**Social Program Evaluations: Strategies and Shared Value**

**Abstract**

Socially responsibility (SR) and Sport-for-Development (SFD) programs require robust evaluations to gauge aggregate effectiveness and demonstrate impact. While the idea of a social program evaluation is not novel, evaluation techniques and delivery methods have improved researcher understanding of how they should be approached. In this article, we package this information into an informative bundle of our processes, mechanisms, and outcomes regarding how we design, deploy, and deliver social evaluations. In doing so, we also illustrate that by maintaining a balance between social and organizational value, organizations can benefit themselves as well society. In addition, we suggest that by combining methodological approaches, program evaluators can construct a stronger programmatic impact narrative.

**Strategies for Social Programs and Evaluations**

* Seek and measure shared value though a balance of meeting societal needs while at the same time yielding organizational benefit.
* Align business objectives with social programming to ensure program viability and sustainability.
* Be proactive, not reactive.
* Theorize social change by identifying the social issue and mapping program characteristics to achieve the desired outcomes.
* Design the evaluation to achieve valid conclusions that employ a blend of research methods and data analysis techniques.

**Social Program Evaluations: Strategies, Methods, and Shared Value**

We have been involved with evaluating numerous social programs that use sport as a mechanism for individual (i.e., personal) development. While the list of possible developmental outcomes is long, the programs we collaborate with focus their interventions on employment training, social mobility, character building, health, and access to formal education. In conducting this work, we have learned many things that have improved our approach to evaluating social programs. The most important realization, however, is that a ‘one size fits all’ blueprint for deploying a program or conducting a program evaluation is not realistic. Each program and the associated evaluation is unique, and depends on a number of items such as organizational and program goals, program structure and size, initiation, and key program tenets. In addition, identifying and working with the intended beneficiary group(s), working closely with the delivery entity, understanding the funding mechanisms and partnerships, and revealing program motives early in the evaluation process are critical to a successful research endeavor (Walker, Hills, & Heere, 2015). We have also learned that by defining the key performance indicators and developmental objectives, and understanding how the organization and funders view impact from day one, are essential pieces of information (Walker, Heere, & Kim, 2013).

The latter (i.e., impact) point has been critical for our work, since many social program evaluations in sport have failed to show “… transparent and substantive evidence” of developmental or social impact (Levermore, 2011, p. 552). This has been the case since third-party evaluations are expensive and many social programs operate with limited resources that primarily go towards staffing and delivery. Moreover, resource limitations encumber robust and holistic research evaluations that should rely on a balance of qualitative and quantitative methods. With limited funds, many social programs only report anecdotal evidence of impact in the forms of heartfelt narratives and evocative images to elicit a ‘warm glow effect’.

Confusing the evaluation process is that many social programs have adopted very broad outcomes, which include reinvigorating communities, reversing childhood obesity, improving youth literacy, enabling gender equality and social scaffolding, peace building, and social capital creation (Walker et al., 2015). While assuaging these issues are worthwhile pursuits, capacity building for social impact must start from, and focus intently on, the individual-level (i.e., program beneficiary). Only if the intervention delivery is sound, the outcomes are well defined, realistic, and measurable, and the program is sustainable, can the program be scaled to achieve broader social impact. This is something that few programs, at least the ones we have studied and observed, can espouse to achieve. To approach social impact, however, a theory of change (ToC) is required to identify the sequence of causes and presumed effects that provide the undergirding to the social intervention (Coalter, 2007). In constructing the ToC, our process has been to hypothesize the relationships between the social problem, program beneficiaries, and program mechanisms and processes to achieve both intermediate impacts (i.e., effect on the beneficiary) and broader outcomes (i.e., individual behavioral or social changes). The ToC provides the structure to the entire evaluation, so we spend considerable time on this aspect in the pre-planning phase of the research partnership. We also work iteratively with program managers to ensure what is being mapped in the ToC aligns with program goals and the outcomes sought. Doing so creates a clear and realistic tone from the outset of the project and helps define the roles of the researchers and delivery partner.

As organizations are increasingly attaching developmental outcomes to their socially responsible (SR) agendas, and governments and non-profits are working to assuage various societal ills via social programming, it is critical for program managers and funders to evaluate the effectiveness of their intervention. Evaluations serve multiple purposes, but moving beyond the appearance of responsibility to demonstrate development and potentially social impact are perhaps most important. In this article, we refer to this as creating social value. However, social value in isolation means little if the organization fails to benefit from leveraging opportunities created by the program. We refer to this as organizational value. Based on these value propositions, we orient our discussion though a shared value lens, which emphasizes creating value for society in a way that also creates organizational value (Kramer & Porter, 2011). This point is important since a number of social programs in sport are claiming impact but empirical, unbiased, third-party data on impacts are often non-existent. In other words, many social investments are simply feel-good responses to external pressures as opposed to drivers of impact, productivity, and program sustainability. In fact, the majority of SR studies in sport have centered on business-level benefits such as reputation, image, and patronage intentions (see Walker, 2013; Walker & Kent, 2009, 2013), and many programs have limited connections to the actual business, leaving them hard to justify in the long-run (Kramer & Porter, 2011).

Based on this commentary, the purpose of this article is to unpack our social program evaluation process using examples of our work in sport that both support and run counter to creating shared value. We reveal our process from the design phase to the deliverable phase so future researchers and social program managers can better understand how cross-pollinating social and organizational value creates optimal conditions to achieve shared value.

**Observations from the Field**

We have seen an evolutionary shift in SR program deployment and the resulting scholarship dedicated to studying SR. Organizations in this context have been devoting financial and human resources to programs focused more on individual development, mainly among disadvantaged youth, via educational training, health and wellness support, and personal/individual (e.g., citizenship, self-efficacy, etc.) development (Walker & Kent, 2013). A number of international organizations are now delivering and supporting programs that blur the SR/Sport-for-Development (SFD) line. For example, Barclays bank uses sport as a vehicle to deploy SR via their ‘Spaces for Sport’ initiative that funds SFD initiatives in United Kingdom (UK) communities focused on youth employment and empowerment (Walker et al., 2015). The Korean Professional Soccer Organization (K-League) supports a physical activity program to bolster youth leadership. Similarly, the National Football League (NFL) promotes ‘Play 60’, a physical activity and character development program aimed at reducing the spread of childhood obesity and strengthening communities. These are a few of the programs operating under an SR umbrella yet use sport as either the direct mechanism to achieve developmental objectives.

On the SFD side, the United Nations Sport for Development and Peace(UNOSDP, 2014) initiative listed eight areas where sport contributes to society: (1) individual development, (2) health promotion, (3) gender equality, (4) social integration and social capital development, (5) peace building, (6) disaster relief, (7) economic development, and (8) social mobilization. While writing this article, this office discontinued its operations, but we digress. Interestingly, Salcines, Babiak, and Walters (2013) published a compendium of SR in sport articles where the contributors discussed each area noted by the UNOSDP. This observation is interesting since, at least at face value, sport organizations are increasingly promoting developmental outcomes within their SR campaigns. Despite this operational shift, the common practical omission has been the lack of evaluations to show whether social or development programs are working to achieve the desired ends. Moreover, research progression in both areas has been encumbered by limited foci. For example, research on SR programs has centered mainly on how strategic ‘paybacks’ (i.e., business-level, instrumental benefits) are realized for the focal organization (Walker & Kent 2009, 2013; Walker & Mercado, 2016; Walker & Parent, 2010). SFD evaluations, on the other hand, have reported positive beneficiary, community, and social returns (i.e., normative-level outcomes as suggested by Walker & Parent, 2010), mainly in the form of qualitative narratives. However, we have recently been engaged in a number of evaluation projects where the programs seek to both enhance developmental outcomes among the beneficiary groups (i.e., social value) while at the same time generate organizational value through outcome marketing.

The caveat here is that many of these programs operate under a SR umbrella, which means that researchers have much to learn from both SR and SFD literature(s) to inform future research progression. For example, beneficiary and community impacts are critical elements largely ignored in SR research, while ‘paybacks’ and leveraging opportunities studied by SR scholars could prove useful in the SFD discussion as funding channels become scarce. By measuring both social impacts and strategic payback, program evaluations have the opportunity “… to demonstrate the elusive win-win proposition of organizations not only benefitting themselves but also program beneficiaries” (Walker et al., 2015, p. 1). Such an approach moves towards creating shared value, which according to Porter and Kramer (2011), enhances organizational competitiveness, while simultaneously advancing economic and social conditions in communities where the organizations are located. We have been fortunate to work on the front lines of this debate by providing research and evaluations to a broad range of sport organizations around the world. Accordingly, and through a shared value lens, we now understand how the SR/SFD intersection can prove useful for program design and delivery. In this article, we offer a number of observations and some resulting strategies that can assist practitioners in social program development and deployment. Based on our work with social program managers, we offer our perspectives for bolstering program delivery, enhancing the outcomes realized for the beneficiary groups, and creating shared organizational value.

**Recommendations for Deliverers**

***Seek and Measure Shared Value***

Striking an appropriate balance between program motives and outcomes is critical to creating shared value. While many motives exist and the majority might seem well-intended, it is our opinion that few organizations truly want to become agents of change, at least on the SR side of the operational discussion. For example, a historic trend in SR reporting has been to characterize social impact in terms of dollars spent and/or the number of participants a program has engaged. In addition, research on SR in both sport and general business has consistently indicated that by acting in responsible ways and being perceived as responsible, business-level outcomes can materialize (e.g., image, reputation, purchases, and investing). While good for business in the short-run, this is not necessarily as good for society or business in the long-term since it misses the point of shared value creation.

On the outcome side, some of the most well intended social programs are seemingly a mile wide and an inch deep. Meaning, they attempt to tackle a lofty and often unrealistic social issue when the focus should more realistically/pragmatically reside on the micro-(i.e., individual), or meso (i.e., local community) as opposed to the macro-(i.e., societal) environment. The National Hockey League’s (NHL) ‘Hockey Fights Cancer’ and the National Basketball Association’s ‘NBA Cares’ programs are examples of these broad approaches. While cancer prevention, youth development, and family development are noteworthy causes, they are far too broad for one organization, let alone a professional sport organization, to meaningfully assuage. The only positive here is that professional sports have high levels of reach and visibility, which provides the power to attract media and public attention to the issues in order engender broad community awareness. However, this only cursorily provides social value. Additionally, many sport organizations support a wide range of social causes that disperse any relative impact that could potentially result from the social program. For example, Walker and Parent (2010) showed that of the ~N=100 sport organizations assessed, 89% supported two or more social initiatives (range=2-17 social programs). In order for social programs to be both practical and impactful, they should be conceptualized in terms of the scope, orientations, and both normative and instrumental outcomes. Not watered down versions of these areas. The latter point requires the shared value balance so both social and organizational value can collectively materialize. Another issue resides with the limited connectedness between the social programming and core business functions, which Kramer and Porter (2011) noted reduces program viability and the benefits realized. For example, reversing the trend of childhood obesity is not something that one American football organization alone can accomplish. It requires a significant global movement.

A tacit, yet imperfect, example of this motives balance is the National Football League’s (NFL) ‘Play 60’ program in London. Broadly, the program theme is to reduce the spread of childhood obesity. This school-based program functions by promoting active play for 60 minutes per day using American football as the sporting ‘hook’. Through American football, the NFL embeds character development education, reinforced by former NFL players and through homework assignments for program participants. These features of the program approach social value since childhood obesity is an issue in both the UK and United States and reinforcing and promoting character is quite socially acceptable. Organizational value was approached through branding and marketing opportunities for the NFL, as they reconceive their market in the UK and enable local clustering of their American football product in UK communities (Kramer & Porter, 2011). Of the six NFL teams that participated in the effort, each was assigned one school in greater London and the school was branded accordingly to reinforce the residual organizational value (i.e., marketing).

The NFL example characterizes the beginning stages of shared value. Our pre-post, mixed-method evaluation data (i.e., inclusive of control group data) indicated significant upticks in character development and physical activity enjoyment and engagement for the experimental groups. While impossible to imagine that steps to reduce childhood obesity were taken, the program nonetheless enhanced perceptions of and engagement with physical activity. However, the intervention lasted only 10-weeks and the NFL failed to pursue any long-term tracking of the significant effects after the program concluded to see if behavior change had materialized. In terms of organizational value, we saw significant differences for American football affinity and team affinity over the intervention period. Again, while programmatic effects were immediately documented, extended research was not conducted to see if ‘fanship’ resulted, or indicators of team loyalty (e.g., team-related purchase) manifested. Although longitudinal data were not collected post-intervention (e.g., weeks/months later), the short-term effects of organization value were enough for the NFL to expand the ‘Play 60’ program for another year, thus providing the opportunity for social value that would not exist if not paired with organization value.

A more conceited use of shared value was seen in Nike’s ‘Designed to Move’ physical activity agenda, which is a call to action for governments, cities, corporations, communities, schools, and individuals to inspire kids to move. The Nike brand was, however, distanced from this agenda in the UK presumably because they were not seeking to enhance their brand image. That is, Nike instead chose to collaborate with a local community center to deliver the physical activity content. It is plausible this connection drove Nike’s arms-length approach since it could have raised consumer cynicism about their SR endeavor (Sartore-Baldwin & Walker, 2011). While we may never truly understand their motivation, the Nike leadership in the UK felt they were best served by focusing on physical activity without risking a consumer backlash because of their non-altruistic motivations (i.e., high social value but low organizational value). That said, with the financial influence of Nike, social value can be achieved regardless of limited altruistic intent. Thus, the pursuit of shared value is still preferred over the pursuit of social value alone.

***Be Proactive, Not Reactive***

From our observations, many social programs have been designed with the ‘squeakiest wheel gets the grease’ mentality, which means that many organizations have employed a reactive as opposed to proactive approach. This has been due in part to public backlash or a socially perceived ill that has generated significant media attention. For example, the character education aspect of the NFL’s ‘Play 60’ was a reactive approach by the NFL as they responded to several scandals highlighting a lack of character among NFL players and coaches. The view from the NFL was to promote their core values to the program beneficiaries via NFL coaches, players, and former players as a way to reinforce this messaging. As an earlier example, in 1991-1992 Nike came under fire for reports detailing issues of low wages and poor working conditions in Indonesian factories where their clothing and other products were produced. Exacerbated by protests at the Barcelona Olympic Games in 1992, negative media attention was devoted to the retail giant over the next five years. Rather than immediately rectifying the issues, Nike expanded its operations and continued its poor worker treatment until 1998 when they announced laudable changes to production policy and employee treatment precipitated by widespread public denouncement of Nike products. This reactive case is important because Nike could have easily modified its structure to simply meet (i.e., rather than exceed) local standards for production and manufacturing in the regions where they conduct their business. Instead, they spearheaded sweeping ‘above and beyond’ change by creating the Fair Labor Association, a non-profit group of companies, and human rights and labor representatives that independently monitor conduct codes of global businesses. Since that time, Nike has conducted annual factory audits and regularly publishes SR reports detailing its commitments, standards, and data to support this movement. We use this example to show that a reactive position can segue to a proactive one if long-term commitment to change is adopted. Moreover, this case exemplifies the shared value position (Kramer & Porter, 2011), whereby Nike produced economic value by offsetting the societal challenges of low wages and poor working conditions in regions where such business practices were commonplace.

**Recommendations to Evaluators**

***Theorize Change***

The starting point for effective program design and evaluation is a ToC, which is a sequence of presumed causes, actions, processes, and effects (Weiss, 1997). There is substantial range to constructing a ToC, but even where deliverers have an understanding of the causes, actions, processes and effects, they are rarely if ever articulated in a written ToC. To facilitate the development of a ToC, we make use of the social marketing framework, which is a process that applies marketing techniques to create, communicate, and deliver value in order to influence target audience behaviors that benefit society (Kotler, Lee, & Rothschild, 2006). We start by identifying a social problem and its associated anti-social behavior (e.g., how obesity is linked to being sedentary). We then describe the mediating element(s) between this connection to achieve the socially-desirable behavior (e.g., sustained physical activity). This basic example illustrates the ‘book end’ approach, whereby the focus is on the processes that add value to the socially desirable behavior. As Breitbarth, Walzel, Anagnostopoulos, and van Eekeren (2015) rightly noted, the process-oriented extension to the traditional content-based approach is a critical next step for creating shared value via SR campaigns and strategies.

We conduct interviews with program designers and deliverers before we write the ToC, and map our preliminary assumptions and hypotheses against the academic literature. Making use of existing knowledge and evidence underpins the strengths and weaknesses of a ToC, allowing ongoing modifications to maximize program success. For example, while developing a ToC for a boxing-based intervention for youth excluded from school, there was an assumption that boxing served as a form of stress release, based on the catharsis theory of aggression. However, our review of the academic literature identified empirical evidence (see Bushman, 2002), which dismissed this assumption and suggested that punching a bag actually increased anger and aggression. As a result, the ToC was modified to align with existing research.

Despite its importance to the evaluation process, not all programs consider a ToC. For example, an SFD charity we worked with recently expanded their operations from India (i.e., Low to Middle Income Country – LMIC) to the UK (i.e., High Income Country – HIC) via a school-based intervention. As we have done at the outset of all our evaluations, we sought to establish a ToC as a blueprint for the evaluation. However, the charity did not have any program objectives, nor did they identify a social problem the program was seeking to assuage. Rather, they wished only to validate their delivery methodology in the HIC context, which was to tackle poverty. However, the type and scope of poverty in areas of India where the charity previously worked was not comparable to what is the case in London. It was, therefore, impossible to orient their methodology in a ToC with intentional design characteristics to address a social problem with interim impacts and broader outcomes (Bruening et al., 2015). Not surprisingly, the resulting evaluation produced limited significant effects on the research variables. A strong ToC would have informed a stronger program design for this charity, and further ensured the intervention was evaluated against variables the design was intentionally influencing.

***Design for Valid Conclusions***

Although program deliverers and funders want to know their program is working, a misguided approach is assuming a ‘cheerleader’ position of reporting positive results and failing to address programmatic shortcomings. Thorough social program evaluations should serve their clients by providing credible evidence-based impact, big or small, positive or negative, and should highlight the strengths and weaknesses that enable success or lack thereof. Without a complete understanding of all program nuances, deliverers and funders cannot improve their offering to achieve shared value. As mentioned at the outset of this article, each evaluation calls for an individualized approach to the research design, which may yield different results for the same program. Accordingly and when a program fails to influence the intended beneficiaries, identifying and remedying deployment features to enhance aggregate program effectiveness is still an important outcome for the evaluation.

For example, our initial evaluation of the NFL ‘Play 60’ in the UK revealed their use of former players and role models had little resonance with and influence on program participants in terms of character education. Since this was a new sport to the UK youth, and few of the participants knew of any former NFL players, role modeling simply did not occur. As a result, we recommended a change to this aspect to include UK-based individuals who were more familiar to the participants. As a result, character education was improved and the time, resources, and efforts by the NFL were eventually recouped as social value. The NFL wanted to use role modeling to bolster organizational value (e.g., marketing effects of high-profile former players) at the expense of any potential social value, which was a problematic approach from the outset.

Structural program elements aside, credible conclusions about program impact cannot be determined from isolated/limited research designs. For example, we have seen both SFD and SR program evaluations drawing conclusions of program impact based on qualitative data alone (e.g., interviews and focus groups). In our opinion, it is difficult to justify a claim of social impact (i.e., effect) based on just participant perceptions. For example, the extent of effect is more precisely measured numerically (e.g., an effect size of 0.78) rather than through words (e.g., “very much so”), making it possible to extrapolate results of effect from a sample to a population (Firestone, 1993). For these reasons, we rely on mixed-methodologies, with a quantitative focus on experimental designs, and a qualitative focus to enable triangulation (i.e., interpretation). The experimental approach enables the researcher greater control of the variables under investigation and the environment where the research variables are observed. This means that when the intervention is introduced, and the associated consequences measured, a more robust case for impact and shared value can be made. In order to make inferences about the effect of a given program, it is necessary to establish comparisons between pre- and post-program measurements, as well as between experimental and control groups. While physically impossible to create the perfect scenario of concurrent evaluations of the same individual in two conditions (Heckman & Smith, 1995), experimental designs offer the most defensible approach for any program evaluation. For example, pre- and post-intervention results establish change from a baseline position, while the experimental and control comparisons establish change when absent the program intervention. This helps the researcher dismiss confounding variables so the effects can be confidently attributed to the intervention.

Now, we are not saying that qualitative methods and cross-sectional designs have no place in program evaluations. Quite the opposite, actually. These methods and resulting data are critical to understanding the processes and mechanisms behind a desired effect. As such, we strongly recommend a mixed approach where experimental designs are paired with robust qualitative inquiry to reveal nuances not captured through purely statistical methods. Processes and mechanisms are subjective in nature and understood through interpretive research to “… make sense of or interpret phenomena in terms of the meanings people bring to them” (Denzin & Lincoln, 2011, p. 3). Furthermore, program mechanisms and processes do not exist in a vacuum, and require contextualizing through a set of material practices that show not just how something works but why something works. Through our work, we have noticed that the more robust our research design, the less positive our findings. In situations where we have evaluated programs post-intervention, and been restricted to qualitative data from interviews and focus groups, perceptions of program impact are largely positive. Thus, a feedback loop for programmatic improvement is non-existent.

For example, we evaluated a grassroots Nike program called ‘Move It’, which was based on the lager ‘Designed to Move’ agenda discussed earlier. We were initially tasked with evaluating the program post year-1 of the intervention but were only able access qualitative data, which restricted us to assess only perceptions of impact. Not ideally, we were not involved with program initiation, which meant a ToC was not formalized and a framework for shared value was not articulated. Moreover, we were limited to conducting post-hoc work, which is a less favorable situation for a program evaluator. Our findings suggested that ‘Move It’ created improved levels of fitness and emotional wellness, which participants and administrators attributed to the program. However, when evaluating year-2 of the program, we were able to collect pre- and post-intervention data, which told a completely different story. A significant difference between pre- and post-measurements was found for fitness, but not self-esteem, self-efficacy, or emotional wellness. The social intervention influenced personal fitness, as expected from keeping participants physically active, but the outcomes for which there was no intentional design were not impacted. In fact, our results contradicted those of the Nike-funded ‘Designed to Move’ research performed by Bailey, Hillman, Arent, and Petipas in 2013. The authors of this evaluation argued the benefits of physical activity were underestimated to extend beyond health to other aspects of the participants’ social lives. However, their evidence for emotional, financial, individual, intellectual, and social outcomes were based on qualitative data alone with no control group to account for external influences. This, however, did not prevent the authors over-selling the perceived value of physical activity, which they used to support their funder’s agenda. This evaluation is a perfect example of achieving organizational value at the expense of social value and over-extrapolating bracketed data as a ‘cheerleading’ endeavor.

Our evaluation of an Indian SFD charity’s initial venture into the UK market underscores the importance of gathering control group data to yield a robust understanding of programmatic influences. For example, participant focus groups revealed positive perceptions of improved conscientiousness, recognition of social bullying, diversity awareness, and enhanced teamwork. However, the pre/post-quantitative evaluation revealed significant and positive effects for only five out of fifteen research variables, none of which were those stated above. The picture became more grim when the results revealed that three of the significant effects were also seen among control group participants. Therefore, of the fifteen research variables examined, we were only able to attribute the intervention to achieving two outcomes. Interviews with teachers in the schools informed why changes were found in the control group, which was because the program duplicated alternative provisions already offered in the schools. This was an enlightening situation that reaffirmed our approach to mixed designs.

**Conclusions**

In sum, the examples herein illustrate how and why a shared value understanding is critical for program deployment. Identifying both social and organizational benefits, and mapping them against an informed ToC from the beginning, are crucial steps to a proper and ultimately robust program evaluation. Managers working in the social program space should seek evaluation partners who approach the research in a holistic and iterative manner. Working to ensure the program is successful, sustainable, and meets societal needs, while at the same time benefitting the organization, are key goals for the design and deployment of any social program. Moreover, program evaluators should use multiple and blended research designs and techniques to claim social impact but also tell the story behind the social impact.

References

Bailey, R., Hillman, C., Arent, S., & Petitpas, A. (2013). Physical activity: An underestimated investment in human capital? *Journal of Physical Activity and Health*, 10(3), 289-308.

Breitbarth, T., Walzel, S., Anagnostopoulos, C., & van Eekeren, F. (2015). Corporate social responsibility and governance in sport: “Oh, the things you can find, if you don’t stay behind!”. *Corporate Governance*, *15*(2), 254-273.

Bruening, J. E., Peachey, J. W., Evanovich, J. M., Fuller, R. D., Murty, C. J. C., Percy, V. E., Silverstein, L. A., & Chung, M. (2015). Managing sport for social change: The effects of intentional design and structure in a sport-based service learning initiative. *Sport Management Review*, *18*(1), 69-85.

Coalter, F. (2007). *A wider social role for sport: who's keeping the score?*. Routledge.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The Sage handbook of qualitative research*. Sage.

Firestone, W. A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational researcher*, *22*(4), 16-23.

Heckman, J. J., & Smith, J. A. (1995). Assessing the case for social experiments. *The Journal of Economic Perspectives*, *9*(2), 85-110.

Kotler, P., Lee, N.R., & Rothschild M. (2006). Cited in Cheng, H., Kotler, P., & Lee, N. (2011). *Social marketing for public health: global trends and success stories*. Jones & Bartlett Learning.

Kramer, M. R., & Porter, M. (2011). Creating shared value. *Harvard business review*, *89*(1/2), 62-77.

Levermore, R. (2011). The paucity of, and dilemma in, evaluating corporate social responsibility for development through sport. *Third World Quarterly*, 32(3), 551-569.

Salcines, J. L. P., Babiak, K., & Walters, G. (Eds.). (2013). *Routledge handbook of sport and corporate social responsibility*. Routledge.

Sartore-Baldwin, M., & Walker, M. (2011). The process of organizational identity: What are the roles of socially responsive behaviors, organizational image, and identification? *Journal of Sport Management,* 25(5), 489-505.

Walker, M., Heere, B., & Kim, C. (2013). The paradox of corporate social responsibility: Putting the “responsibility” back in CSR through program evaluation (p. 309-316). In J. Salcines, K. Babiak, & G. Walters (Eds.), *The Routledge Handbook of Sport and Corporate Social Responsibility.* London, UK: Taylor & Francis.

Walker, M., Hills, S., & Heere, B. (2015). Evaluating a socially responsible employment program: Beneficiary impacts and stakeholder perceptions. *Journal of Business Ethics,* 143(1), 53-70.

# Walker, M., & Kent, A. (2009). Do fans care? Assessing the influence of corporate social responsibility on consumer attitudes in the sport industry. Journal of Sport Management, 23(6), 743-769.

Walker, M., & Kent, A. (2013). The roles of credibility and social consciousness in the corporate philanthropy–consumer behavior relationship. *Journal of Business Ethics,* 116(2), 341-353.

Walker, M., & Mercado, H. U. (2016). Environmentally responsible value orientations: Perspectives from public assembly facility managers. *Corporate Social Responsibility and Environmental Management,* 23, 271-282.

Walker, M., & Parent, M. M. (2010). Toward an integrated framework of corporate social responsibility, responsiveness, and citizenship activities in sport. *Sport Management Review,* 13(3), 198-213.

Weiss, C. H. (1997). Theory‐based evaluation: Past, present, and future. *New directions for evaluation*, *1997*(76), 41-55.