

RESEARCH ARTICLE OPEN ACCESS

# Looking Under the Hood Challenges and Remedies for Cross-Country Surveying of Nonprofit Organizations

Ben Suykens<sup>1,2,3</sup>  | Johan Hvenmark<sup>2</sup> | ChiaKo Hung<sup>4</sup>  | Bram Verschuere<sup>1</sup>

<sup>1</sup>Department of Public Governance and Management, Ghent University, Ghent, Belgium | <sup>2</sup>Department of Civil Society and Religion, Center for Civil Society Research, Marie Cederschiöld University, Stockholm, Sweden | <sup>3</sup>The Gradel Institute of Charity, New College, University of Oxford, Oxford, UK | <sup>4</sup>School of Social Policy and Practice, University of Pennsylvania, Philadelphia, Pennsylvania, USA

**Correspondence:** Ben Suykens ([ben.suykens@ugent.be](mailto:ben.suykens@ugent.be))

**Received:** 19 January 2023 | **Revised:** 20 January 2025 | **Accepted:** 19 February 2025

**Funding:** This research was supported by the Research Foundation Flanders (research grant FWO-1293122N).

**Keywords:** business-like | comparative method | measurement equivalence | organizational effectiveness | survey error

## ABSTRACT

Large-N cross-country surveying is a promising, yet methodologically challenging way to contextualize existing research, as it allows researchers to formulate and test hypotheses that verify under what conditions relationships are (un)likely to materialize. Integrating insights from the survey error- and equivalence literature with our own recent experience with cross-country surveying, this article offers guidance on how to navigate the methodological challenges associated with cross-country survey sampling and –measurement. Doing so, we provide a learning case that stimulates further debate about and increased engagement with large-N cross-country comparison in nonprofit research.

## 1 | Introduction

Increasingly, calls for large-N cross-country comparison appear in nonprofit management research (e.g., Anheier 2023; de Moraes Holanda et al. 2023; Simsa and Brandsen 2020). A potent way to engage in research contextualization, large-N cross-country comparison essentially enables researchers to consider a phenomenon in relation to the situation in which it exists (Verhoest et al. 2018). In specific, it allows researchers to formulate and test hypotheses to verify to what extent the drivers, manifestations and effects of a certain phenomenon hold under particular contextual conditions (cf. middle-range theory) or across different contextual settings (cf. grand theory) (O'Toole Jr. and Meier 2015; Zhao et al. 2022). This is important, as too often, “context is typically written cursorily in the abstract and Introduction of a journal article (e.g., “we studied phenomenon A in country B”), in the Methodology (e.g., “our strategic research site is country/

city X”) or occasionally mentioned as a Limitation (e.g., “we only studied country X”) and be left as an avenue for Future Research (e.g., “future research should examine whether these findings apply to other contexts”) (Chandra and Kerlin 2021, 135).” A parsimonious approach to context limits nonprofit research threefold. First, if not properly accounted for, contextual effects are likely to threaten internal validity due to omitted variable bias. Among others, Johns (2006) argues that systematic inclusion of context can (a) change causal direction, (b) reverse signs (e.g., an assumed negative association turns out to be a positive one), and/or (c) prompt curvilinear effects. Second, underreporting of contextual conditions produces weak external validity claims, which hinders systematic research integration (Abner et al. 2017). In turn and third, this leads to a limited application of research into practice (Johns 2017; Perry 2012). While a particular practice might work purposefully in the studied context, it might be an outright failure in another.

Johan Hvenmark and ChiaKo Hung contributed equally to this study.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). Nonprofit Management & Leadership published by Wiley Periodicals LLC.

This said, large-N cross-country comparisons come with formidable methodological challenges, and thus, may appear unattractive to researchers for more pragmatic reasons (Pollitt 2011). Indeed, given its complexity, it is not unusual that the bulk of resources are invested in data collection efforts, leaving data valorization suboptimal at best or a mere afterthought at worst (Verhoest et al. 2018). Hence, cross-country comparison may be considered “high risk, high reward” in the minds of many. Drawing on our recent experience surveying nonprofit organizations (NPOs) across different welfare states for illustration purposes, this article offers a methodological reflection on cross-country sampling and measurement.

By providing a look under the proverbial hood, we contribute twofold to the literature. First, it ties in with the recent calls for a more context-sensitive approach in the nonprofit literature (e.g., Chandra and Kerlin 2021; Searing et al. 2022; Zhao et al. 2022), and broadly, the fields of public (e.g., Abner et al. 2017; O’Toole Jr. and Meier 2015) and organizational management (e.g., Bamberger 2008; Johns 2017) to validate, refine, or extend commonly used theories. Second, by discussing methodological challenges, we aim to provide a learning case that aims to stimulate further debate about and increased engagement with large-N cross-country comparison in nonprofit research. In the remainder, we integrate insights from the survey error and equivalence literature to discuss the main challenges and remedies in terms of survey sampling, administration, and measurement. Before doing so, we set the scene by contrasting large-N cross-country comparison to other methods used for research contextualization and briefly outline the research project used for illustration purposes.

## 2 | Literature Review

In addition to large-N cross-country comparison, bibliometric analyses, systematic literature reviews, and meta-analyses are heralded as potent methods for knowledge integration. In light of more contextual appreciation, however, these are limited in

important ways (see Figure 1). For one, bibliometric analysis typically assesses the state of research in a given journal (e.g., Kang et al. 2021) or on a particular topic (e.g., Macías Ruano et al. 2021). Rather than verifying to what extent research findings hold beyond their studied setting, the evolution of and patterns in knowledge production are central in bibliometric analysis.

Systematic literature reviews are typically limited to *building* context theory by engaging in middle-range theorizing, yet leave the theory testing to other studies (e.g., O’Toole Jr. and Meier 2015; Post et al. 2020). Meta-analyses, in turn, engage in context theory *testing*, albeit often to a limited extent (Bamberger 2008). Contextual detail greatly varies between large-N studies, forcing meta-analysts to focus on contextual moderators that are most commonly reported or inherent to the research design, for example, do different measures of a concept result in different findings? In contrast, large-N cross-country research designs are uniquely placed to engage in both context theory building and testing, as these allow researchers to engage in middle-range theorizing by drawing on the existing literature, and subsequently collect data on the phenomenon at interest, its drivers and effects, and the contextual variation in which it appears.

Despite this promise, large-N cross-country comparison lost much of its momentum since the heyday of the Johns Hopkins Comparative Nonprofit Sector Project (Anheier 2023), inducing recent observations that cross-country studies remain relatively rare in the nonprofit field (Simsa and Brandsen 2020). Taking a closer look at so-called exemplar cross-country studies, de Morais Holanda et al. (2023) suggest that data convenience may be part of the explanation, as most cross-country studies refrain from taking an organizational perspective by drawing on individual-level and/or second-source data to study volunteering (e.g., Enjolras 2021) and giving behavior (e.g., Einolf 2017). Often, such studies are characterized by

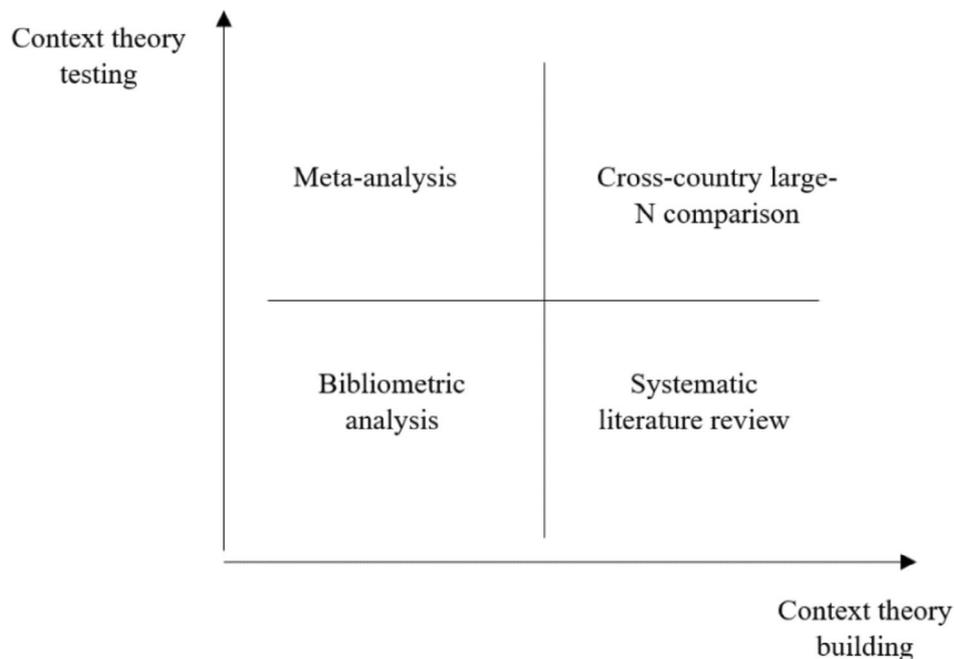


FIGURE 1 | Research contextualization methods.

methodological pragmatism. Wiepking et al. (2021) for instance merge datasets built in nineteen different countries based on different survey questionnaires that were administered at different points in time to study how institutional context affects philanthropy. Doing so, they offer the following justification: “There exist methodological weaknesses; that is different timeframes, sampling methods and questionnaires were used. These differences may lead to different estimated relationships between factors of institutionalization and philanthropic giving. However, until other micro-level data are collected, the IIPD [i.e., name of the constructed database] is the best available database to test relationships between institutional contexts and individual philanthropy across a range of countries (Wiepking et al. 2021, 704–705).” Turning to cross-country studies drawing on organizational-level survey data, similar pragmatism can be observed. For example, Yanagi et al. (2021) draw on survey data collected among NPOs across Japan, South Korea, the Philippines, and the USA to study nonprofit advocacy. Here, every participating country used its own data sources to build a sample that varied in size (ranging from 24,403 NPOs to 3300 NPOs), scope (ranging from national to regional focus), time (ranging from 2004 to 2010) and administration mode (ranging from postal service, email, interview to telephone). Overall, this “make do” approach should not be confused for opportunism, but rather, understood as important pioneering work indicating a need for more in-depth methodological reflection on how to “do” large-N cross-country comparison of nonprofit behavior. In the remainder, we aim to offer a learning case by sharing and reflecting on the meaningful difficulties that arose in our recent experience with cross-country surveying.

### 3 | Opening the Cross-Country Survey Hood

#### 3.1 | The Learning Case: A Cross-Country Survey on Nonprofit Marketization, Organizational Performance, and Context

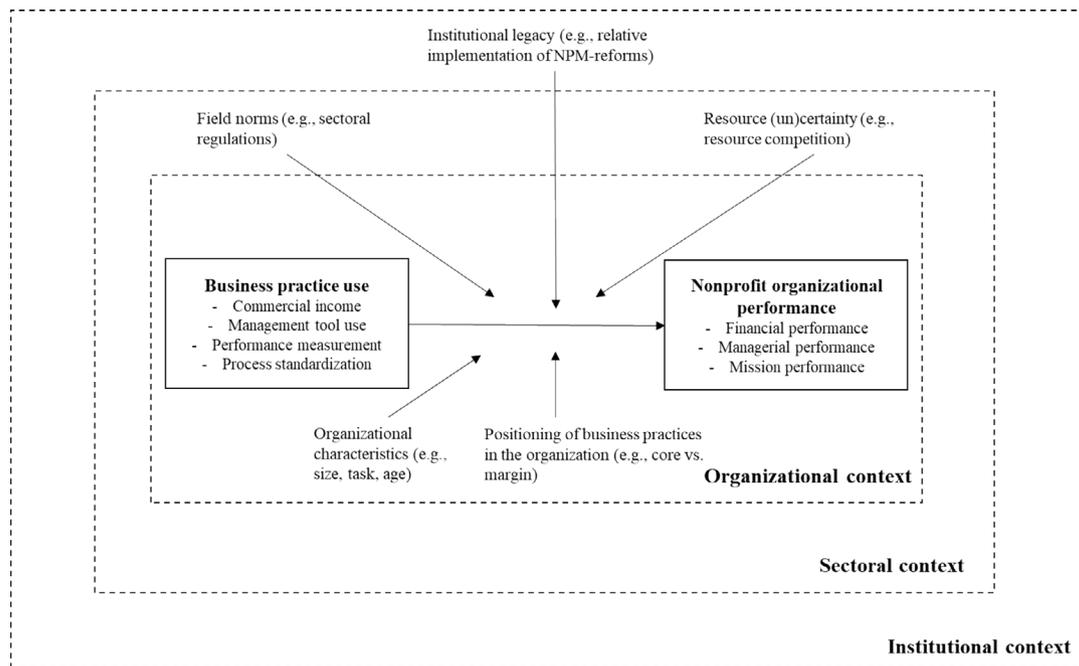
Our cross-country survey focused on a key topic in the nonprofit management debate: NPOs adopting business-like practices, commonly understood as nonprofit marketization (Eikenberry and Kluver 2004). Propelled by New Public Management, proponents argue that business practices can contribute to NPOs’ financial stability by means of business modeling (e.g., Hung and Suykens 2023), as well as their day-to-day functioning through extensive engagement with corporate management tools (e.g., Hvenmark 2013), performance measurement (e.g., Carnochan et al. 2014) and process standardization (e.g., Baines et al. 2011). These, in turn, are argued to strengthen the ability of NPOs to fulfill their prosocial goals. Although agreeable at first sight, this argument is highly contested. Critics warn that extensive use of business practices by NPOs risks crowding out donative income (Hung 2020a), is likely to erode traditional nonprofit values such as equity, equality, and participation (Eikenberry 2009), and, more importantly, keeps NPOs from fulfilling their key societal roles (Eikenberry and Kluver 2004). Seemingly risky for some while promising for others, existing research fails to provide insight into the conditions under which business practices are (un)likely to work purposefully in a nonprofit setting.

Shedding light on how contextual conditions impact the relationship between business-like behavior and key dimensions of nonprofit performance is important. Not to cast doubt on the validity of previous research, but to foster a more evidence-based debate by contextualizing said body of research.

Research contextualization arguably constitutes the logical next step in this debate. In broad terms, two distinct phases of knowledge production can be discerned. First, early work on NPOs becoming business-like is predominantly conceptual-theoretical in nature and generally portrays NPOs becoming business-like as a fundamental change largely detrimental to NPOs (see e.g., Bush 1992; Dart 2004b; Eikenberry and Kluver 2004; Salamon 1993). Although conceptually refining how NPOs can behave more business-like (Dart 2004a) as well as describing potential effects on nonprofit performance (Eikenberry and Kluver 2004; Salamon 1993; Weisbrod 2000), these early, yet influential studies largely remained stuck in a search for universal patterns and effects and suffered from a lack of empirical testing. Gradually, studies became more evidence-based. This second phase of research showed that the empirical plurality underlying the concepts “nonprofit organization” and “becoming business-like” basically disallows a black-or-white positioning regarding the notion that NPOs becoming business-like might be a wrong-headed idea from the outset. While some corroborate often-theorized effects such as financial motives taking precedence over prosocial goals (e.g., Hustinx and De Waele 2015) and extensive performance measurement eroding workplace democracy (e.g., Baines et al. 2011), others point to managerial (e.g., LeRoux and Wright 2010) and financial advantages of becoming more business-like (e.g., Gras and Mendoza-Abarca 2014). Despite providing a more balanced, evidence-based account, these findings largely remain fragmented due to a lack of integrative efforts testing the generalizability of these studies beyond their original research context.

Building on these insights, the aim of our cross-country comparison is to herald a distinct third phase of knowledge production by developing a theory of context in order to test how contextual conditions impact the relationship between business practice use and nonprofit performance (see Figure 2). In specific, our cross-country survey focused on key manifestations of business-like behavior such as NPOs generating commercial income through the profitable sale of goods and services (i.e., nonprofit commercialism) and adopting managerial practices like corporate management tools and performance management (i.e., nonprofit managerialism) (Hung 2020b; Hvenmark 2016). This approach allows us to examine this phenomenon for a large-N sample of NPOs, as all NPOs can reflect on (the changing nature of) their organizational income and management despite the many organizational differences that exist across it. Furthermore, in line with the most-cited effects of NPOs’ business-like behavior (Maier et al. 2016), the research project focuses on three inter-related, yet distinct dimensions of nonprofit performance, that is, (a) financial health, (b) mission drift, and (c) organizational effectiveness (Sowa et al. 2004).

Similar to business-like behavior and nonprofit performance, context is a multi-layered concept (Johns 2006). Based on a thorough reading of the nonprofit marketization literature, one can discern three relevant contextual levels that arguably



**FIGURE 2** | Towards a multi-level context-based approach.

impact the business practice use—nonprofit performance nexus (see Figure 2). At the organizational level, for example, several studies indicate that large(r) organizations benefit from management tool use and performance measurement to keep track of their activities (e.g., Suykens et al. 2021) and have the professional capacity present to set up commercial activity (e.g., Suykens, George, et al. 2020). In turn, commercial income can increase organizational hiring and retention rates and thus contribute to capacity-building efforts (Dart 2004a). Turning to the field level, resource uncertainty is traditionally identified as a driver of nonprofit commercialization (e.g., Guo 2006). Consequently, whereas many NPOs perceive commercialization as a financial strategy of last resort, others emphasize how commercial venturing can contribute to goal fulfillment. Brown (2018) for instance, describes how commercialization was historically viewed by social workers as a means to infuse a sense of empowerment in the professional-client relationship. Consequently, the impact of commercialization on organizational effectiveness might be contingent on the observation of whether commercial income generation is an imposed or a voluntary choice by the NPO at hand. Finally, at the institutional level, recent studies show that the adoption of business-like practices by NPOs holds true to different extents in different welfare state types (Bode 2011; Suykens, De Rynck, et al. 2020; Vaceková et al. 2017; Wang 2020). Essentially, this phenomenon is arguably most outspoken in liberal welfare states (e.g., USA) in comparison to social-democratic (e.g., Sweden) and corporatist welfare states (e.g., Belgium). An often-cited explanation for this variation is the notion that the NPM movement gained a strong foothold in liberal welfare states (Bode 2011; Pollitt and Bouckaert 2017). Building on this, a challenging question is to what extent different variations in business practice use impact nonprofit performance differently. Suykens et al. (2022) for instance, find that managerial practices do not hamper but marginally contribute to societal role uptake in a corporatist welfare

regime like Flanders, which sharply contrasts with the many critical sounds emerging from research on liberal welfare states (e.g., Baines et al. 2011).

This overview illustrates that context constitutes a major part of the puzzle as to why business-like behavior might be evident for some NPOs, while outright questionable for others. In order to paint a broad-brush picture of how contextual factors impact the business-like behavior—nonprofit performance nexus, we launched an online cross-country survey questionnaire to nonprofit executives active in the aforementioned welfare state types: Sweden (social-democratic), Belgium (corporatist) and the US (liberal) (Salamon and Anheier 1998). We now turn to the methodological issues encountered with regard to survey sampling and measurement.

## 3.2 | Methodological Challenges and Remedies

### 3.2.1 | Challenge #1. Survey Sampling

Compared to cross-country comparisons on public agencies (Verhoest et al. 2016), top public sector executives (Hammerschmid et al. 2016) and local governments (Kim et al. 2013), NPOs are arguably a more fluid, volatile unit of analysis to compare internationally. Oftentimes, NPOs are defined by what they are not, that is, private founded as opposed to public-owned, nonprofit as opposed to the for-profit motive of businesses, and formally organized as opposed to the informal ties characteristic of the community sphere. This tends to bunch together NPOs as a residual and heterogeneous collection of organizations with no clear conceptual borders (Salamon and Sokolowski 2016). The composition of the nonprofit domain in the three countries selected reflects this diversity (Salamon and Anheier 1998) (see Table 1).

**TABLE 1** | Three “ideal-typical” welfare state types.

	<b>Corporatist welfare state</b>	<b>Liberal welfare state</b>	<b>Social-democratic welfare state</b>
Level of public welfare spending	High	Low	High
Public role of the nonprofit domain	Public service delivery—Advocacy	Public service delivery—Civic engagement	Civic engagement—Community building
Composition of the nonprofit domain			
– Size	Large	Large	Small
– Main funding	Public funding	Private donations—commercial income	Membership fees—commercial income
– Workforce	Professional-based	Mix professionals—volunteers	Volunteer-based

Belgium resonates with the corporatist welfare state model. Here, NPOs are historically intertwined with the state in terms of funding, regulation, and service delivery. Resonating with the idea of third-party government (Salamon and Toepler 2015), NPOs deliver public services in response to particular human needs in exchange for public funding. This high level of publicness hints that hybridity is primarily situated at the nonprofit-state intersection (Hustinx et al. 2014). Sweden fits the social-democratic welfare state model. In this setting, generally speaking, the state delivers most public services, leaving little room for NPOs to do so. Consequently, nonprofit initiative historically focused on expressive functions such as civic engagement and political advocacy, as well as community building (see e.g., Wijkström 2011). NPOs are primarily self-financed through membership fees, often in combination with private donations and earned income. Most organizations are not heavily professionalized and are largely managed and governed by volunteers that are members of the organization (see e.g., Hvenmark 2008). In turn, these organizations are often grouped in larger federations that represent their interests towards the state. Hence, although hybridity can be observed across all nonprofit sector borderlands (i.e., state, community, market), it is primarily situated at the nonprofit-community interface in this context (see e.g., Lundberg 2020). The United States, finally, corresponds with the liberal welfare state model. NPOs traditionally deliver public services parallel to and more distant from the state. Here, commercial income through the profitable sale of services and goods is common practice in the nonprofit domain (Brown 2018). Hybridity in this context is primarily situated at the nonprofit-market intersection (Billis 2010).

In order to systematically capture this empirical diversity, we draw on the operational-structural definition originally developed by Salamon and Anheier (1992). Drawing on the insights of the Johns Hopkins Comparative Nonprofit Sector Project, they discern five key nonprofit characteristics:

- Formally organized, that is, this can materialize either through legal incorporation or having some degree of organizational permanence. One-off, ad hoc initiatives are excluded.
- Private, that is, NPOs are founded by citizens. The state can financially support NPOs, as well as sit on their boards.

- Self-governing, that is, NPOs should have a large extent of managerial autonomy. This does not exempt them from external control from resource holders.
- Voluntary, that is, this can materialize through the presence of (a) in-kind donations and/or (b) volunteers at the operational or board level.
- Nonprofit distributing, that is, NPOs are forbidden from returning profits to their organizational owners or directors. Instead, profits should be reinvested in the organizational mission. In this regard, however, we extend the traditional understanding of the structural-operational definition by adhering to a limited distribution constraint (rather than a “no” distribution constraint) (Evers and Laville 2004). This allows the inclusion of traditional European nonprofit actors like work integration social enterprises (WISE), mutuals, and cooperatives as part of the nonprofit sphere.

To ensure equivalence, the organizations included in our survey sample need to correspond with the inclusion criteria advanced by this definition. After defining our inclusion criteria, we turned to the actual selection of the country samples. In line with Wiepking et al. (2021) and similar to the John Hopkins Comparative Nonprofit Sector Project, this cross-country data collection was executed by a research consortium consisting of at least one researcher per participating country. At the time of formation, every partner had published impactful contributions to the debate on NPOs becoming more business-like (e.g., Hung 2020a, 2020b; Hvenmark 2013, 2016; Suykens et al. 2019a, 2019b). Considering the data sources available, we collaboratively developed a customized sampling approach for every participating country. The goal was to compose a sample reflective of the composition of the nonprofit domain at hand within the boundaries of the modest financial means of the project.<sup>1</sup>

- For Belgium, we departed from an existing population database of Flemish NPOs active in the human well-being ( $N=603$ ), sociocultural ( $N=1713$ ) and social economy sectors ( $N=159$ ). This database was composed by four PhD researchers in 2017 via desktop research and elite interviews, the first author being one of them.<sup>2</sup> These NPOs are predominantly active at the regional level, meaning

TABLE 2 | Cross-country samples.

	Original sample framework	Sample framework with contact details	Sample framework with accurate contact details	Survey responses (full responses)	Response rate (response rate when accounting only for the full responses)
Belgium	2475	2230	2050	559 (375)	27.27% (18.29%)
Sweden	3500	1731	1534	535 (254)	34.88% (16.56%)
United States	3530	1960	1688	349 (201)	20.68% (11.91%)

that they (a) provide services beyond the local community, (b) engage in politicization or advocacy efforts aimed at the supra-local level, and/or (c) are umbrella organizations covering the Flemish region. Human well-being NPOs include care initiatives targeting minors, people with a disability, or underprivileged people. The social economy sector consists of WISE, which provide vocational training to disadvantaged workers. The sociocultural sector, on the contrary, is more diverse. Here, we included professional associations, patient associations, social rights movements, political organizations, sociocultural associations, youth associations, ethnic-cultural associations, and faith-based associations. The choice to depart from this mapping database—and not: public databases—has arguments for and against. On a critical note, this sample is more limited than the United States and Swedish samples. Several fields of activity are excluded (e.g., hospitals, child care, elderly care, leisure initiatives), as well as those NPOs that merely operate at the local level. On a positive note, this approach ensures a more comprehensive listing of NPOs. Public databases often exist because of taxation or formal recognition purposes, and therefore, are skewed in favor of larger, more established NPOs (see e.g., Grønberg et al. 2010). Yet, several studies point out that business-like behavior such as commercial venturing is important for small-scale initiatives balancing on the verge of (in)formality, as these typically rely on mixed income streams (e.g., Suykens, George, et al. 2020).

- For Sweden, population data for the nonprofit domain is available, yet not public. Consequently, we bought a sample from Statistics Sweden, the public agency responsible for developing, producing, and disseminating official statistics. After checking what Statistics Sweden included in their definition of the nonprofit sector, which was consistent with our inclusion criteria, we ordered a sample framework of 3500 NPOs that was proportionally stratified on (a) primary field of activity and (b) organizational size. Sweden Statistics removed inactive organizations from the sample.
- For the United States, we derived a sample framework from the Business Master File managed by the Internal Revenue Service (IRS), which is publicly accessible. We proportionally stratified our sample on field of activity (NTEE) and organizational size and limited our selection to 3530 NPOs. NPOs reporting zero annual income and zero personnel were considered inactive and excluded from the sample. Furthermore, we decided to limit our

selection to NPOs based in Hawaii ( $N = 1750$ ) and Arizona ( $N = 1750$ ), which are in many aspects “most different” states. Respectively, these two states differ in terms of political color (i.e., typically democrat vs. republican), industry (i.e., little vs. high diversified economy), race (i.e., high vs. low ethnic diversity), and of particular interest here, composition of the nonprofit population (i.e., the largest nonprofit subsector in Hawaii is Arts, Culture, and Humanities whereas in Arizona it is Human Services). Doing so, we aimed to capture field dynamics (e.g., resource competition, nonprofit-business relationships) while ensuring variation at the level of the institutional context.

On the whole, 9505 NPOs were sampled across three countries. Illustrating the limited application of second-source data for international comparison (Breen 2013; Searing et al. 2022), both our sampling variables were measured in different ways in different countries. Whereas Statistics Sweden lists the number of paid employees by means of class intervals, indicators of organizational size provided in the BMF are financial in nature (e.g., asset amount, income amount). In a similar vein, every country holds a different classification of nonprofit fields of activity due to the particularities of their legal systems. In order to ensure equivalence, we decided to re-question our sampling variables. Whereas we measured organizational size as the amount of paid employees and volunteers (in classes), we turned to the International Classification of Nonprofit Organizations (ICNPO) to allow self-categorization of the primary field of activity (Salamon and Anheier 1996). Subsequently, job students were hired to identify contact details, that is, the email address of the organizational leader and/or the general email address of the organization—via systematized desk research. This endeavor took over 3 months and resulted in a final sample of 5921 NPOs (62.3%) (see Table 2).

Whereas the search for contact details removed ca. 50% of the Swedish and US sample, the more narrow Belgian sample ( $N = 2288$ ) boasted a 92% retrieval rate, which was expected since the organizations in this database were already checked for online presence when originally composed in 2017. Although two out of three country samples were stratified in terms of field of activity and organizational size, the subsequent process to identify (valid) contact details made any claim to representativeness short-lived, implying considerable sampling error (Lee et al. 2012). Often loosely used in popular and scientific publications (Kruskal and Mosteller 1979a, 1979b), the quest for representativeness in cross-country nonprofit samples

is arguably a Sisyphean task. Aside from starting from biased secondary sources (e.g., Grønberg et al. 2010; Ma 2021; Searing et al. 2022), new nonprofit initiatives rapidly emerge while others demise (Heylen et al. 2018). Statistically, representativeness can be claimed by verifying whether there are significant differences between the population and the survey respondents on key parameters like organizational size and field of activity. Hence, statistical representativeness is premised on the (a) acknowledgement that the sample is inherently biased due to the flawed nature in secondary sources, and (b) the availability of accurate information on key parameters for the population at large. Such information is country contingent and often lacking. Although forsaking representativeness rules out valid cross-country comparison at the descriptive level (Hager 2013), inferential statistics allow researchers to effectively control for over and undersampled contextual characteristics due to sampling- and non-response bias (Verhoest et al. 2018). Rather than having the correct variation, researchers first and foremost need sufficient variation on key variables in order to make valid inferences (cf. Kim et al. 2013).

This line of argument shifts the attention to survey administration, and more specifically, maximizing survey response rates. Between March and September 2022, our survey was sent out digitally to nonprofit executives (one per organization). Given the diversity represented in our country samples, nonprofit executives were broadly understood as individuals at the helm of the organization who held deep knowledge of organizational practices, strategy, and performance. This can be a board member, a director, a middle manager, or a key volunteer (in case of non-professionalized initiatives). To control whether different types of nonprofit executives responded differently to our questions, respondents were asked at the beginning of the survey to indicate their position within the organization. In each country, we conducted a soft launch—that is, an initial send-out to 100 randomly drawn organizations—to verify that everything worked technically. Subsequently, survey invitations were sent out in the name of the national project partner involved. Four digital reminders were sent in every country. Additionally, to maximize our response rate, we (a) invited once those NPOs that only had a contact form listed on their website, (b) offered an intellectual—(receive research updates for the length of the project) and a monetary incentive (a \$1 dollar donation to a good cause for every full survey response), and (c) sent out postal surveys in both Belgium ( $N=300$ ) and the United States ( $N=500$ ) of a random sample of organizations that failed to respond after several digital reminders.<sup>3</sup> We received, respectively, 62 and 30 full responses back.

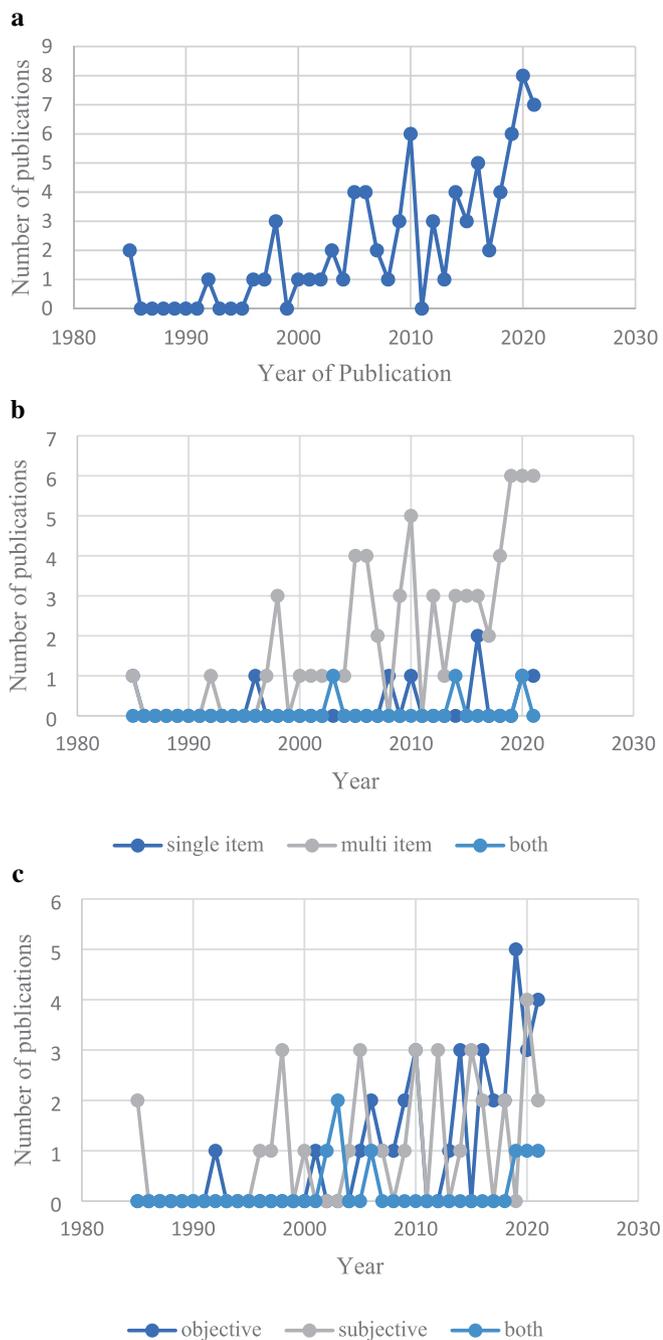
Ultimately, we collected 559 responses in Belgium (375 full responses), 349 responses in the United States (201 full responses), and 535 responses in Sweden (254 full responses) (see Table 2). Depending on the calculation method, our survey responses average around 27% (all responses) or alternatively 15% (limited to full responses). Either way, our response rates deviate from the reported 35% response rate average in organizational studies (Baruch and Holtom 2008) or the 50% response rate desired for nonprofit management studies (Hager 2013). Falling short of these marks resonates with the global decline in survey response rates (Brick and Williams 2013), and basically points to a growing survey fatigue among nonprofit executives. Already two decades ago, Hager et al. (2003, 263) for instance note that

“nonprofit executives are inundated with surveys. [...] If at one time survey science inspired awe in our publics, and respondents felt honored to be drawn into survey samples, those days are gone”. On the whole, both sampling error introduced by the search for up-to-date contact details as well as low response rates imply that nonresponse bias cannot be excluded. Although the effect of non-response bias is likely small when conducting explanatory analysis (Verhoest et al. 2018), we should be careful to generalize our descriptive analyses to the broader nonprofit population (cf. Hager 2013). Part of the explanation for these limitations lies in the realization that our research capacity (i.e., four researchers) was too small to (a) conduct thousands of pre-survey phone calls informing respondents about our research project (e.g., Hager 2013), (b) send out thousands of surveys by courier services like FedEx (e.g., Hager et al. 2003), and/or (c) provide considerable financial compensation for survey participation (e.g., Hager 2013). Future research with more budgetary means should take these strategies into consideration to increase response rates.

### 3.2.2 | Challenge #2. Survey Measurement

Given the heterogeneity of our sample, we aimed to strengthen measurement validity by (a) using statements understandable and relevant to all respondents (cf. Willems et al. 2014), (b) where possible, adopting validated measures forwarded by previous research, and (c) checking for conceptual equivalence through rigid translations and piloting in all countries involved. To illustrate the nitty-grittiness underlying this aim, we zoom in on our approach to measuring nonprofit performance. A perennial debate (see e.g., Herman and Renz 1999; Lacey et al. 2012; Willems et al. 2014), nonprofit organizational performance is essentially a multi-dimensional concept that can be measured both subjectively and objectively (Sowa et al. 2004; Willems et al. 2014). As a starting point, we reviewed the measures that have been used in nonprofit performance studies. More specifically, we searched within the three top-tier nonprofit management journals—Nonprofit Voluntary Sector Quarterly (NVSQ), Nonprofit Management & Leadership (NML), and VOLUNTAS—for articles with either “performance” or “effectiveness” in the title. Book reviews were excluded for further analysis. This systematic search was conducted between December 2021 and January 2022. Clearly on the rise (see Figure 3a), 77 articles were identified: 30 in NML, 34 in NVSQ, and 13 in VOLUNTAS. The Supporting Information details the articles and their year of publication, journal, whether the performance measure was single- or multi-item, and whether the author(s) used an objective/subjective measure.

Unsurprisingly, multi-item measures emerge as the norm for the study of nonprofit organizational performance (see Figure 3b). The objective/subjective divide constitutes a more salient issue. Objective performance measures are often viewed as the gold standard (Andrews et al. 2011). Subjective performance measures are considered second-tier, as measurement bias can be introduced through the method (i.e., common method bias [CMB]) and the respondent (i.e., social desirability bias [SDB]). Our concise literature review seems to resonate with these concerns: although subjective performance measures are dominant, objective measures are increasingly used (see Figure 3c).



**FIGURE 3** | Trends in nonprofit organizational performance measurement published in top-tier nonprofit management journals. (a) Evolution of publications containing nonprofit organizational performance measures. (b) Single versus multi-item measures. (c) Objective versus subjective measures.

However, objective performance measures are no panacea. Four arguments run against objective measurement, and ultimately motivated our decision to use subjective measures. First, closer inspection of our review learns that the rise of objective performance measures is explained by either a focus on (a) financial performance for a heterogeneous nonprofit sample, or (b) mission, board, and/or managerial performance for a homogenous nonprofit sample. Hence, the use of objective performance measures forces nonprofit scholars to either limit nonprofit performance to financial indicators or constrain contextual richness.

Second, several management studies suggest that social desirability concerns associated with subjective performance measurements are overdrawn, as studies using both objective and subjective measures show no meaningful difference in their findings (Andrews et al. 2011; Singh et al. 2016). Third, objective performance measures are commonly derived from secondary data sources, for which data quality lies beyond the control of researchers (George and Pandey 2017). More importantly, performance indicators derived from said sources are country specific, and reflect the ideological-political agenda of powerful coalitions in the country at hand. Hence, similar to their subjective counterparts, objective performance measures are equally “socially constructed in that decisions are made about what performance data should and should not be record[ing]” and how so (Andrews et al. 2011, 229–230; Searing et al. 2022). Accordingly, more often than not, such indicators are unsuitable for cross-country comparison (Verhoest et al. 2018). Fourth, measurement of nonprofit leaders' perceptions on organizational performance like organizational effectiveness, goal fulfillment and financial performance is highly relevant, as these perceptions ultimately drive organizational behavior regardless whether these perceptions are (in)accurate (Liu et al. 2014). In the end, we adopted the validated organizational effectiveness measure developed by Willems (2016), and added the item “Impacting public policy” to account for organizational advocacy effectiveness. This choice was motivated by the notion that this scale was validated, and that all items were clearly formulated and relevant to our respondents. In a similar vein, subjective measures for mission drift and financial performance were formulated by drawing on the nonprofit management literature (Hersberger-Langlosh et al. 2020; Prentice 2016) (see Supporting Information).

Already mentioned above, both SDB and CMB are in their own right potential sources of measurement error (Favero and Bullock 2015; George and Pandey 2017; Larson 2019; Lee et al. 2012). SDB occurs when respondents “change their answers for impression management (to look better than others), self-deception (to feel good about themselves), or identity definition” (Larson 2019, 534). Surveying organizational leaders on organizational performance is prone to SDB (Suykens et al. 2021), as leaders are ultimately responsible for (the lack of) organizational performance (Enticott et al. 2009). In order to counteract this dynamic, we added the following disclaimer when introducing our organizational performance questions to our respondents:

With the next set of questions, we would like to learn more about the performance of your organization. We would kindly like to remind you that these questions are by **no means an evaluation of your personal performance**. Your answers are **confidential**, and will be **anonymously** analyzed.

In addition, we (a) ensured subject anonymity in the survey introduction, (b) adhered to a neutral line of questioning, (c) administered the survey online/via mail in order to create physical distance between the interviewer and respondent (Lee and Woodliffe 2010), and (d) measured the time used to answer each survey page. In terms of the latter, the argument is that people giving socially desirable answers spend less time reflecting

on survey questions (Larson 2019). This allows researchers to conduct robustness tests by excluding outliers in response time to verify whether their findings hold. Lastly, we refrained from including a specific measure to control for SDB (see e.g., Stöber 2001), as (a) these items are likely considered off-topic at best and offensive at worst by the respondents<sup>4</sup>, and (b) it only adds to survey length, and thus, is likely to increase response fatigue.

CMB, in turn, is a risk when both the independent and dependent variables are collected by means of the same instrument (Favero and Bullock 2015). CMB is argued to inflate correlations, as one “is answering the questions at the same sitting, so the characteristics of the respondent may be a common source of measurement of measurement error across items” (Hager 2019, 470). Following the prescriptions of Podsakoff et al. (2012) and George and Pandey (2017), we applied spatial separation (i.e., buffer items between the independent and dependent variables), and used varying scale properties. For our performance measures, we used a nine-point Likert scale (−4 = “strongly disagree” to “4 = strongly agree”) (cf. Willems et al. 2021). We chose this scale to allow respondents to provide nuanced answers, and to counteract cognitive bias stemming from the social context. For instance, in the Belgian educational system, test scores typically range from “0” to “10.” This ingrained habit likely nudges Belgian respondents to effectively use the answer options “5” to “9” when presented with a “0–10” anchored Likert scale as they do not like to fail (score under 5) nor believe in a perfect score (10).

This brings us naturally to the issue of equivalence (Verhoest et al. 2018), which was verified ex-ante and ex-post survey administration. In the design phase, ensuring measurement equivalence was central to the translation of the survey questionnaire. Instead of word-for-word translations, both in-team discussions and survey pilots in the participating countries validated that our items (a) were perceived as relevant for all respondents (i.e., content equivalence), (b) held the same meaning after translation (semantic equivalence), and (c) were appropriate to measure the same latent variable (conceptual equivalence) (Tsai et al. 2022). This approach led to important adjustments in wording (e.g., volunteer has different translations and meanings in Swedish), answer options (e.g., ranking different governmental levels is country contingent), and survey length, i.e., removal of non-essential, ambiguous questions). Overall, we corroborate the need to have researchers involved with expert knowledge on the topic and target language at hand to accurately translate and adapt a survey instrument beyond its original research context (cf. Searing et al. 2022; Tsai et al. 2022; Verhoest et al. 2018).

Once the data is collected, measurement equivalence or invariance can be verified empirically (Jilke et al. 2015). We take our organizational effectiveness items derived from Willems (2016) as our case in point. Measured on a 9-point ordinal scale, we apply multigroup item response theory (IRT) models using STATA to compare item scores across the three countries. Specifically, we run multi-group models (RSMs) by adding the group() option to the IRT command. We examine the hypothesis that our items are perceived equally by respondents across the countries using a likelihood-ratio test. The results indicate that respondents

across countries interpreted some items equally (Chi-square not significant—1 and 8) while others are slightly different (Chi-square border significant—i.e., item 5 and 6) or different (Chi-square significant—i.e., item 2, 3, 4, 7) (see Table 3).

How to handle item bias stemming from measurement non-equivalence? On the one hand, researchers can decide to use single item variables that meet equivalence standards. This choice resonates with the argument posited by Willems et al. (2014, 1656) that “a unidimensional or even single criterion perspective on effectiveness deserves more attention, especially if the aim is to (1) improve generalizability across contexts, (2) enhance analytically robustness, [...]”. On the other hand, researchers can move forward using latent variables, yet control for nonequivalence by calculating the sum of the nonequivalent item scores and add these as a covariate to their model (Jilke et al. 2015).

## 4 | Discussion and Conclusion

This article started from the observation that calls for more cross-country comparisons increasingly appear in the nonprofit literature (e.g., Anheier 2023; de Moraes Holanda et al. 2023; Simsa and Brandsen 2020). Typically, large-N cross-country comparison is often considered a promising endeavor for research contextualization, yet it is marred by methodological challenges (Pollitt 2011; Verhoest et al. 2018). Drawing on our recent experience with cross-country surveying, we took the opportunity to write this article as a learning opportunity by sharing how we remedied the challenges tied to cross-country survey sampling and measurement (see Table 4). Doing so, we aim to stimulate future research to engage more with cross-country surveying, as well as trigger further dialogue on how

**TABLE 3** | Results from likelihood-ratio test for differences across groups.

Nonprofit effectiveness items	Chi-squared	p
1. Optimal use of available resources	11.24	0.26
2. Deployment of a predefined strategy	26.26	0.01
3. Bringing the planned strategic actions to a good end	18.74	0.02
4. Keeping the vision and mission up to date	13.38	0.04
5. Creating efficient output from limited means available	13.10	0.07
6. Enforcing changes to our society	16.79	0.08
7. Impacting public policy	34.79	0.01
8. Delivery of services to service recipients	6.83	0.23

*Note:* Because not all of the questions have the same number of levels of responses for each country, we remove some responses to make the number of levels of responses the same across the countries to run multiple-group IRT models (11 responses are removed from Question 1; 13 from Question 2; 11 responses from Question 3; 25 from Question 4; 8 from Question 5; 10 from Question 6; 34 from Question 7; 35 from Question 8).

**TABLE 4** | Cross-country survey design: Challenges and remedies.

Survey sampling	Equivalence	1. Define clear inclusion—and exclusion criteria for survey sampling purposes. Our approach: operational-structural definition of Salamon and Anheier (1992).
	Survey error (sampling/non-response)	2. Re-survey sampling variables for standardization purposes. Our approach: re-survey the primary field of activity in accordance to the ICNPO categorization (Salamon and Anheier 1996), as well as organizational size by asking to indicate the amount of professionals and volunteers using an ordinal scale.
		3. Statistical representativeness can be claimed when (a) sampling variables are available at the population level so a random proportional stratified sample can be drawn, and (b) Chi square tests indicate that there are no significant differences between the distribution of the sampling variables of the population at large, and full responses to the survey. If not, the survey suffers from sampling error, which in turn might cause non-response error.  Our approach: due to the search for up-to-date contact details, we abandoned any claim of statistical representativeness. Following the survey prescriptions for (nonprofit) organizational research (Baruch and Holtom 2008; Hager 2013; Hager et al. 2003), we aimed to maximize response rate to minimize error in regression estimates. This minimizes the impact of both sampling- and non-response error.
Survey measurement	Equivalence	4. Remedy measurement equivalence ex-ante and ex-post survey administration. Our approach: <ul style="list-style-type: none"> <li>• Ex-ante: (1) identify measures that have been tested and validated by previous research, (2) justify choices underlying objective/subjective and single/multi-item measurement (Willems et al. 2014), (3) validate the accuracy of survey translation via survey piloting in all participating countries, and in-team discussions (Tsai et al. 2022). Here, it is essential to have experts on both the topic and target language at hand.</li> <li>• Ex-post: verify measurement equivalence via multiple-group confirmatory factor analysis (continuous variables) or item response theory (Jilke et al. 2015).</li> </ul>
	Survey error (social desirability bias/ common method bias)	5. Follow the recommendations of Larson (2019) and Lee and Woodliffe (2010) to counteract social desirability bias. Our approach: ensure anonymity, neutral questions, online survey administration & survey response timing.
		6. Following the recommendations of George and Pandey (2017); Podsakoff et al. (2012) to counteract common method bias. Our approach: spatial separation between dependent and independent variables, varying scale properties. Dependent variables scored on a 9-point Likert scale ranging from ‘-4’ to ‘4’ in order to omit cognitive bias due to the particularities of the social context.

to do it best. Rekindling the momentum from the seminal Johns Hopkins Comparative Nonprofit Sector Project is arguably only the beginning, as we encourage future research to go beyond the descriptive approach of the Johns Hopkins project by systematically collecting cross-country survey data on independent, interaction, and dependent variables to apply inferential statistics to learn to what extent key mechanisms underlying nonprofit behavior are context (in)sensitive.

In addition to our methodological reflections, and tying in with calls for more research transparency (Kim and Raggo 2022), we

offer our survey questionnaire with concise explanatory notes detailing the logic and purpose of our questions for third-party use (see online [Supporting Information](#)). Researchers can use both as resources to build their research design when engaging in country comparison. This is important, as academic “conversations often focus solely on research findings, not the decisions in research design or data collection processes leading to those results” (Willems et al. 2014, 1665). Making survey questionnaires available for reuse and being transparent about the decisions underlying its design and administration ultimately enables critical evaluation of methodological strengths and

weaknesses, and thus, is likely to encourage both methodological innovation and research replication (Bekkers et al. n.d.). After all, better cars are not built by gazing at the catalog, but by taking a good look under the hood.

---

### Acknowledgments

We would like to thank the editor and reviewers for their constructive feedback. This research was supported by the Research Foundation Flanders (research grant FWO-1293122N). Johan Hvenmark wishes to acknowledge the importance of the supportive and creative academic environment at SCANCOR at the Weatherhead Center, Harvard University, USA, during the development of this manuscript.

### Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

### Endnotes

<sup>1</sup>This data collection was part of a three-year research project with a yearly bench fee of €10,000.

<sup>2</sup>The empirical value of this sample framework is already validated via A1 publications in *Nonprofit Voluntary Sector Quarterly* and *Public Management Review*.

<sup>3</sup>Postal distribution of the survey was funded on budgetary extras for a value of ca. \$1500 in each country. The difference in N is explained by different costs tied to postal services in both countries. Given that the survey distribution in Sweden was delayed due to ethical review of our research project (which was unanimously approved), we decided against sending out our survey via post in Sweden.

<sup>4</sup>For example, consider the following yes/no statements as proposed by Stöber (2001): 'I sometimes litter', 'In traffic, I am always polite and considerate for others', 'I have tried illegal drugs (e.g., marijuana, cocaine, etc.)', 'I take out my bad moods on others now and then'.

### References

Abner, G. B., S. Y. Kim, and J. L. Perry. 2017. "Building Evidence for Public Human Resource Management: Using Middle Range Theory to Link Theory and Data." *Review of Public Personnel Administration* 37, no. 2: 139–159.

Andrews, R., G. Boyne, and R. M. Walker. 2011. "The Impact of Management on Administrative and Survey Measures of Organizational Performance." *Public Management Review* 13, no. 2: 227–255.

Anheier, H. K. 2023. "Comparative Research on Nonprofit Organizations and Sectors: Looking Back and Looking Forward." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 34, no. 6: 1115–1121.

Baines, D., I. Cunningham, and H. Fraser. 2011. "Constrained by Managerialism: Caring as Participation in the Voluntary Social Services." *Economic and Industrial Democracy* 32, no. 2: 329–352. <https://doi.org/10.1177/0143831X10377808>.

Bamberger, P. 2008. "From the Editors Beyond Contextualization: Using Context Theories to Narrow the Micro-Macro Gap in Management Research." *Academy of Management Briarcliff Manor* 51, no. 5: 839–846.

Baruch, Y., and B. C. Holtom. 2008. "Survey Response Rate Levels and Trends in Organizational Research." *Human Relations* 61, no. 8: 1139–1160.

Bekkers, R., C. M. Chapman, A. M. Prosser, P. Wiepking, and M. Meyer. n.d. "Let's Be Clear: Why and How to Demonstrate Data and Methods Transparency in Nonprofit Research." Preprint.

Billis, D. 2010. *Hybrid Organizations and the Third Sector: Challenges for Practice, Theory and Policy*. Palgrave Macmillan.

Bode, I. 2011. "Creeping Marketization and Post-Corporatist Governance: The Transformation of State–Nonprofit Relations in Continental Europe." In *Governance and Regulation in the Third Sector*, 123–149. Routledge.

Breen, O. B. 2013. "The Disclosure Panacea: A Comparative Perspective on Charity Financial Reporting." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 24, no. 3: 852–880.

Brick, J. M., and D. Williams. 2013. "Explaining Rising Nonresponse Rates in Cross-Sectional Surveys." *Annals of the American Academy of Political and Social Science* 645, no. 1: 36–59.

Brown, M. 2018. "The Moralization of Commercialization: Uncovering the History of Fee-Charging in the US Nonprofit Human Services Sector." *Nonprofit and Voluntary Sector Quarterly* 47, no. 5: 960–983.

Bush, R. 1992. "Survival of the Nonprofit Spirit in a for-Profit World." *Nonprofit and Voluntary Sector Quarterly* 21, no. 4: 391–410. <https://doi.org/10.1177/089976409202100406>.

Carnochan, S., M. Samples, M. Myers, and M. J. Austin. 2014. "Performance Measurement Challenges in Nonprofit Human Service Organizations." *Nonprofit and Voluntary Sector Quarterly* 43, no. 6: 1014–1032.

Chandra, Y., and J. A. Kerlin. 2021. *Social Entrepreneurship in Context: Pathways for New Contributions in the Field*. Vol. 14, 135–151. Taylor & Francis.

Dart, R. 2004a. "Being 'Business-Like' in a Nonprofit Organization: A Grounded and Inductive Typology." *Nonprofit and Voluntary Sector Quarterly* 33, no. 2: 290–310. <https://doi.org/10.1177/0899764004263522>.

Dart, R. 2004b. "The Legitimacy of Social Enterprise." *Nonprofit Management & Leadership* 14, no. 4: 411–424.

de Moraes Holanda, B., H. Kim, and S. Appe. 2023. "Comparative Civil Society and Third Sector Research in Voluntas." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 34: 1122–1133.

Eikenberry, A. M. 2009. "Refusing the Market." *Nonprofit and Voluntary Sector Quarterly* 38, no. 4: 582–596. <https://doi.org/10.1177/0899764009333686>.

Eikenberry, A. M., and J. D. Kluver. 2004. "The Marketization of the Nonprofit Sector: Civil Society at Risk?" *Public Administration Review* 64, no. 2: 132–140. <https://doi.org/10.1111/j.1540-6210.2004.00355.x>.

Einolf, C. J. 2017. "Cross-National Differences in Charitable Giving in the West and the World." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 28: 472–491.

Enjolras, B. 2021. "Explaining the Varieties of Volunteering in Europe: A Capability Approach." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 32, no. 6: 1187–1212.

Enticott, G., G. A. Boyne, and R. M. Walker. 2009. "The Use of Multiple Informants in Public Administration Research: Data Aggregation Using Organizational Echelons." *Journal of Public Administration Research and Theory* 19, no. 2: 229–253.

Evers, A., and J.-L. Laville. 2004. "Defining the Third Sector in Europe." In *Third Sector in Europe*, 11–42. Edward Elgar Publishing.

Favero, N., and J. B. Bullock. 2015. "How (Not) to Solve the Problem: An Evaluation of Scholarly Responses to Common Source Bias." *Journal of Public Administration Research and Theory* 25, no. 1: 285–308.

George, B., and S. K. Pandey. 2017. "We Know the Yin—But Where Is the Yang? Toward a Balanced Approach on Common Source Bias in Public Administration Scholarship." *Review of Public Personnel Administration* 37, no. 2: 245–270.

- Gras, D., and K. I. Mendoza-Abarca. 2014. "Risky Business? The Survival Implications of Exploiting Commercial Opportunities by Nonprofits." *Journal of Business Venturing* 29, no. 3: 392–404.
- Grønbjerg, K. A., H. K. Liu, and T. H. Pollak. 2010. "Incorporated but Not IRS-Registered: Exploring the (Dark) Grey Fringes of the Nonprofit Universe." *Nonprofit and Voluntary Sector Quarterly* 39, no. 5: 925–945.
- Guo, B. 2006. "Charity for Profit? Exploring Factors Associated With the Commercialization of Human Service Nonprofits." *Nonprofit and Voluntary Sector Quarterly* 35, no. 1: 123–138. <https://doi.org/10.1177/0899764005282482>.
- Hager, M. A. 2013. "Gathering Information From Nonprofits: Toward Representative Survey Samples." *Journal of Nonprofit Education and Leadership* 3, no. 1: 47–59.
- Hager, M. 2019. "Editor's Notes." In *Nonprofit Management and Leadership*. Wiley.
- Hager, M. A., S. Wilson, T. H. Pollak, and P. M. Rooney. 2003. "Response Rates for Mail Surveys of Nonprofit Organizations: A Review and Empirical Test." *Nonprofit and Voluntary Sector Quarterly* 32, no. 2: 252–267.
- Hammerschmid, G., S. Van de Walle, R. Andrews, and P. Bezes. 2016. *Public Administration Reforms in Europe: The View From the Top*. Edward Elgar Publishing.
- Herman, R. D., and D. O. Renz. 1999. "Theses on Nonprofit Organizational Effectiveness." *Nonprofit and Voluntary Sector Quarterly* 28, no. 2: 107–126.
- Hersberger-Langloh, S. E., S. Stühlinger, and G. von Schnurbein. 2020. "Institutional Isomorphism and Nonprofit Managerialism: For Better or Worse?" *Nonprofit Management and Leadership* 31, no. 3: 461–480. <https://doi.org/10.1002/nml.21441>.
- Heylen, F., B. Fraussen, and J. Beyers. 2018. "Live to Fight Another Day? Organizational Maintenance and Mortality Anxiety of Civil Society Organizations." *Nonprofit and Voluntary Sector Quarterly* 47, no. 6: 1249–1270.
- Hung, C. 2020a. "Commercialization and Nonprofit Donations: A Meta-Analytic Assessment and Extension." *Nonprofit Management and Leadership* 31: 287–309.
- Hung, C. 2020b. "Decomposing the Effect of Commercialization on Nonprofit Donations." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 32: 448–459.
- Hung, C., and B. Suykens. 2023. "Donors' Giving Decisions Toward Nonprofit Commercialization: Do Commercial Form and Intensity Matter?" *Nonprofit Management and Leadership* 34, no. 2: 293–315.
- Hustinx, L., and E. De Waele. 2015. "Managing Hybridity in a Changing Welfare Mix: Everyday Practices in an Entrepreneurial Nonprofit in Belgium." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 26, no. 5: 1666–1689.
- Hustinx, L., B. Verschuere, and J. De Corte. 2014. "Organisational Hybridity in a Post-Corporatist Welfare Mix: The Case of the Third Sector in Belgium." *Journal of Social Policy* 43, no. 2: 391–411. <https://doi.org/10.1017/S0047279413000962>.
- Hvenmark, J. 2008. "Reconsidering Membership: A Study of Individual members' Formal Affiliation With Democratically Governed Federations." PhD thesis, Ekonomiska Forskningsinstitutet, Handelshögskolan i Stockholm.
- Hvenmark, J. 2016. "Ideology, Practice, and Process? A Review of the Concept of Managerialism in Civil Society Studies." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 27, no. 6: 2833–2859.
- Hvenmark, J. 2013. "Business as Usual? On Managerialization and the Adoption of the Balanced Scorecard in a Democratically Governed Civil Society Organization." *Administrative Theory & Praxis* 35, no. 2: 223–247. <https://doi.org/10.2753/ATP1084-1806350203>.
- Jilke, S., B. Meuleman, and S. Van de Walle. 2015. "We Need to Compare, but How? Measurement Equivalence in Comparative Public Administration." *Public Administration Review* 75, no. 1: 36–48.
- Johns, G. 2006. "The Essential Impact of Context on Organizational Behavior." *Academy of Management Review* 31, no. 2: 386–408.
- Johns, G. 2017. "Reflections on the 2016 Decade Award: Incorporating Context in Organizational Research." *Academy of Management Review* 42, no. 4: 577–595.
- Kang, C. H., Y. M. Baek, and E. H.-J. Kim. 2021. "Half a Century of NVSQ: Thematic Stability Across Years and Editors." *Nonprofit and Voluntary Sector Quarterly* 51, no. 3: 8997640211017676.
- Kim, M., and P. Raggio. 2022. "Taking Stock on How We Research the Third Sector: Diversity, Pluralism, and Openness." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 33, no. 6: 1107–1113.
- Kim, S., W. Vandenberghe, B. E. Wright, et al. 2013. "Investigating the Structure and Meaning of Public Service Motivation Across Populations: Developing an International Instrument and Addressing Issues of Measurement Invariance." *Journal of Public Administration Research and Theory* 23, no. 1: 79–102.
- Kruskal, W., and F. Mosteller. 1979a. "Representative Sampling, I: Non-Scientific Literature." *International Statistical Review/Revue Internationale de Statistique* 47, no. 1: 13–24.
- Kruskal, W., and F. Mosteller. 1979b. "Representative Sampling, II: Scientific Literature, Excluding Statistics." *International Statistical Review/Revue Internationale de Statistique* 47: 111–127.
- Larson, R. B. 2019. "Controlling Social Desirability Bias." *International Journal of Market Research* 61, no. 5: 534–547.
- Lecy, J. D., H. P. Schmitz, and H. Swedlund. 2012. "Non-governmental and Not-for-Profit Organizational Effectiveness: A Modern Synthesis." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 23, no. 2: 434–457.
- Lee, G., J. Benoit-Bryan, and T. P. Johnson. 2012. "Survey Research in Public Administration: Assessing Mainstream Journals With a Total Survey Error Framework." *Public Administration Review* 72, no. 1: 87–97.
- Lee, Z., and L. Woodliffe. 2010. "Donor Misreporting: Conceptualizing Social Desirability Bias in Giving Surveys." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 21, no. 4: 569–587. <https://doi.org/10.1007/s11266-010-9153-5>.
- LeRoux, K., and N. S. Wright. 2010. "Does Performance Measurement Improve Strategic Decision Making? Findings From a National Survey of Nonprofit Social Service Agencies." *Nonprofit and Voluntary Sector Quarterly* 39, no. 4: 571–587.
- Liu, G., S. Takeda, and W.-W. Ko. 2014. "Strategic Orientation and Social Enterprise Performance." *Nonprofit and Voluntary Sector Quarterly* 43, no. 3: 480–501. <https://doi.org/10.1177/0899764012468629>.
- Lundberg, E. 2020. "Toward a New Social Contract? The Participation of Civil Society in Swedish Welfare Policymaking, 1958–2012." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 31, no. 6: 1371–1384.
- Ma, J. 2021. "Automated Coding Using Machine Learning and Remapping the US Nonprofit Sector: A Guide and Benchmark." *Nonprofit and Voluntary Sector Quarterly* 50, no. 3: 662–687.
- Macías Ruano, A. J., J. Milán-García, M. E. Marruecos Rumi, and J. De Pablo Valenciano. 2021. "Scientific Production on the Social Economy: A Review of Worldwide Research." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 32, no. 5: 925–943.
- Maier, F., M. Meyer, and M. Steinbereithner. 2016. "Nonprofit Organizations Becoming Business-Like: A Systematic Review." *Nonprofit and Voluntary Sector Quarterly* 45, no. 1: 64–86. <https://doi.org/10.1177/0899764014561796>.

- O'Toole, L. J., Jr., and K. J. Meier. 2015. "Public Management, Context, and Performance: In Quest of a More General Theory." *Journal of Public Administration Research and Theory* 25, no. 1: 237–256.
- Perry, J. L. 2012. "How Can We Improve Our Science to Generate More Usable Knowledge for Public Professionals?" *Public Administration Review* 72: 479–482.
- Podsakoff, P. M., S. B. MacKenzie, and N. P. Podsakoff. 2012. "Sources of Method Bias in Social Science Research and Recommendations on How to Control It." *Annual Review of Psychology* 63: 539–569.
- Pollitt, C. 2011. "Not Odious but Onerous: Comparative Public Administration." *Public Administration* 89, no. 1: 114–127.
- Pollitt, C., and G. Bouckaert. 2017. *Public Management Reform: A Comparative Analysis-Into the Age of Austerity*. Oxford University Press.
- Post, C., R. Sarala, C. Gattrell, and J. E. Prescott. 2020. "Advancing Theory With Review Articles." *Journal of Management Studies* 57, no. 2: 351–376.
- Prentice, C. R. 2016. "Why So Many Measures of Nonprofit Financial Performance? Analyzing and Improving the Use of Financial Measures in Nonprofit Research." *Nonprofit and Voluntary Sector Quarterly* 45, no. 4: 715–740.
- Salamon, L. M. 1993. "The Marketization of Welfare: Changing Nonprofit and For-Profit Roles in the American Welfare State." *Social Service Review* 67, no. 1: 16–39.
- Salamon, L. M., and H. K. Anheier. 1992. "In Search of the Non-profit Sector. I: The Question of Definitions." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 3, no. 2: 125–151.
- Salamon, L. M., and H. K. Anheier. 1996. *The International Classification of Nonprofit Organizations: ICNPO-Revision 1, 1996*. Baltimore: Johns Hopkins University Institute for Policy Studies.
- Salamon, L. M., and H. K. Anheier. 1998. "Social Origins of Civil Society: Explaining the Nonprofit Sector Cross-Nationally." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 9, no. 3: 213–248.
- Salamon, L. M., and S. W. Sokolowski. 2016. "Beyond Nonprofits: Re-Conceptualizing the Third Sector." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 27, no. 4: 1515–1545. <https://doi.org/10.1007/s11266-016-9726-z>.
- Salamon, L. M., and S. Toepler. 2015. "Government–Nonprofit Cooperation: Anomaly or Necessity?" *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 26, no. 6: 2155–2177. <https://doi.org/10.1007/s11266-015-9651-6>.
- Searing, E. A., N. J. Grasse, and A. Rutherford. 2022. "The Promise and Perils of Comparing Nonprofit Data Across Borders." *Nonprofit and Voluntary Sector Quarterly* 52, no. 1: 8997640221114140.
- Simsa, R., and T. Brandsen. 2020. *The Evolution of Third Sector Research and the Journal Voluntas: The Editors' Impressions*. Springer.
- Singh, S., T. K. Darwish, and K. Potočník. 2016. "Measuring Organizational Performance: A Case for Subjective Measures." *British Journal of Management* 27, no. 1: 214–224.
- Sowa, J. E., S. C. Selden, and J. R. Sandfort. 2004. "No Longer Unmeasurable? A Multidimensional Integrated Model of Nonprofit Organizational Effectiveness." *Nonprofit and Voluntary Sector Quarterly* 33, no. 4: 711–728.
- Stöber, J. 2001. "The Social Desirability Scale-17 (SDS-17): Convergent Validity, Discriminant Validity, and Relationship With Age." *European Journal of Psychological Assessment* 17, no. 3: 222.
- Suykens, B., F. De Rynck, and B. Verschuere. 2019a. "Examining the Influence of Organizational Characteristics on Nonprofit Commercialization." *Nonprofit Management and Leadership* 30, no. 2: 339–351.
- Suykens, B., F. De Rynck, and B. Verschuere. 2019b. "Nonprofit Organizations in Between the Nonprofit and Market Spheres: Shifting Goals, Governance and Management?" *Nonprofit Management and Leadership* 29, no. 4: 623–636.
- Suykens, B., F. De Rynck, and B. Verschuere. 2020. "Examining the Extent and Coherence of Nonprofit Hybridization Toward the Market in a Post-Corporatist Welfare State." *Nonprofit and Voluntary Sector Quarterly* 49, no. 5: 899764020908344. <https://doi.org/10.1177/0899764020908344>.
- Suykens, B., B. George, F. De Rynck, and B. Verschuere. 2020. "Determinants of Non-profit Commercialism. Resource Deficits, Institutional Pressures or Organizational Contingencies?" *Public Management Review* 23, no. 10: 1–23. <https://doi.org/10.1080/14719037.2020.1764083>.
- Suykens, B., F. Maier, M. Meyer, and B. Verschuere. 2022. "Business-Like and Still Serving Society? Investigating the Relationship Between NPOs Being Business-Like and Their Societal Roles." *Nonprofit and Voluntary Sector Quarterly* 52, no. 3: 8997640221106979.
- Suykens, B., K. Meyfroidt, S. Desmidt, and B. Verschuere. 2021. "Does Performance-Based Accountability Impact How Non-profit Directors Perceive Organizational Performance? Insights From Rational Planning." *Public Management Review* 24, no. 9: 1355–1382.
- Tsai, T.-I., L. Luck, D. Jefferies, and L. Wilkes. 2022. "Challenges in Adapting a Survey: Ensuring Cross-Cultural Equivalence." *Nurse Researcher* 30, no. 2: 28–32.
- Vaceková, G., V. Valentinov, and J. Nemeč. 2017. "Rethinking Nonprofit Commercialization: The Case of The Czech Republic." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 28, no. 5: 2103–2123.
- Verhoest, K., S. Van Thiel, G. Bouckaert, P. Lægread, and S. Van Thiel. 2016. *Government Agencies: Practices and Lessons From 30 Countries*. Springer.
- Verhoest, K., J. Wynen, W. Vandenebeele, and S. Van de Walle. 2018. "Challenges for Large-Scale International Comparative Survey-Based Research in Public Administration." In *The Palgrave Handbook of Public Administration and Management in Europe*, edited by K. Verhoest, J. Wynen, W. Vandenebeele, and S. Van de Walle, 1147–1168. Springer.
- Wang, X. L. 2020. "Marketization in a Statist-Corporatist Nonprofit Sector: The Case of Hong Kong." *International Review of Administrative Sciences* 88, no. 2: 20852320925867.
- Weisbrod, B. 2000. *To Profit or Not to Profit: The Commercial Transformation of the Nonprofit Sector*. Cambridge University Press.
- Wiepking, P., F. Handy, S. Park, et al. 2021. "Global Philanthropy: Does Institutional Context Matter for Charitable Giving?" *Nonprofit and Voluntary Sector Quarterly* 50, no. 4: 899764021989444. <https://doi.org/10.1177/0899764021989444>.
- Wijkström, F. 2011. "'Charity Speak and Business Talk' The On-Going (Re) hybridization of Civil Society." Paper presented at the Nordic Civil Society at a Cross-Roads, 2011.
- Willems, J. 2016. "Building Shared Mental Models of Organizational Effectiveness in Leadership Teams Through Team Member Exchange Quality." *Nonprofit and Voluntary Sector Quarterly* 45, no. 3: 568–592.
- Willems, J., S. Boenigk, and M. Jegers. 2014. "Seven Trade-Offs in Measuring Nonprofit Performance and Effectiveness." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 25, no. 6: 1648–1670.
- Willems, J., L. Faulk, and S. Boenigk. 2021. "Reputation Shocks and Recovery in Public-Serving Organizations: The Moderating Effect of Mission Valence." *Journal of Public Administration Research and Theory* 31, no. 2: 311–327.

Yanagi, I., Y. Kobashi, R. J. Pekkanen, and Y. Tsujinaka. 2021. "Distinguishing Providing Public Services From Receiving Government Funding as Factors in Nonprofit Advocacy." *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations* 32: 534–547.

Zhao, Y., J. Galaskiewicz, and E. Yoon. 2022. "Reconciling Theory and Context in Comparative Nonprofit Research." *Nonprofit and Voluntary Sector Quarterly* 51, no. 1: 5–30.

### **Supporting Information**

Additional supporting information can be found online in the Supporting Information section.