be seen as a proxy for the salience of business-like management. In particular, the WMS conceptualizes state-of-the art "monitoring" practice as implementation of formal, ongoing, consequence driven evaluation (Delfgaauw et al., 2011; Van Reenen & Bloom, 2007). Importantly, many of these practices have found currency in the HSNs through marketization (Sandberg et al., 2019), the McDonaldization of service delivery (Dustin, 2016), and the rise of audit cultures (Moxham, 2010; Shore & Wright, 2015). The WMS provides a standardized reference point to dig deeper into the evaluation practices of nonprofit managers, allowing for standardized comparison and contextualization of its results.

SAMPLING AND DATA COLLECTION

Sampling was conducted using a mixture of probability and purposive techniques (Teddlie & Yu, 2007). First, a random geographic sampling of states was conducted to ensure variety among YCAS nonprofits. States in both countries in immediate proximity to one another were eliminated. Subsequently, a selective sample of managers was contacted by email. These managers worked at medium and large nonprofits in the states selected that operated in the YCAS field. Guidestar, other relevant industry databases, and the websites of nonprofit conglomerates were used as a reference for finding participant managers. Sampling was restricted to managers who had direct contact with front line staff and are responsible for one or more YCAS program. The geographic sample, the size of HSNs, and limitations to managers' duties assured a broad level of comparability while still allowing for some variety in terms of organizational practices and environmental conditions. Ethics approval was obtained in advance using the appropriate procedures for graduate thesis research at the author's institution of study, a large public university in Germany. All interviewees gave oral consent to take part in the study and be recorded in advance of the interview.

In total, 32 telephone interviews were conducted and recorded (16 in Germany, anonymized as DE1-DE16; 16 in the United States, anonymized as US1-US16). Interview questions from the "monitoring" WMS section (five question areas, see appendix) were analyzed for this study. The items were scored using a standardized rubric on a scale of one to five, with five being the score closest to ideal for-profit management practices identified by Van Reenen and Bloom (2007), and thus closest to business-like evaluation practices.

A digital memo and other handwritten notes were produced for each recorded interview. The purpose of these memos was to summarize the most important management practices addressed in each question, along with the WMS score for each question. Sections of each interview in which interview partners related reasons for using specific management practices were transcribed. All data was anonymized and stored securely to protect participants' privacy.

DATA ANALYSIS

Recorded answers to each area of the "monitoring" section were first scored by the author using

the standardized rubric to ascertain quantitative data for each HSN (see appendix). In addition to the average scores and standard deviations for the "monitoring" management area, an independent sample, two-tailed Student's *t*-test was conducted to test for statistical significance, complimented by reporting the effect size using Cohen's *d*. Student's *t*-test was selected due to the small sample size of the study (Nishishiba, Jones, & Kraner, 2014). The outputs from this test are available in the appendix.

Description and validation of management practices from the interviews was analyzed qualitatively. This mixed methods approach provided added texture and depth to the discussion of management micropractices and their possible associations with institutional contexts. The qualitative data was later triangulated with the quantitative results to animate this discussion. The author utilized a systematic, open-ended codes-to-themes technique based on Bazeley (2013) to analyze the qualitative data. Managers' reasoning for evaluation practice, represented by short contextualized quotes, was first sorted into similar codes. Codes were continuously refined throughout the coding process via extensive reflective note taking, allowing for an inductive approach to the data. As key words and phrases in each code were selected, a comparative process across codes was conducted. Each code was thus further developed during this process by constant comparison to other key phrases. Codes were subsequently reanalyzed for contrasts and similarities between countries. Codes were then summarized and synthesized into themes. Themes represent a reduction that portrays shared elements of several codes, expressed by a specific phrase. Next, themes were split into country level manifestations based on the previous comparative analysis of codes. This allowed for an emphasis on both contrasts and similarities across the data between countries. All German interview quotes presented in this article were translated by the author.

FINDINGS
Quantitative
United States nonprofit managers (*n* = 16) scored an average 2.69 (standard deviation [SD] 0.55) and their German counterparts (*n* = 16) an average 2.24 (SD 0.41). Equal variances were assumed. The independent sample Student's *t*-test results indicate differences in evaluation practices are statistically significant ($t(30) = 2.48$, $p < .05$), showing a large effect size ($d = 0.87$). United States YCAS nonprofit managers are therefore more likely than German nonprofit managers to implement formal, ongoing, consequence-driven evaluation practices.

Qualitative findings
The qualitative data was sorted into five codes, which were in turn synthesized into two distinct themes with differing presentations in German and US nonprofits (see Table 1). Evaluation practices were validated by interview partners as *a means of demonstrating accountability to external bodies*. For managers in both countries, evaluation practices served an important function vis-à-vis funders and accrediting bodies. DE7 explained that their organization's evaluation practices were linked to premiums. US13 validated the "constant" evaluation practice of their organization by describing the unpredictability of external audits. Dichotomous reward and punishment were particularly apparent for the German nonprofit managers, who emphasized the contractual nature of evaluation practices, both in terms of (continued) funding distribution and accreditation. A further important aspect of

this theme is the notion of public-facing accountability. United States nonprofit managers, in particular, framed their evaluation practices by describing the organization's mission in providing a benefit to the local community. Other participants reasoned that their evaluation practices cut against implications of impropriety. US13 noted that "we put out annual reports that report on each [quality indicator], so none of it's a secret." Transparent evaluation allowed them to publicize the efficacy of their work and show public money put to good use.

Table 1: Evaluation practices in German and US nonprofits

	Germany	**USA**
Average	2.24	2.69
Standard Deviation	0.41	0.55
t (df)	2.48 (30)	
p-value	0.02	
Cohen's *d*	0.87	

Evaluation practices were further validated as *a means of internal affirmation, that is, as a method of demonstrating identity*. This was often described by HSN managers in both countries by emphasizing what made their management (and direct service work) unique. Broadly speaking, participants framed the special internal role of evaluation in confident or doubtful terms. A limited number of US managers expressed the former. Following descriptions of specialized software (e.g., Salesforce), data organizing tools (e.g., dashboards), and results dissemination forums (e.g., regular meetings with stakeholders, brainstorming sessions), these managers framed evaluation as an integral part of motivating and improving their work at all levels. US1 explained the development in their organization: "we tracked with an excel sheet for a long time. And it was fine for sending out numbers to funders . . . but it wasn't as good for monitoring our actual success and identifying any red flags for students." The development and implementation of specialized or expert management toolsets, often with the intention of reducing complexity, appeared to be integral to these efforts. US10 elaborated that "our program dashboard does track all of the key measures of each contract, because we have about 16 contracts and they each measure something different."

Most study participants expressed skepticism about the efficacy and axiology of evaluation practice. This doubt about evaluation practice was articulated as a unique facet of (sometimes challenging) internal affirmation. To state this theme differently, managers distanced themselves from mainstream, often prescribed evaluation practice as a means of demonstrating that the relational element of their work was salient. The tacit nature of human services was used as a frequent justification by German managers. DE5 explained that "we see it differently and more specifically [than the indicators suggest], breaking it down to the perspective of young people." DE13 took this logic one step further, arguing that front line work is "individual, regarding the relationships between participating youths, social workers, master tradespeople, and teachers. There are so many emotional topics. There are issues of education and personal development ['Bildung']. Saying that you can determine quality criteria is quite presumptuous." Managers in both countries also relayed

insecurities about using data in decision making processes, a measure used in the WMS survey. US9 maintained that "I would say people tend towards the anecdotal. So it's much easier for people on the board to say 'this is what I've observed' as opposed to, say, the specific scientific data." Many other participants asserted that the wide variety of indicators and data management requirements made using evaluation results for organizational learning difficult.

DISCUSSION
Isomorphism and public sector funding

In previous studies, the role of public sector funding has been shown to produce strong isomorphic effects on nonprofit organizations to adopt particular practices (Hersberger-Langloh, Stühlinger, & Schnurbein, 2020; Suárez, 2010). A key aspect of the American welfare regime in international comparison is its use of means-testing to limit service provision (Esping-Andersen, 1990; Lynch, 2014). United States study participants noted that means-testing impacted evaluation practices in significant ways, in particular, US YCAS agencies acting as gatekeepers and promoters of services to potential service users. Monitoring of services played an important role in this dynamic, with compliance with the norms of state funding being a key issue. Human service nonprofit organizations collected extensive demographic data and tailored services to attract service users who would bring in funding to the non-profit agencies to maintain financial viability, strongly implying that that HSNs are tacitly encouraged to select service users with less complex problems ("creaming"), based on the potential for positive impact (Considine, Lewis, & O'Sullivan, 2015; Greer, Schulte, & Symon, 2018). This presents an example of vendorism, that is, a distortion of the values-based mission of a nonprofit organization in order to pursue revenue (Salamon, 1987).

German managers frequently validated their own evaluation practices through inflexible public sector contracts. Question-level answers from the WMS survey indicate that these practices are less rigorous in Germany and often oriented toward success or failure in meeting prescribed targets. This was often explained utilizing a legalistic vocabulary, mentioning "contracts," "service performance description," and "contract award documents." Requirements were often fixed for several years, leading to limited agency and flexibility in operation and revision of evaluation practices. Whereas organizational flexibility is identified by Salamon (1987) as a key advantage of nonprofit service provision, this rigid contracting dynamic appears to mirror principles in traditional German public administration that emphasize rules-based order and the "proper fulfillment of regulatory functions" (Pollitt & Bouckaert, 2017, p. 298). Though not apparent in the data from this study, scholarship on the role of HSN accreditation agencies in the United States indicates that they may play a similarly (normative or coercive) isomorphic role in some service areas (Ford & Andersson, 2021; Lee & Clerkin, 2017).

This coercive isomorphism from public sector bodies in Germany contrasts with attempts by US managers to assert a unique identity for organizational and managerial practice. Some US managers described use of specialized software in these terms, calling attention to potential benefits of technology to systematize service user and employee outcomes. This behaviour can be interpreted as a form of mimetic isomorphism mirroring tech-forward private sector business management and similar nonprofit management practices (Lee & Clerkin, 2017; Zorn, Flanagin, & Shoham, 2011). Mimetic isomorphism is not accomplished by simply grafting ideas and practices onto a new organ-

ization. As US1 explained, "we really are expecting our staff to report on outcomes and identify students who are making progress for outcomes. [The software is] good for monitoring our actual success and identifying any red flags for students." Here, the US manager has stripped the software of its original identifiers from the private sector, re-contextualizing and re-narrating its applications in the YCAS field. This presents a form of knowledge and practice translation (Nielsen, Mathiassen, & Newell, 2014; Wedlin & Sahlin, 2017).

Constructs of accountability
The rise of audit culture and redefinition of accountability are vitally important recent develop-ments for HSNs (Anheier, 2009; Benjamin, 2021; Mourey, 2021; Shore & Wright, 2015). Findings from this international comparative project indicate differing understandings of accountability, which may be associated with the respective institutional environments. For example, German managers seemed concerned with bipolar, vertically structured accountability to public sector funders (often the Federal Employment Office—*Agentur für Arbeit*). Benchmarking was often con-ducted by the funders, with only nominal participation by managers themselves. In other words, this accountability was instrumental in that it served functional and inflexible aims to legitimize organizational practice to an outside group (Knutsen & Brower, 2010). In contrast, some US man-agers, most notably those who received significant funding by donations, explained accountability in multiple dimensions, including community groups and other nonprofits within a local community of care. One US manager's explanation of accountability stood out, demonstrating near complete public-facing transparency. The participant offered to answer questions of curious stakeholders and small donors. Knutsen and Brower (2010, p. 597) characterize this form of "expressive" com-munity based nonprofit accountability as "organizations' self-perception of their community roles, often altruistic and value driven."

For US managers in the study, accountability was portrayed in a manner that implied the importance of a unique organizational value proposition to the public and, sometimes, community stakeholders. This reflects Young's (2010) argument that US nonprofits have been forced to rethink the way they run, as businesses and service providers. Dart (2004) similarly proposes that results-oriented man-agement is a key area of business-like management practice in HSNs. The specific practices de-scribed by study participants can be viewed within broader institutional paradigms such as marketization (Salamon et al., 2017; Sandberg et al., 2019).

The data confirm that funders influenced and determined evaluation practices for some US nonprofit managers in the research project, yet many also emphasized the role of communities and philan-thropists. These contrasting responses echo Benjamin's (2008) conceptualization of nonprofit eval-uation efforts as a risky undertaking between durable and instrumental accountabilities in relationship building. Heterogenous, complex, and often fraught funding requirements, as well as the mission-based commitments of some nonprofits, may combine to create a greater focus on the immediate demands of stakeholders for the study participants. US2 explain that their organization is "held accountable for [goals] by outside groups because in applying for a grant we'll say, 'we're going to serve 200 youth in the course of a year.' If you don't serve 200 youth, you have to explain why." United States managers seemed to subscribe more fully to the notion that they needed to

demonstrate public accountability through rationalized and transparent methods. This idea echoes some aspects of Ebrahim's (2009) technocratic and adaptive accountability regimes.

In contrast, German nonprofit managers structured accountability in a bifurcated manner, possibly relying on the constructs common in the public sector. While HSNs are, nominally speaking, able to operate freely in terms of their service provision, they remain interdependent on the state for their legal framework, funding, and policy direction (Archambault, Priller, & Zimmer, 2014). Strachwitz (2010) and Pollitt and Bouckaert (2017) contend that the conservative framework informing public sector management, with its emphasis on corporatist stability, has largely remained intact. The findings indicate that demonstrable public-facing accountability remained absent from most attempts to monitor performance.

Instead, as Strachwitz (2010) argues, responsibility for results is vertically structured between the nonprofit and state funders. Using Ebrahim's (2009) accountability typology, these attributes of German HSN evaluation practice can be characterized as coercive. As DE5 explained unequivocally, "our success is defined quite narrowly by the contracting party, which means there is only success and failure. Either a service user starts employment or they don't." Skepticism of the efficacy of such zero-sum quantitative measurement, especially in conjunction with individual rewards and punishment for employees, was common among German managers. This suggest that although the current logic of the German public sector vests trust in the expertise of the state apparatus to shape society, there exists an inherent criticism of "the depredations of globalized neo-liberalism" (Pollitt & Bouckaert, 2017, p. 121).

Limited translations of global(ized) New Public Management have impacted the German nonprofit sector, leading to potential hybrid forms of management practice (Grunwald & Roß, 2017). Increased control by funders, over the methods used to monitor outputs and the output targets themselves, combined with an emphasis on rules and procedures and the primacy of legalistic state dominance (Pollitt & Bouckaert, 2017) may have led to German managers scoring lower on the quantitative portion of the study, which rewards rigorous, agile practice over consistency. The lower standard deviation for German managers could suggest bureaucratic consistency. When ultimate responsibility for evaluation lies elsewhere, there is little reason for HSNs to use precious resources to develop their own sophisticated evaluation capacities (Despard, 2016). German managers' consistent but simple evaluation efforts might be compared with ritualized practice, which has been noted in other scholarship on German HSN management (Bode, 2019). While this study's results point to these aspects regarding limited evaluation capacity being more apparent in German HSN management practice, relevant US scholarship indicates serious concerns with regard to evaluation capacity in the US context as well, including lack of resources, internal and external buy-in, and necessary expertise (Carman & Fredericks, 2010; Mitchell & Berlan, 2018).

CONCLUSION

German and US HSN managers practice evaluation differently. United States managers evaluate and use evaluation more similar to for-profit organizations than their German counterparts. German managers' evaluation practices are less business-like and may be closer to prevailing public sector

evaluation. Building on these results and the discussion, two arenas for further critical scholarship will be explained.

The first area concerns financial and service provision accountability between the state and HSNs. The complex, multifaceted relationship between state and third sector informs evaluation practice in nonprofits, as demonstrated by the study's findings and other scholarship (e.g., Hersberger-Langloh et al., 2020; Moxham, 2010). Especially in countries where welfare governance has shifted towards contracting (marketized) services, the state has been essential in assuring the survival and enlarged role of nonprofit organizations. Reliance on state funding, when it is stable and combined with consistent capacity building, may be preferable to private philanthropy in terms of its ethical implications and increased professionalism driven by state contracts (Salamon, 1987; Salamon & Toepler, 2015; Suárez, 2010). However, HSNs run the risk of becoming an extension of state control apparatuses, sacrificing the historically important democratic role of nonprofits as advocacy organizations and drivers of social justice agendas (Almog-Bar & Schmid, 2013; Eikenberry & Kluver, 2004; Guo, 2007). This study analyzes the evaluation practices in HSNs providing career assistance to youth funded by a range of sources. The results imply that no source of funding is better or worse, but rather that the dynamics of demonstrating accountability to funders inform evaluation practice differently. The complex and sometimes contradictory "translations" of evaluation practice in the HSN sector require additional inquiry.

Especially within internationally comparative scholarship, issues of nonprofit-state relations remain underexamined, especially those emerging from Anglo-American discourse, which are not necessarily generalizable (Schubert & Boenigk, 2019). What types of evaluation are practiced in hybridized HSNs embedded in different institutional contexts (Pache & Santos, 2013)? How do local differences in neoliberal public sector reforms translate comparatively in nonprofit management settings (Baines et al., 2014)? How can service users be involved more meaningfully in evaluation and helping processes generally, in order to fulfill the third sector's democratizing aspirations (Benjamin, 2021; Benjamin & Campbell, 2015)? These topics, and in particular the institutional processes that drive them, present intriguing areas of future research.

Power and agency in comparative HSN management presents a second key area of scholarship that is both undertheorized and under researched (Coule, Dodge, & Eikenberry, 2020). More research into the practical, qualitative dimensions of agency and power in HSNs is needed to understand their micro and macro practical implications as they apply to evaluation. This undertaking might include a phenomenological focus on the *Habitus* of nonprofit managers as they co-construct contested paradigms, such as accountability (Albrecht, 2018) and professionalism (King, 2017). Exploring burgeoning HSNs from an institutional perspective could be another avenue to explore in future evaluation research. New organizational forms, technologies, and relational formats represent an evocative, multilevel example of visible agency (Nicholls, 2010; Tracey, Phillips, & Jarvis, 2011). Related key issues might include: How do organizational actors in the nonprofit sector experience their own power and (dis)empowerment with regard to shifting and contested institutional templates of accountability? (Åberg, 2015; King, 2017). Considering that one key aspect of evaluation is its axiological component (Mertens, 2016), how are management and organizational structures reframed within specific institutional contexts to embody social justice values? (Sandberg & Elliott, 2019).

The findings presented in this article indicate significant and meaningful differences between HSN evaluation practices in Germany and the United States. The qualitative analysis yielded varying reasoning strategies for management practices, especially in managers' relationships to the public sector and to constructs of accountability. These managements are translated across sectors within each country and internationally by embedded actors, aided by cultural, social, political, and economic frames of reference. Although the results of this study point to US nonprofit evaluation practices being more business-like, and German nonprofit evaluation being similar to that of public administration, this offers only a limited snapshot of HSN management practices in both countries. Further research of this nature would increase understanding of the important and complex relationships between institutional structures at the macro and mezzo levels, and nonprofit management micro-practice.

REFERENCES

Åberg, P. (2015). Myths and traditions as constraints or resources? Path dependency and decoupling strategies among civil society organizations. *Journal of Civil Society, 11*(1), 19–38. doi:10.1080/17448689.2015.1009695

Abramovitz, M., & Zelnick, J. (2018). The logic of the market versus the logic of social work: Whither the welfare state? *Social Work & Society, 16*(2), 1. URL: http://nbn-resolving.de/urn:nbn:de:hbz:464-sws-1503 [October 02, 2022].

Albrecht, K. (2018). Institutional logics and accountability: Advancing an integrated framework in nonprofit–public partnerships. *Journal of Public and Nonprofit Affairs, 4*(3), 284–305. doi:10.20899/jpna.4.3.284-305

Almog-Bar, M., & Schmid, H. (2013). Advocacy activities of nonprofit human service organizations. *Nonprofit and Voluntary Sector Quarterly, 43*(1), 11–35. doi:10.1177/0899764013483212

Anheier, H.K. (2009). What kind of nonprofit sector, what kind of society?: Comparative policy reflections. *The American Behavioral Scientist, 52*(7), 1082–1094. doi:10.1177/0002764208327676

Anheier, H.K. (2010). Comparing third sector policy: A postscript and six theses. In B. Gidron & M. Bar (Eds.), *Policy initiatives towards the third sector in international perspective* (pp. 213–216). New York, NY: Springer. doi:10.1007/978-1-4419-1259-6

Anheier, H.K., Hass, R., & Beller, A. (2013). Accountability and transparency in the german nonprofit sector: A paradox? *International review of public administration, 18*(3), 69–84. doi:10.1080/12294659.2013.10805264

Anheier, H.K., & Toepler, S. (2020). Nonprofit Management. In H.K. Anheier & S. Toepler (Eds.), *The Routledge companion to nonprofit management* (First ed.). New York, NY: Routledge. doi:10.4324/9781315181011

Archambault, E., Priller, E., & Zimmer, A. (2014). European civil societies compared: Typically German—typically French? *Voluntas, 25*(2), 514–537. Manchester, UK. doi:10.1007/s11266-013-9349-6

Arvidson, M., & Lyon, F. (2014). Social impact measurement and non-profit organisations: Compliance, resistance, and promotion. *Voluntas 25*(4), 869–886. Manchester, UK. doi:10.1007/s11266-013-9373-6

Baines, D., Charlesworth, S., & Cunningham, I. (2014). Fragmented outcomes: International comparisons of gender, managerialism and union strategies in the nonprofit sector. *Journal of Industrial Relations, 56*(1), 24–42. doi:10.1177/0022185613498664

Bazeley, P. (2013). *Qualitative data analysis: practical strategies*. London,UK: SAGE.

Benjamin, L.M. (2008). Bearing more risk for results: Performance accountability and nonprofit relational work. *Administration & Society, 39*(8), 959–983. doi:10.1177/0095399707309357

Benjamin, L.M. (2021). Beyond programs: Toward a fuller picture of beneficiaries in nonprofit evaluation. In P. Dahler-Larsen (Ed.), *A research agenda for evaluation*. Cheltenham, UK: Edward Elgar Publishing. doi:10.4337/9781839101083.00011

Benjamin, L.M., & Campbell, D.A. (2020). Evaluation and performance measurement. In H.K. Anheier & S. Toepler (Eds.), *The Routledge companion to nonprofit management* (First ed.). New York, NY: Routledge. doi:10.4324/9781315181011

Benjamin, L.M., & Campbell, D.C. (2015). Nonprofit performance: Accounting for the agency of clients. *Nonprofit and Voluntary Sector Quarterly, 44*(5), 988–1006. doi:10.1177/0899764014551987

Bloom, N., Lemos, R., Sadun, R., & Van Reenen, J. (2015). Does management matter in schools? *The Economic journal (London), 125*(584), 647–674. doi:10.1111/ecoj.12267

Bloom, N., & Van Reenen, J. (2010). Why do management practices differ across firms and countries? *Journal of Economic Perspectives, 24*(1), 203–224. doi:10.1257/jep.24.1.203

Bloom, N., & Van Reenen, J. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly journal of economics, 122*(4), 1351-1408. doi:10.1162/qjec.2007.122.4.1351

Bode, I. (2011). Creeping marketization and post-corporatist governance: The transformation of state-nonprofit relations in Continental Europe. In S. Phillips & S.R. Smith (Eds.), *Governance and regulation in the third sector: international perspectives* (pp. 115–141). New York: Routledge. doi:10.4324/9780203835074

Bode, I. (2019). Let's count and manage—and forget the rest. Understanding numeric rationalization in human service provision. *Historical social research (Köln), 44*(2(168)), 131–154.

Bode, I., & Turba, H. (2020). Schizophrenic sensemaking as (non-)response to institutional ambiguity: The case of the German child protection industry. *Human service organizations, management, leadership & governance, 44*(1), 9–31. doi:10.1080/23303131.2019.1696907

Bromley, P., & Meyer, J.W. (2017). "They are all organizations": The cultural roots of blurring between the nonprofit, business, and government sectors. *Administration & Society, 49*(7), 939–966. doi:10.1177/0095399714548268

Carman, J.G., & Fredericks, K.A. (2010). Evaluation capacity and nonprofit organizations: Is the glass half-empty or half-full? *The American journal of evaluation, 31*(1), 84–104. doi:10.1177/1098214009352361

Carnochan, S., Samples, M., Myers, M., & Austin, M.J. (2014). Performance measurement challenges in nonprofit human service organizations. *Nonprofit and Voluntary Sector Quarterly, 43*(6), 1014–1032. doi:10.1177/0899764013508009

Casey, J. (2016a). A bridge between US and international perspectives: Deliberate relations between governments and nonprofits. In J. Brothers (Ed.), *Rebalancing public partnership: Innovative practice between government and nonprofits from around the world* (pp. 93–107). Milton Park, UK: Taylor and Francis. doi:10.4324/9781315603896

Casey, J. (2016b). Comparing nonprofit sectors around the world: What do we know and how do we know it? *Journal of Nonprofit Education and Leadership, 6*(3). doi:10.18666/JNEL-2016-V6-I3-7583

Considine, M., Lewis, J.M., & O'Sullivan, S. (2015). Australia's competitive contract model. In M. Considine, J.M. Lewis, & S. O'Sullivan (Eds.), *Getting welfare to work: street-level governance in Australia, the UK, and the Netherlands*. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780198743705.001.0001

Coule, T.M., Dodge, J., & Eikenberry, A.M. (2020). Toward a typology of critical nonprofit studies: A literature review. *Nonprofit and Voluntary Sector Quarterly*, 89976402091980. doi:10.1177/0899764020919807

Dart, R. (2004). Being "business-like" in a nonprofit organization: A grounded and inductive typology. *Nonprofit and Voluntary Sector Quarterly, 33*(2), 290–310. doi:10.1177/0899764004263522

Delfgaauw, J., Dur, R., Propper, C., & Smith, S.L. (2011). Management practices: Are not for profits different? *CMPO Working Papers, 11*. URL: www.cepr.org/pubs/dps/DP8498.asp [September 29 2022].

Despard, M.R. (2016). Challenges in implementing evidence-based practices and programs in nonprofit human service organizations. *Journal of evidence-informed social work, 13*(6), 505–522. doi:10.1080/23761407.2015.1086719

Dustin, D. (2016). *The McDonaldization of social work*. Milton Park, UK: Taylor and Francis. doi:10.4324/9781315555676

Ebrahim, A. (2009). Placing the normative logics of accountability in "thick" perspective. *The American behavioral scientist (Beverly Hills), 52*(6), 885–904. doi:10.1177/0002764208327664

Ebrahim, A. (2019). *Measuring social change: Performance and accountability in a complex world*. Stanford, CA: Stanford Business Books. URL: https://go.exlibris.link/GDSzYnwh [October 01 2022].

Eikenberry, A.M. (2015). Nonprofit organizations, philanthropy, and democracy in the United States. In *Democracy and Public Administration* (pp. 169–193). New York, NY: Routledge.

Eikenberry, A.M., & Kluver, J.D. (2004). The marketization of the nonprofit sector: Civil society at risk? *Public Administration Review, 64*(2), 132–140. doi:10.1111/j.1540-6210.2004.00355.x

Emirbayer, M., & Mische, A. (1998). What Is agency? *The American Journal of Sociology, 103*(4), 962–1023. doi:10.1086/231294

Esping-Andersen, G. (1990). *The three worlds of welfare capitalism*. Cambridge, UK: Polity Press.

Farmer, R.N., & Richman, B.M. (1964). A model for research in comparative management. *California management review, 7*(2), 55–68. doi:10.2307/41162165

Ford, M.R., & Andersson, F.O. (2021). Sources of isomorphism in the Milwaukee voucher school sector. *Public policy and administration, 36*(1), 89–114. doi:10.1177/0952076719838298

Greenwood, R., Oliver, C., Lawrence, T.B., & Meyer, R.E. (2017). *The SAGE handbook of organizational institutionalism* (2nd ed.). London, UK: Sage Reference.

Greer, I., Schulte, L., & Symon, G. (2018). Creaming and parking in marketized employment services: An Anglo-German comparison. *Human relations (New York), 71*(11), 1427–1453. doi:10.1177/0018726717745958

Grunwald, K., & Roß, P.-S. (2017). Sozialmanagement als Steuerung hybrider sozialwirtschaftlicher Organisationen. In A. Wöhrle, A. Fritze, T. Prinz, & G. Schwarz (Eds.), *Sozialmanagement — Eine Zwischenbilanz* (pp. 171–184). Wiesbaden, Germany: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-658-14896-6_11

Guo, C. (2007). When government becomes the principal philanthropist: The effects of public funding on patterns of nonprofit governance. *Public Administration Review, 67*(3), 458–473. doi:10.1111/j.1540-6210.2007.00729.x

Hargrave, T.J., & Van De Ven, A.H. (2009). Institutional work: actors and agency in institutional studies of organizations. In T.B. Lawrence, R. Suddaby, & B. Leca (Eds.), *Institutional work: Actors and agency in institutional studies of organizations*. New York: Cambridge University Press. doi:10.1017/CBO9780511596605

Hasenfeld, Y. (2015). What exactly is human services management? *Human Service Organizations Management, Leadership & Governance, 39*(1), 1–5. doi:10.1080/23303131.2015.1007773

Hasenfeld, Y., & Garrow, E.E. (2012). Nonprofit human-service organizations, social rights, and advocacy in a neoliberal welfare state. *The Social service review (Chicago), 86*(2), 295–322. doi:10.1086/666391

Henriksen, L.S., Smith, S.R., & Zimmer, A. (2012). At the eve of convergence? Transformations of social service provision in Denmark, Germany, and the United States. *Voluntas (Manchester, England), 23*(2), 458–501. doi:10.1007/s11266-011-9221-5

Hersberger-Langloh, S.E., Stühlinger, S., & Schnurbein, G. (2020). Institutional isomorphism and nonprofit managerialism: For better or worse? *Nonprofit management & leadership, 31*, 461–480. doi:10.1002/nml.21441

Hicks, A.M., & Swank, D.H. (1992). Politics, institutions, and welfare spending in industrialized democracies, 1960–1982. *The American political science review, 86*(3), 658–674. doi:10.2307/1964129

Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, CA: Sage Publications.

House, R.J., Hanges, P.J., Javidan, M., Dorfman, P.W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. London, UK: Sage.

House, R.J., Dorfman, P.W., Javidan, M., Hanges, P.J., & Sully de Luque, M.F. (2014). *Strategic leadership across cultures: The GLOBE study of CEO leadership behavior and effectiveness in 24 countries*. London, UK: Sage.

Howe, K.R. (2012). Mixed methods, triangulation, and causal explanation. *Journal of Mixed Methods Research, 6*(2), 89–96. doi:10.1177/1558689812437187

James, E. (1987). The nonprofit sector in comparative perspective. In W.W. Powell (Ed.), *The nonprofit sector: A research handbook* (pp. 397–415). New Haven, CT: Yale University Press.

Jantz, B., Christensen, T., & Lægreid, P. (2015). Performance management and accountability: The welfare administration reform in Norway and Germany. *International Journal of Public Administration, 38*(13–14), 947–959. doi:10.1080/01900692.2015.1069838

Keller, G.F. (2011). Comparing the affects of management practices on organizational performance between for-profit and not-for-profit corporations in Southeast Wisconsin. *Journal of business & economics research, 9*(3), 29. doi:10.19030/jber.v9i3.4127

King, D. (2017). Becoming business-like: Governing the nonprofit professional. *Nonprofit and Voluntary Sector Quarterly, 46*(2), 241–260. doi:10.1177/0899764016663321

Knutsen, W.L. (2012). Adapted institutional logics of contemporary nonprofit organizations. *Administration & Society, 44*(8), 985–1013. doi:10.1177/0095399712438371

Knutsen, W.L., & Brower, R.S. (2010). Managing expressive and instrumental accountabilities in nonprofit and voluntary organizations: A qualitative investigation. *Nonprofit and Voluntary Sector Quarterly, 39*(4), 588–610. doi:10.1177/0899764009359943

Kopaneva, I.M., & Cheney, G. (2019). Organizational identity formation in alternative organizations: A study of three benefit corporations. *Management Communication Quarterly, 33*(4), 484–511. doi:10.1177/0893318919858684

Krashinsky, M. (1986). Transactions costs and a theory of the non-profit organization. In S. Rose-Ackerman (Ed.), *The economics of nonprofit institution* (pp. 114–132). Oxford, UK: Oxford University Press.

Lee, C., & Clerkin, R.M. (2017). Exploring the use of outcome measures in human service nonprofits: Combining agency, institutional, and organizational capacity perspectives. *Public Performance & Management Review, 40*(3), 601–624. doi:10.1080/15309576.2017.1295872

Liket, K. (2017). Challenges for policy-makers: Accountability and cost-effectiveness. In B. Greve (Ed.), *Handbook of social policy evaluation* (pp. 183–203). Cheltenham, UK: Edward Elgar Publishing. doi:10.4337/9781785363245.00017

Lynch, J. (2014). A cross-national perspective on the American welfare state. In D. Béland, K.J. Morgan, & C. Howard (Eds.), *Oxford Handbook of U.S. Social Policy* (Vol. 1, pp. 112-130). New York, NY: Oxford University Press. doi:10.1093/oxfordhb/9780199838509.013.023

Maier, F., Meyer, M., & Steinbereithner, M. (2016). Nonprofit organizations becoming business-like: A systematic review. *Nonprofit and Voluntary Sector Quarterly, 45*(1), 64–86. doi:10.1177/0899764014561796

McConnell, K.J., Hoffman, K.A., Quanbeck, A., & McCarty, D. (2009). Management practices in substance abuse treatment programs. *Journal of Substance Abuse Treatment, 37*(1), 79–89. doi:10.1016/j.jsat.2008.11.002

Mendell, M. (2017). Reflections on the evolving landscape of social enterprise in North America. *Policy and Society, 29*(3), 243–256. doi:10.1016/j.polsoc.2010.07.003

Mertens, D.M. (2016). Assumptions at the philosophical and programmatic levels in evaluation. *Evaluation and Program Planning, 59*, 102–108. doi:10.1016/j.evalprogplan.2016.05.010

Meyer, J.W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *The American Journal of Sociology, 83*(2), 340–363. doi:10.1086/226550

Mitchell, G.E., & Berlan, D. (2018). Evaluation in nonprofit organizations: An empirical analysis. *Public Performance & Management Review, 41*(2), 415–437. doi:10.1080/15309576.2017.1400985

Mourey, D. (2021). Twists and turns of the practice of accountability in the nonprofit sector: The thin line between accountancy and organizational learning. In G. Donnely-Cox, M. Meyer, & F. Wijkström (Eds.), *Research handbook on nonprofit governance* (pp. 391–408). Cheltenham, UK: Edward Elgar Publishing. doi:10.4337/9781788114912.00032

Moxham, C. (2010). Help or hindrance?: Examining the role of performance measurement in UK nonprofit organizations. *Public Performance & Management Review, 33*(3), 342–354. doi:10.2753/PMR1530-9576330302

Nicholls, A. (2010). The legitimacy of social entrepreneurship: Reflexive isomorphism in a pre–paradigmatic field. *Entrepreneurship Theory and Practice, 34*(4), 611–633. doi:10.1111/j.1540-6520.2010.00397.x

Nielsen, J.A., Mathiassen, L., & Newell, S. (2014). Theorization and translation in information technology institutionalization: Evidence from danish home care. *MIS Quarterly, 38*(1), 165–186. doi:10.25300/MISQ/2014/38.1.08

Nishishiba, M., Jones, M., & Kraner, M. (2014). *Research methods and statistics for public and nonprofit administrators: a practical guide*. London, UK: SAGE. doi:10.4135/9781544307763

Oehme, A. (2016). Teilhabegerechtigkeit und Inklusion als Handlungsorientierung der Jugendberufshilfe. *Sozial Extra, 40*(3), 28–32. doi:10.1007/s12054-016-0052-x

Pache, A.-C., & Santos, F. (2013). Inside the hybrid organization: Selective coupling as a response to competing institutional logics. *Academy of Management journal, 56*(4), 972–1001. doi:10.5465/amj.2011.0405

Pandey, S.K., & Johnson, J.M. (2019). Nonprofit management, public administration, and public policy: Separate, subset, or intersectional domains of Inquiry? *Public Performance & Management Review, 42*(1), 1–10. doi:10.1080/15309576.2018.1557382

Perrin, B. (2018). Accountabilities two solitudes and the questions it raises. In J.-E. Furubo & N. Stame (Eds.), *The evaluation enterprise: A critical view* (pp. 126–158). New York, NY: Routledge, Taylor & Francis Group. URL: https://go.exlibris.link/QWFcZPdZ [October 02 2022].

Pollitt, C., & Bouckaert, G. (2017). *Public management reform: A comparative analysis: into the age of austerity* (4th ed). Oxford, UK: Oxford University Press. URL: https://ebookcentral.proquest.com/lib/ubc/detail.action?docID=4891273 [October 01 2022].

Post, M., & Dodge, J. (2019). The promise of qualitative and participatory approaches to performance assessment: A critical perspective. In A.M. Eikenberry, R. Mirabella, & B. Sandberg (Eds.), *Reframing nonprofit organizations: Democracy, inclusion, and social change* (pp. 138–153). Irvine, CA: Melvin & Leigh.

Roy, M.J., Eikenberry, A.M., & Teasdale, S. (2021). The marketization of the third sector? Trends, impacts and implications. In G. Donnely-Cox, M. Meyer, & F. Wijkström (Eds.), *Research handbook on nonprofit governance* (pp. 371–390). Cheltenham, UK: Edward Elgar Publishing. doi:10.4337/9781788114912.00031

Ryan, P. (2001). The school-to-work transition: A cross-national perspective. *Journal of Economic Literature, 39*(1), 34–92. doi:10.1257/jel.39.1.34

Salamon, L.M. (1987). Of market failure, voluntary failure, and third-party government: Toward a theory of government-nonprofit relations in the modern welfare state. *Journal of Voluntary Action Research, 16*(1-2), 29–49. doi:10.1177/089976408701600104

Salamon, L.M. (1995). *Partners in public service: Government-nonprofit relations in the modern welfare state*. Baltimore, MC: Johns Hopkins University Press. URL: https://go.exlibris.link/k9Sww0HD [October 01 2022].

Salamon, L.M. (1999). The nonprofit sector at a crossroads: The case of America. *Voluntas (Manchester, England), 10*(1), 5–23. doi:10.1023/A:1021435602742

Salamon, L.M., & Anheier, H.K. (1998). Social origins of civil society: Explaining the nonprofit sector cross-nationally. *Voluntas, 9*(3), 213–248. Manchester, UK. doi:10.1023/A:1022058200985

Salamon, L.M., Sokolowski, S.W., & Haddock, M.A. (2017). *Explaining civil society development: A social origins approach*. Baltimore, MD: Johns Hopkins University Press. URL: http://ebookcentral.proquest.com/lib/ubc/detail.action?docID=4862708 [October 01 2022.

Salamon, L.M., & Toepler, S. (2015). Government–nonprofit cooperation: Anomaly or necessity? *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, 26*(6), 2155–2177. doi:10.1007/s11266-015-9651-6

Sandberg, B., & Elliott, E. (2019). Toward a care-centered approach for nonprofit management in a neoliberal era. *Administrative Theory & Praxis, 41*(3), 286–306. doi:10.1080/10841806.2019.1621661

Sandberg, B., Elliott, E., & Petchel, S. (2019). Investigating the marketization of the nonprofit sector: A comparative case study of two nonprofit organizations. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations, 31*(3), 494–510. doi:10.1007/s11266-019-00159-8

Sanders, M.L. (2012). Theorizing nonprofit organizations as contradictory enterprises. *Management Communication Quarterly, 26*(1), 179–185. doi:10.1177/0893318911423761

Savickas, M.L., & Savickas, S. (2019). A history of career counselling. In J.A. Athanasou & H.N. Perera (Eds.), *International Handbook of Career Guidance* (pp. 25–43). New York, NY: Springer International Publishing. doi:10.1007/978-3-030-25153-6_2

Schoonenboom, J., & Johnson, R.B. (2017). How to construct a mixed methods research design. *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 69*(Suppl 2), 107–131. https://doi.org/10.1007/s11577-017-0454-1

Schröder, M. (2013). *Integrating varieties of capitalism and welfare state research: A unified typology of capitalisms*. London, UK: Palgrave.

Schubert, P., & Boenigk, S. (2019). The nonprofit starvation cycle: Empirical evidence from a german context. *Nonprofit and Voluntary Sector Quarterly, 48*(3), 467–491. doi:10.1177/0899764018824669

Scott, W.R. (2017). Institutional theory: Onward and upward. In Greenwood, R., Lawrence T.B., & Meyer R.E. (Ed.), *The Sage handbook of organizational institutionalism* (2 ed.), pp. 853–869). SAGE Publications Ltd. doi:10.4135/9781526415066

Shore, C., & Wright, S. (2015). Governing by numbers: Audit culture, rankings and the new world order. *Social anthropology, 23*(1), 22–28. doi:10.1111/1469-8676.12098

Simon, H.A. (1979). Rational decision making in business organizations. *The American Economic Review, 69*(4), 493–513.

Skelcher, C., & Smith, S.R. (2015). Theorizing hybridity: Institutional logics, complex organizations, and actor identities: The case of nonprofits. *Public Administration, 93*(2), 433–448. doi:10.1111/padm.12105

Sozialgesetzbuch. (1990). *Achtes Buch. Kinder und Jugendhilfe*. URL: https://www.sozialgesetzbuch-sgb.de /sgbviii/1.html [October 10, 2022].

Starke, P., Obinger, H., & Castles, F.G. (2008). Convergence towards where: In what ways, if any, are welfare states becoming more similar? *Journal of European Public Policy, 15*(7), 975–1000. doi:10.1080/13501760802310397

Stockmann, R., & Meyer, W. (2020). Germany. In R. Stockmann, W. Meyer, & L. Taube (Eds.), *The Institutionalisation of Evaluation in Europe* (pp. 167–198). Springer International Publishing. doi:10.1007/978-3-030-32284-7_7

Strachwitz, R.G. (2010). Policy initiatives towards civil society in Germany. In B. Gidron & M. Bar (Eds.), *Policy initiatives towards the third sector in international perspective* (pp. 67–85). New York, NY: Springer.

Suárez, D.F. (2010). Collaboration and professionalization: The contours of public sector funding for nonprofit organizations. *Journal of Public Administration Research and Theory, 21*(2), 307–326. doi:10.1093/jpart/muq049

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research, 1*(1), 77–100. doi:10.1177/2345678906292430

Tracey, P., Phillips, N., & Jarvis, O. (2011). Bridging institutional entrepreneurship and the creation of new organizational forms: A multilevel model. *Organization Science (Providence, R.I.), 22*(1), 60–80. doi:10.1287/orsc.1090.0522

Tsai, T.C., Jha, A.K., Gawande, A.A., Huckman, R.S., Bloom, N., & Sadun, R. (2015). Hospital board and management practices are strongly related to hospital performance on clinical quality metrics. *Health Affairs, 34*(8), 1304–1311. doi:10.1377/hlthaff.2014.1282

Van Kersbergen, K., & Kremer, M. (2008). Conservatism and the welfare state: Intervening to preserve. In W. van Oorschot, M. Opielka, & B. Pfau-Effinger (Eds.), *Culture and welfare state: Values and social policy in comparative perspective* (pp. 71–88). Cheltenham, UK: Edward Elgar Publishing.

Wedlin, L., & Sahlin, K. (2017). The imitation and translation of management ideas. In R. Greenwood, C. Oliver, T.B. Lawrence, & R.E. Meyer (Eds.), *The Sage handbook of organizational institutionalism* (pp. 102–127). London, UK: Sage Publications. doi:10.4135/9781446280669.n5

Weisbrod, B.A. (1977). *The voluntary nonprofit sector: An economic analysis*. Lexington, MA: Lexington Books.

Young, D.R. (2010). Nonprofits and public policy in the United States: The evolution of accountability. In B. Gidron & M. Bar (Eds.), *Policy Initiatives towards the third sector in international perspective* (pp. 45–66). New York, NY: Springer.

Zorn, T.E., Flanagin, A.J., & Shoham, M.D. (2011). Institutional and noninstitutional influences on information and communication technology adoption and use among nonprofit organizations. *Human Communication Research, 37*(1), 1–33. doi:10.1111/j.1468-2958.2010.01387.x

ABOUT THE AUTHOR

Karl Urban is a doctoral candidate at the University of British Columbia, School of Social Work. Email: kshoreurban @gmail.com

APPENDIX

Modified World Management Survey Section ("Monitoring Practices")
Based on Bloom and van Reenen (2007) and Delfgaauw et al. (2011), modified for nonprofit managers engaged in the field of youth career assistance services.

1) Performance tracking
Tests whether the overall performance of the agency is tracked using meaningful metrics and with appropriate regularity.

> a. What kind of performance/quality indicators do you use to keep track of how your agency is performing (such as work placements, college/training program completion, organizational goals)?
>
> b. How do these indicators apply to individual members of staff?
>
> c. How frequently is the performance of the agency measured? Who gets to see the performance information?

Scoring grid:

> 1: No clear idea of how overall performance is measured (other than government targets). Performance measurement is ad-hoc.
>
> 3: Most important performance indicators are tracked formally; tracking is overseen by senior staff.
>
> 5: Performance is continuously tracked and communicated against most critical measures, both formally and informally, to all staff using a range of visual management tools

2) Performance review
Tests whether the performance of staff is reviewed in a comprehensive way.

> a) How do you review your main performance indicators?
>
> b) Can you describe a recent performance review meeting? How often do these meetings take place?
>
> c) Who is involved in these meetings? Who within the agency then gets to see the results of this review?
>
> d) Assuming there is a follow-up plan after such meetings: what does it look like?

Scoring grid:

> 1: Performance is reviewed infrequently or in an unmeaningful way e.g., only success or failure is noted.

3: Performance is reviewed periodically with both successes and failures identified. No clear follow-up plan is adopted.

5: Performance is continually reviewed, based on the indicators tracked. All aspects are followed up to ensure continuous improvement.

3) Performance dialogue
Tests the quality of review conversations.

a) How are these performance review meetings structured?

b) When you meet, do you generally find that you do have enough information on performance data? How is this data used?

c) When a problem is discussed during these meetings, how do you identify the root cause?

Scoring grid:

1: The right information for a constructive discussion is often not present or the quality is too low; conversations focus overly on data that is not meaningful. Clear agenda is not known and purpose is not explicitly. Next steps are not clearly defined.

3: Review conversations are held with the appropriate data present. Objectives of meetings are clear to all participating and a clear agenda is present. Conversations do not drive to the root causes of the problems, next steps are not well defined.

5: Regular review/performance conversations focus on problem solving and addressing root causes. Purpose, agenda, and follow-up steps are clear to all. Meetings are an opportunity for constructive feedback and coaching.

4) Consequence Management
Tests whether differing levels of (personal) performance lead to different consequences (good or bad).

a) Let's say you've agreed to a follow up plan at one of your meetings, what would happen if the plan was not enacted?

b) How long is it between when a problem is identified to when it is solved (again using the example of an inappropriate work placement or other)? Can you give me a recent example?

c) How do you deal with repeated failures?

Scoring grid:

1: Failure to achieve agreed objectives does not carry any consequences.

3: Failure to achieve agreed results is tolerated for a period before action is taken.

Urban   (2023)

5: A failure to achieve agreed targets drives retraining in identified areas of weakness or moving individuals to where their skills are appropriate.

5) Clarity and comparability of targets
Tests how easily understandable performance measures are and whether performance is openly communicated.

a) If I asked your staff directly whether they had been given individual performance targets, what would they tell me?

b) Do you conduct benchmarking with other organizations involved in youth employment assistance? If so, how is this data used?

Scoring grid:

1: Performance measures are complex and not clearly understood, or only relate to government targets.

3: Performance measures are well defined and communicated; organizational performance is public, but comparisons are discouraged.

5: Performance measures are well defined, strongly communicated and reinforced at all reviews; organizational performance and rankings are made public.