The Importance of Theorizing Social Change in Sport for Development: A Case Study of Magic Bus in London

Stephen Hills
London Metropolitan University

Matthew Walker
University of North Texas

Marlene Dixon
Texas A&M University

For sport for development practitioners, a theory of change document is a critical first step to map how program inputs yield the desired program outcomes. Yet, in our experience, this document is rarely created in practice. Accordingly, this study makes use of the case of an award-winning sport for development charity that expanded their operations from India to London to illustrate the pejorative implications resulting from failing to create a theory of change. A mixed-methods, quasi-experimental approach was utilized to understand program mechanisms, program processes, and how these influenced the aggregate participant experience. The quantitative analyses yielded no significant effects. Triangulating the qualitative data revealed that personal, social, health, and economic education was a competing product to the program. The limited effects are attributed to a failure to identify and attempt to assuage a local social problem. In addition, alignments with stakeholder expectations, program context, and legal requirements were also derailing. A discussion of the results, implications, and recommendations for establishing and implementing a theory of change are provided.

Keywords: theory of change, social prescribing, social outcomes

The importance of a theory of change to guide a sport for development (SFD) program toward achieving its desired outcomes has been established in the literature (Coalter, 2007, 2012; Bruening et al., 2015; Walker, Hills, & Heere, 2017). For example, Walker et al. (2017) evaluated a soccer-based employability program delivered by an English Premier League club. The results from this work indicated that the lack of hard employment outcomes achieved was attributable to limited synergy between the program delivery and goals and a lack of direct ties between the desired outcomes and factors related to those outcomes. These results are not uncommon in the SFD space. Throughout its history, SFD has been encumbered by "overly romanticised, communitarian generalisations about the 'power' of sport for development" (Coalter, 2010, p. 1386) and dependence "on the supposed inherent properties of sport to achieve desired outcomes" (Coalter, 2013, p. 609). This historic SFD trend, while advancing, has resulted in failures to theorize how sport-based programs impart individual and societal change (Coalter, 2013; Hartmann, 2003; Kruse, 2006; Lyras & Welty Peachey, 2011; Schulenkorf, 2012). More recently, SFD academics have developed SFD theory, articulating the cause and effect between sport interventions and social change. However, social change theorizing has not aggregately extended to all SFD practitioners. To illustrate these implications, we present a case study of the Magic Bus Explorer program in the United Kingdom, to demonstrate how the intended beneficiary effects did not materialize because the theory of change was not established, as the social problem was misunderstood.

SFD Theory

When the desired social outcomes are mapped against program curriculum, sport can contribute to individual change on a number of social and psychological levels (Sugden, 2008). For example, Lyras and Welty Peachey (2011) attempted to theorize how sport interventions can effectively promote social change and development. They surmised that sport practices should be based on moral principles, mixed teams, and traditional and nontraditional sports, and that a variety of sport and physical activities should be provided to attract and sustain more representative populations. Moreover, SFD programs should be educational, and coaches and instructors should serve as agents of change. While informative for the SFD discussion, such generic principles for social interventions, not specific to sport, do not constitute a SFD theory (Hills, Gomez Velasquez, & Walker, 2018). This positionality reflects what Coalter (2015) labeled the problem of displacement of scope, whereby microlevel (i.e., program level) mechanisms are incorrectly generalized as macro-level SFD outcomes. Rather than attempt to produce an SFD "theory of everything" at a macro level, theorizing should occur at the program level, to understand

Hills is with the Guildhall School of Business and Law, London Metropolitan University, London, United Kingdom. Walker is with the Department of Management in the G. Brint Ryan College of Business at the University of North Texas, USA. Dixon is with the Department of Health and Kinesiology, Texas A&M University, College Station, TX, USA. Hills (S.Hills@londonmet.ac.uk) is the corresponding author.

what works for whom, when, and under what conditions (Coalter, 2007). In other words, context is everything for SFD practitioners.

Sport in itself does not cause social change. For SFD, this means that participation in sport is a necessary but not sufficient condition for change (Coalter, 2013). For example, sport may serve as the context for socialization opportunities, rather than the cause of socialization outcomes, or it may work as a hook to engage a target group before other program components treat the problem (Coakley, 1993; Green, 2008). That is not to say that sport cannot play a bigger role than being the context or the hook, as sport can be used as an analogy to replace other mechanisms/pathways of social development. Hills et al. (2018) demonstrated this very idea when describing the "Seedbeds of Peace" program in Medellin, Colombia, which uses soccer drills and games to teach life principles. This use of sport embeds meaning into a program's theory of change, placing sport within a conceptual framework that provides a direct relationship between sport and the desired social/life outcomes. In either case, and in addition to simply increasing participation in sport, there is a need to identify the sufficient conditions (i.e., mechanisms, processes, and experiences) through which an SFD intervention can result in contextually bracketed individual or social change (Coalter, 2015). According to Coalter (2012), without this understanding, we are left with a "black box" view of SFD interventions as being able to effectively change a range of values, attitudes, and behaviors, with little understanding of the processes behind any change. Accordingly, the need to understand the relationship between program concept, delivery, and outcomes is ever present in SFD practice (Coalter, 2012).

SFD Theory of Change

For an SFD intervention to achieve the desired outcomes, a theory of change is essential. According to Coalter (2012), a theory of behavior change or program theory is a sequence of causes and presumed effects underpinning interventions that articulate hypotheses about the relationships between social problems, participants, program mechanisms and processes, intermediate impacts (i.e., the effect on participants), and broader outcomes (i.e., individual behavioral or social changes). Similarly, Bruening et al. (2015) referred to the concept of intentional design, whereby program processes, mechanisms, and experiences add value to a desired behavior. Intentional design elements bridge social problems, span program mechanisms and processes, and should ultimately lead to broader outcomes. By mapping the causal factors important for success (World Bank, 2004) and understanding the relationship and pathways among program concept, delivery, and outcomes (Coalter, 2012), a theory of change serves two primary purposes: (a) informing program design and practice so to maximize the possibility of achieving the desired outcomes (Coalter, 2007) and (b) serving as a theoretical framework for evaluation (Coalter, 2012).

Coalter (2012) studied the mechanisms underpinning four SFD programs seeking to address gang membership, racism, at-risk youth, and conflict through in-depth interviews with participants, delving into their experiences and the program elements perceived to have had the greatest impact on their values, attitudes, and behavior. His study identified limitations with regard to the use of a theory of change in the programs. For example, the author found a lack of clarity in defining a target group and flawed at-risk assumptions about the group. In other words, the programs allowed participants to self-select as opposed to the program leaders targeting individuals for whom there was at-risk evidence. The programs recruited participants on the basis of an implicit deficit model based on an environmental fallacy that all youth from high-crime areas are at risk. Coalter (2012) found that there was little systematic analysis of the presumed deficits as a preprogram measure, creating a paradoxical danger of well-meaning projects being based on negative stereotypes, leading to misconceived provisions and inappropriate performance indicators. He also found that sport was presumed to play an important role in change but that the participants rarely mentioned the sport components when discussing the mechanisms that resulted in perceived changes. Rather, relationships with role models were reported as the key mechanism in achieving intended outcomes, whereby support was provided by role models beyond the sport setting, and relationships were built based upon respect, trust, and reciprocity.

Coalter (2012) advocated that a theory of change should start by defining desired outcomes and work backward, identifying the mechanisms that serve as causal factors, rather than assuming a cause-and-effect relationship between participation and desired outcomes. Similarly, Walker and Hills (2017) applied social marketing principles to show that a theory of change should start by identifying a substantial and measurable social need or problem. Working back from this social problem or localized need, SFD practitioners should identify a target group and their behaviors that are associated with the social problem. From there, corresponding behaviors that benefit participants or society should be established as desired outcomes, thus establishing "book ends" within which processes, mechanisms, and experiences can create, communicate, and attach value (i.e., meeting needs and removing barriers) to the desired behaviors, which replace the negative behaviors associated with the social problem.

The importance of following such a process is illustrated via a case study of Magic Bus. One limitation of Coalter's work is that he studied mechanisms across four programs, despite acknowledging that "given the diversity of participants, programmes, processes, relationships and desired outcomes it is not possible to develop a definitive or prescriptive programme theory—each programme requires its own programme theory to reflect its context" (Coalter, 2012, p. 607). This study advances work in the area by leveraging Coalter's arguments while focusing on a single case study.

Study Context

Originating in India, Magic Bus is an SFD charity that works with children to provide "a better life with better awareness, better life skills, and better opportunities" (Magic Bus, 2016). In 2015, Magic Bus expanded their operations to the United Kingdom to extend their work from the low-to-middle income country of India to the high-income country of the United Kingdom to help children make the right choices and have more control in their lives (Magic Bus UK, 2016). The Explorer program is the Magic Bus pilot program in the United Kingdom. As a school-based program, it is delivered within physical education (PE) lessons (i.e., a blend of play and sport-based) as a vehicle to teach and reinforce life lessons. Although the program was delivered in PE classes, it was not intended to deliver a physical activity curriculum. Rather, the designers of the program attempted to challenge the assumption that these classes can only be used to achieve sport-based outcomes. At the outset of the project, a needs analysis established that children from disadvantaged families in the London borough of Lambeth suffer from poor social skills and emotional management, which manifested in a negative school environment, a lack of cultural integration, and bullying. According to Magic Bus, poor social skills and emotional management in youth are detrimental to

development, which is amplified as pupils transition between school years and levels.

Within the Explorer program, sport and play are used as the hook to engage participants in social–emotional education, when they might otherwise be disengaged. The value neutrality of sport and play is leveraged, so tensions created by gender, class, and ethnicity are minimized (Green, 2008). Furthermore, the inherent fun in sport and play is used to ensure active beneficiary participation and motivation for task persistence. Games are used metaphorically, whereby hidden social–emotional messages are embedded into games, which are thought to be transferable to other contexts. According to Magic Bus, this approach adds a rhetorical effect to the message delivery, as it is first experienced and accepted as a game and then applied to other contexts (e.g., a classroom, playground, or home). Magic Bus believes that, if properly done, this process can increase the likelihood of the message or idea being internalized by participants, manifesting in new socially desirable attitudes.

To achieve the intended educational and social ends, a reflective (i.e., debriefing) exercise is undertaken after each session to discuss the embedded messages/ideas. Through direct questions from the coach, participants are encouraged to identify the messages hidden in the games, and then identify where and how they can be applied to other social contexts. Measuring the beneficiary effect and explaining the mechanisms, processes, and experiences behind the effects (or lack of) illustrate the challenge that SFD practitioners might face in establishing and implementing a theory of change. Accordingly, the following research questions guided the investigation:

Research Question 1: What effect did the Explorer program have on participant relationships (i.e., social competence, conflict management, diversity awareness and attitudes, and bullying)?

Research Question 2: What effect did the Explorer program have on participant goal-setting?

Research Question 3: What effect did the Explorer program have on broader outcomes (i.e., sense of community in school, and emotional wellness)?

Research Question 4: What Explorer program mechanisms, processes, and experiences explain the effects or lack of effects?

Methods

Procedure

We began the mixed-methods sequential design with quantitative methods to measure the influence of the Explorer program before moving to a second qualitative phase to explain the mechanisms, processes, and experiences behind the intended effects (Creswell, Plano, Clark, Gutmann, & Hanson, 2003). Data collection in the quantitative phase consisted of pre- (i.e., 1 week before) and postprogram (i.e., 1 week after) questionnaires. Data collection in the qualitative phase consisted of participant focus groups and administrator interviews. The relative strengths of the quantitative data were leveraged to answer research questions one through three, which measure program influence. Concurrently, the relative strengths of the qualitative data were leveraged to answer research question four, which detailed the mechanisms, processes, and experiences of the participants.

There are several benefits of utilizing a mixed-methods approach. First, statistical quantitative analyses allowed for the extrapolation and generalization of the findings from the sample to the population (Firestone, 1993), while the richness, detail, depth, and nuances that are inherent in qualitative data allowed for an understanding of the complexities in the mechanisms, processes, and experiences (Creswell, 2012). Second, the use of quantitative data provided a perspective based upon quantifiable and measurable "facts," while the qualitative analysis made sense of and interpreted the mechanisms, processes, and experiences in terms of the meanings participants and administrators brought to them (Denzin & Lincoln, 2011). Finally, the quantitative phase used an experimental design in an attempt to control and keep constant everything other than the treatment variable (i.e., the Explorer program) in order to strengthen conclusions about the effects (Burtless, 1995). In contrast, the qualitative phase used focus groups to study the students in their natural setting (Denzin & Lincoln, 2011).

The use of both qualitative and quantitative data provided "a plurality of interests, voices, and perspectives" (Greene & Caracelli, 1997, p. 14) to answer distinct research questions. Specifically, the quantitative data were used as part of a quasi-experimental design, so as to answer the research questions seeking to establish the *effect* of the Explorer program, but this does not answer *how* or *why* the treatment variables work (Deaton, 2010), leading to a black box view of causality (Imai, Keele, Tingley, & Yamamoto, 2011). Therefore, the qualitative data from the participant focus groups and administrator interviews were required to answer the final research question to understand the mechanisms, processes, and experiences that explain the effects or lack of effects.

Quantitative Technique

The Explorer participants and a control group of participants were asked to complete a questionnaire related to relationships, values, growth, and broader outcomes before and after the program. This approach allowed us to draw parallels between the pre- and posttreatment measurements, as well as between the experimental and control groups for the research variables. This way, any change from pre- to posttreatment could be attributed to the program if the change was not present in the control group (Hakim, 2000). The participants were nonrandomly assigned to the experimental and control group conditions. The program was delivered in school and consisted of two classes per year. Randomly assigning one class to the experimental condition and another class to the control condition in order to eliminate selection bias was not feasible due to participation target numbers. Therefore, a control group was constructed from a school that was planning to participate in the program the following year. According to Sefton, Byford, McDaid, Hills, and Knapp (2002), constructing a control group from a waiting list limits selection bias and is a valid means of constructing a nonrandom control group when a randomized means of construction is not possible. In this case, both the experimental and control groups were targets of the intervention, but they experienced it at different points in time. Furthermore, as an in-school intervention where recruitment and participation goes through the teachers and parents, the participants did not volunteer or self-select for the program, removing a source of selection bias (Hakim, 2000). In total, four classes from two schools (i.e., two classes per school) formed the experimental group, and one class from another school formed the control group. All three schools were in close proximity within the London borough of Lambeth. The pre- and postprogram questionnaire data were analyzed using

descriptive statistics, paired samples *t* tests to measure within-subjects effects independently (i.e., time; pre- to postprogram), and a two-way, mixed analysis of variance to measure the interaction of within-subject effects (i.e., time; pre- to postprogram) and between-subjects effects (i.e., condition; experimental vs. control). Although the between-subjects analysis provided the requisite information to assess the influence of the intervention, the within-subject effects are presented illustratively to demonstrate the different picture when ignoring contextual factors that are revealed by changes in a control group. In addition to the main effects, effect sizes for the significant results were used to determine the level of practical significance.

Quantitative Measures

Relationships. The effects on beneficiary relationships were measured using the following constructs: (a) social competence—defined as the ability to achieve goals, appreciate the perspectives of others, establish and maintain positive relationships, and handle interpersonal situations constructively (Harter, 1982); (b) conflict management—defined as the behavioral and attitudinal disposition taken when responding to a disagreement (Rubenstein & Feldman, 1993); (c) diversity awareness and attitudes—defined as respect and appreciation for diversity leading to a reduction in stereotypes and discrimination (Larke, 1990); and (d) bullying—defined as when someone is deliberately and repeatedly hurting or frightening someone weaker than themselves for no good reason (Bond, Wolfe, Tollit, Butler, & Patton, 2007). The following scales were used: social competence (five items, $\alpha = .81$; Harter, 1982), conflict management (eight items, $\alpha = .65$; Rubenstein & Feldman, 1993), diversity awareness and attitudes (three items, $\alpha = .70$; Larke, 1990), and bullying (six items, $\alpha = .81$; Bond et al., 2007). An item was dropped from the conflict management scale to improve internal consistency.

Goal-setting. Goal-setting, defined as the developing of an action plan designed to motivate and guide a person or group toward a goal, was measured using Hansen, Larson, and Dworkin's (2003) goal-setting scale (four items, $\alpha = .63$).

Broader outcomes. Broader outcomes was measured using the constructs of (a) sense of community in school, defined as a reflection of needs fulfillment, membership in a group, influence, and emotional connections (Vieno, Santinello, Pastore, & Perkins, 2007) and (b) emotional wellness, defined as one's ability to handle emotions constructively, enabling a positive emotional state (Hansen et al., 2003). The following scales were used: sense of community in school (six items, $\alpha = .70$; Vieno et al., 2007) and emotional wellness (five items, $\alpha = .63$; Hansen et al., 2003). All of the questionnaire items were adapted for age appropriateness and were pilot tested prior to the evaluation taking place (See Table 1 for the constructs and sample items).

Variables lacking internal consistency. Due to a lack of internal consistency, the data were collected but not used for the following variables: prosocial norms (three items, $\alpha = .50$; from Hansen et al., 2003), sportspersonship (three items, $\alpha = .52$; from Vallerand, Deshaies, Cuerrier, BriÈre, & Pelletier, 1996), aspirations (three items, $\alpha = .48$; from Meuleners, Lee, Binns, & Lower, 2003), help-seeking (four items, $\alpha = .32$; from Ryan & Pintrich, 1997), and psychological quality of life (three items, $\alpha = .40$; from Meuleners et al., 2003). Although the standard criteria for sufficient internal consistency is a Cronbach's alpha equal to or greater than .7, this study accepted scales with alphas as low as .63 on the basis that the scales used contained a limited number of items (e.g., goal-setting,

four items, $\alpha = .63$) and that alpha depends on the number of items in a scale (Nunnally & Bernstein, 1994; Sijtsma, 2009). Therefore, we argue that a limited number of items in the scales limited the scales' alpha, and the fit of the measures to this particular age group could have mildly limited the comprehension of some items.

Qualitative Technique

Four participant focus groups (i.e., one per participating class, two per school) were conducted. There were four boys and four girls selected for each focus group, which were conducted in a separate classroom in the presence of a teaching assistant, meeting the recommendation of 6–10 participants by Morgan (1996). The initial questions focused on general effect, making use of open questions (e.g., *What have you learned from participating in the Magic Bus program?*) before probing questions sought to establish the mechanisms behind the effects (e.g., *How did you learn this?*). Next, the questions delved into the specific beneficiary impacts (e.g., *Tell me about bullying in your school Has this changed in any way since participating in the program?*), which were followed with probing questions for cases where change was reported (e.g., *Why is there less bullying than there was before?*). Finally, general questions about delivery were asked (e.g., *What was your general experience of Magic Bus? What parts of the program benefited you the most?*).

Six semistructured teacher interviews, lasting approximately 45 min, were also conducted (i.e., three per participating school, two with class teachers, one with head teacher). The initial questions focused on their general perceptions of the program, making use of open questions (e.g., *What is your general perception of the success of the program?*) before evaluative questions about program delivery (e.g., *What challenges did the program face in being delivered in your school? What opportunities exist that the program could take advantage of?*) to help Magic Bus refine its delivery protocols in the future. Finally, two semistructured administrator interviews, lasting approximately 1 hr each, were conducted with the program designer and the program coach. These interviews sought to probe the theory of change behind the Explorer program. The initial questions focused on the social problem and desired outcomes (e.g., *How would you describe the social problem or problems that the Explorer program was seeking to tackle? Why was the social problem or problems important to Magic Bus? What were the desired outcomes of the Explorer program?*), before probing about how the mechanisms and processes were designed to achieve desired outcomes (e.g., *How was Explorer designed to achieve its desired outcomes? What differentiates the Magic Bus methodology from other developmental endeavors?*).

Finally, evaluative questions about program delivery were posed (e.g., *What challenges were encountered in the delivery of Explorer? How would you improve Explorer in a subsequent delivery?*). Collecting qualitative data from participants, teachers, and administrators provided a triangulation of perspectives.

The focus groups and semistructured interviews were professionally transcribed. The data were then thematically analyzed to uncover salient themes (Attride-Stirling, 2001). An inductive coding strategy was used, whereby themes and quotes were extracted, focusing on the mechanisms, processes, and experiences of the program. Initially, two researchers individually analyzed the transcriptions to reduce researcher bias (Maxwell, 2012) using a line-by-line open coding procedure to "expose the thoughts, ideas, and meanings contained therein" (Strauss & Corbin, 1998, p. 102). Next, axial coding was undertaken that sorted, synthesized,

organized, and grouped the data into larger abstract categories (Creswell & Miller, 2000), reassembling or reconstructing the data in a meaningful or comprehensible fashion (Jorgensen, 2015) so as to bring order to the data (Brewer, 2000) in a way that answers the final research question. The two researchers read each other's memos and compared the coding and categories before reaching a consensus on the themes and representative quotes.

Results

Quantitative Results

A total of 157 individuals (134 treatment group participants; 23 control group individuals) completed the pre- and postprogram questionnaires. The ages ranged from 6 to 10 years (Treatment: $M = 8.38$, $SD = 1.00$; Control: $M = 7.65$, $SD = .49$). The ethnicity of the sample groups was diverse (Treatment: White, 31%; mixed, 31%; Asian, 6%; Black, 26%; Control: White, 13%; mixed, 34%; Asian, 4%; Black, 48%), as was the gender of the groups (Treatment: boys, 51%; girls, 49%; Control: boys, 48%; girls, 52%). The participant year in school was well distributed in the experimental group, but the control group was constructed from a single class in a single year (Treatment: $n = 45$ in Year 3; $n = 42$ in Year 4; $n = 47$ in Year 5; Control: $n = 23$ in Year 3). The number of household members ranged from two (i.e., single parent) to eight (Treatment: $M = 4.29$, $SD = 1.44$; Control: $M = 3.61$, $SD = 1.23$).

Independent sample t tests were used to descriptively test for differences between the participants and the control group on preprogram measures in order to characterize the strength of causal inferences. There was no statistically significant difference in the preprogram measures for six of the research variables: social competence, $t(153) = -1.214$, $p = .227$; conflict management, $t(146) = -.247$, $p = .805$; diversity awareness and attitudes, $t(100) = 1.96$, $p = .053$; goal-setting, $t(151) = .588$, $p = .558$; sense of community in school, $t(152) = .989$, $p = .324$; and emotional wellness, $t(153) = .356$, $p = .722$. However, there was a significant difference between the participants and the control group for bullying, $t(155) = -3.88$, $p = .000$. As such, it is not possible to make causal inference claims for bullying because any difference (or absence of) may be due to differences between the children in the participant and control groups, rather than the program.

Paired samples t tests. Descriptive statistics were calculated (see Table 2), and a series of paired samples t tests were conducted to test the effect of time (i.e., pre to post) on the dependent variables. Dependent variable (i.e., relationships, values, growth, and broader outcomes) means were compared between the pre- and postprogram measurements (i.e., time; the within-subjects independent variable). The following analyses are conceptually grouped according to the research questions (i.e., relationships, goal-setting, and broader outcomes effects), and a report of all the effects appears in Table 3.

For Research Question 1, time had no significant effect on social competence, $t(127) = .036$, $p = .972$; conflict management, $t(120) = .152$, $p = .880$; diversity awareness and attitudes, $t(53) = .036$, $p = .539$; or bullying, $t(131) = .787$, $p = .433$. Descriptive

Table 1 Constructs and Example Items

Variable	Example items
Social competence[a]	I have a lot of friends.
	I am easy to like.
Conflict management[b]	[When you disagree with a parent or friend about something important to you, how often do you . . .] Get mad and start yelling [*Reversed item*]
	[When you disagree with a parent or friend about something important to you, how often do you . . .] Try to work out a solution so everyone is happy
Diversity awareness and attitudes[a]	It is important to understand different backgrounds.
	I prefer to socialize with children from the same background as me. [*Reversed item*]
Bullying[b]	[How often did any of these things happen to you at school this year?] Called hurtful names
	[How often did any of these things happen to you at school this year?] Kicked or hit
Prosocial norms[b]	I think about helping others.
	I think about changing my school and community for the better.
Sportspersonship[b]	I respect the rules when I play sport and games.
	I obey the referee or my teachers when I play sport or games.
Aspirations[a]	I like learning new things.
	I like getting a good education.
Goal-setting[b]	I set goals for myself.
	I think about ways to achieve my goals.
Help-seeking[b]	If I need help doing something, I skip it. [*Reversed item*]
	I think that asking questions helps me learn more.
Sense of community in school[a]	I belong at my school.
	When I need extra help, I get it from my teacher.
Psychological quality of life[a]	I feel good about myself.
	I feel good about my future.
Emotional wellness[a]	[I am able to. . .] Keep myself from feeling sad
	[I am able to. . .] Look for something good in a bad situation

Note. [a]Anchored by a 5-point scale from 1 = *strongly disagree* to 5 = *strongly agree*. [b]Anchored by a 5-point scale from 1 = *never* to 5 = *often*.

Table 2   Independent Samples' *t* Test

Preprogram variables	Standard error (df) Treatment → control	95% confidence interval Lower	95% confidence interval Upper	Treatment → control t value[a]	Treatment → control p value
Social competence	0.192 (153)	−0.614	0.147	−1.214	.227
Conflict management	0.159 (146)	−0.354	0.276	−0.247	.805
Diversity awareness and attitudes	0.167 (100)	−0.004	0.660	1.960	.053
Bullying	0.240 (155)	−1.404	−0.457	−3.880	.000[b]
Goal-setting	0.225 (151)	−0.312	0.576	0.588	.558
Sense of community in school	0.141 (152)	−0.139	0.419	0.989	.324
Emotional wellness	0.207 (153)	−0.336	0.484	0.356	.722

Note. [a]Absolute value of the *t* statistic. [b]A significant difference.

Table 3   Paired Samples' *t* Test

Paired variables	Standard error (df) Pre → post	95% confidence interval Lower	95% confidence interval Upper	Pre → post t value[a]	Pre → post p value	Effect size r
Pair 1: Social competence	0.088 (127)	−0.170	0.177	0.036	.972	–
Pair 2: Conflict management	0.075 (120)	−0.137	0.160	0.152	.880	–
Pair 3: Diversity awareness and attitudes	0.100 (53)	−0.138	0.262	0.619	.539	–
Pair 4: Bullying	0.099 (131)	−0.119	0.275	0.787	.433	–
Pair 5: Goal-setting	0.104 (122)	−0.750	−0.339	−5.249	.000[b]	.270
Pair 6: Sense of community in school	0.061 (128)	−0.126	0.115	−0.085	.932	–
Pair 7: Emotional wellness	0.077 (130)	−0.248	0.155	−1.843	.068	–

Note. [a]Absolute value of the *t* statistic. [b]A positive and significant difference.
r = .1 is a small effect, *r* = .3 is a moderate effect, and *r* = .5 is a large effect.

analyses showed that preprogram means for all relationships variables were positive, exceeding the midpoint of 3 on the 5-point Likert scale (social competence: $M = 3.86$; conflict management: $M = 3.64$; diversity awareness and attitudes: $M = 4.50$). Bullying was measured as a count of act of bullying; thus, there was no midpoint. However, $M = 2.16$ acts of bullying in a year can be considered low; thus, it is also already positive.

For Research Question 2, time had a significant and moderate effect on goal-setting, $t(122) = −5.249$, $p = .000$, $r = .270$. Descriptive analyses showed that the preprogram mean for goal-setting was already positive, exceeding the midpoint on the 5-point Likert scale ($M = 3.12$).

For Research Question 3, time had no significant effect on the sense of community in school, $t(128) = −.085$, $p = .932$, or emotional wellness, $t(130) = −1.843$, $p = .068$. Descriptive analyses

showed that the preprogram means for both broader outcomes variables were already positive, exceeding the midpoint of 3 on the 5-point Likert scale (sense of community in school: $M = 4.28$; emotional wellness: $M = 3.50$). In sum, only one (i.e., goal-setting) out of the seven variables changed from pre to post. All six Likert-scale variables had preprogram means above the midpoint of 3, two of which had preprogram means above 4 out of 5.

Analysis of variance.   Descriptive statistics were calculated (see Table 2), and a series of repeated-measures analysis of variance analyses were conducted in order to test the Time × Condition interaction on the dependent variables (see Tables 4 and 5). The dependent variable (i.e., relationships, goal-setting, and broader outcomes) means were compared between the pre- and postprogram measurements (i.e., time; the within-subject independent

Table 4   Treatment and Control Groups' Descriptive Statistics

Constructs	Preprogram (treatment), M (SD)	Postprogram (treatment), M (SD)	Preprogram (control), M (SD)	Posprogram (control), M (SD)
Social competence	3.86 (.84)	3.83 (1.00)	4.10 (.90)	3.78 (.88)
Conflict management	3.64 (.70)	3.63 (.63)	3.68 (.71)	3.72 (.60)
Diversity awareness and attitudes	4.50 (.66)	4.60 (.64)	4.17 (.86)	4.45 (.94)
Bullying	2.16 (1.04)	2.08 (.80)	3.09 (1.16)	2.65 (.89)
Goal-setting	3.12 (1.03)	3.63 (.77)	2.99 (.76)	3.56 (.75)
Sense of community in school	4.28 (.60)	4.29 (.66)	4.14 (.77)	4.07 (.65)
Emotional wellness	3.50 (.90)	3.64 (.74)	3.43 (1.03)	3.58 (.82)

Table 5 Two-Way, Mixed Analysis of Variance

Time × Condition	F value (df)	p value	d
Social competence	1.97 (1)	.16	–
Conflict management	0.076 (1)	.78	–
Diversity awareness and attitudes	2.06 (1)	.16	–
Bullying	1.18 (1)	.28	–
Goal-setting	0.002 (1)	.97	–
Sense of community in school	0.183 (1)	.67	–
Emotional wellness	0.005 (1)	.94	–

Note. [a]A negative and significant interaction.
$d = 0.2$ is a small effect, $d = 0.5$ is a moderate effect, and $d = 0.8$ is a large effect.

variable) and between the experimental and control groups (i.e., condition; the between-subjects independent variable). We conceptually grouped the analyses according to the research questions.

For Research Question 1, there was not a significant time × interaction for social competence, $F(1) = 1.97$, $p = .16$; conflict management, $F(1) = 0.76$, $p = .78$; diversity awareness and attitudes, $F(1) = 2.06$, $p = .16$; or bullying, $F(1) = 1.18$, $p = .28$. For Research Question 2, there was not a significant time × interaction for goal-setting, $F(3) = 0.002$, $p = .97$. For Research Question 3, there was not a significant time × interaction for the sense of community in school, $F(1) = 0.183$, $p = .67$, or emotional wellness, $F(1) = 0.005$, $p = .94$. In sum, there were zero significant and positive effects for the interactions. Time had no measurable effect on help-seeking, but there was a significant negative Time × Condition interaction. There was a statistically significant difference between conditions on the preprogram measure of bullying. Therefore, it is possible that the lack of effect on bullying could be due to the differences between the participants in the experimental and control groups rather than due to the lack of effect of the Explorer program. There was no significant Time × Condition interaction for goal-setting, for which a time effect had previously been found because there was also a positive and significant change from pre- to postprogram in the control group, suggesting that something other than the program is causing this change.

Qualitative Results

Theme #1: A lack of focus on a clearly defined and measurable social need. When adopting a social marketing approach, a theory of change should commence with identifying and defining a clear and measurable social need or problem for an intervention to focus upon (Walker & Hills, 2017). However, despite undertaking a needs assessment with the schools prior to the program implementation, Magic Bus leaders failed to identify (or agree upon) a consistent focus for the program. For example, in our initial discussions with the program leaders, the Magic Bus personnel listed in excess of 20 outcomes they wished to achieve. Following much discussion, the list was reduced to 12 outcomes and corresponding measurement variables. These outcomes were then grouped into three interim impact areas of relationships, values, and growth, which in turn, were mapped against the broader outcomes of sense of community in school, psychological quality of life, and emotional wellness. Despite this reorganization and reduction, it remained difficult to identify the core mission of the Explorer program in relation to overcoming a specific social problem. This was evident in the administrators being unable to clearly articulate a social problem that the program was intended to

assuage. Rather, in responding to the interview question of what social problem was being tackled, there was a tendency to list several outcomes without identifying a corresponding social problem. For example, the program coach stated, "The social problems that the program was seeking to tackle were lack of confidence, self-esteem, fear of failure, fear of mistakes, and fear of socializing." Similarly, the program administrator stated the following:

> The problems that Explorer was seeking to tackle was pupils' self-confidence and the emotional well-being of the children. In the schools we spoke to, they identified problems with bullying, anger management, and attention-seeking. . . . The key outcomes were increased resilience, self-efficacy and confidence, [and] prosocial behavior.

In these statements alone, nine different problems and four different outcomes were mentioned as being core to the program, with the only actual overlap being a lack of self-confidence. Without clear agreement on the core needs of the participants, the program struggled to maintain a focus on how to meet those needs. From our view, the main priority of Magic Bus was to validate their delivery methodology in the new context of London, rather than understand the needs of a specific group in need of an intervention. Indeed, Magic Bus proposed a list of outcomes against which they had achieved success in areas of India where they had previously worked. However, the preprogram measurements, which averaged above the midpoint for all of the research variables, suggest that participants were not in need of an intervention focused on these outcomes. Hence, a lack of program effect was inevitable from the start. As a result, by prioritizing validation and failing to consider context, Magic Bus gained little credibility in the eyes of the London schools.

Theme #2: Misalignment between program goals. A theory of change should provide a road map of the mechanisms and processes and experiences that move participants from an undesirable behavior to a desirable behavior, effectively applying available resources to best achieve desired outcomes. However, having distilled the long list of desired outcomes to 12, there was evidence of "outcome creep." That is, we observed that the program curriculum contained sessions related to 15 outcomes, six of which (i.e., self-awareness, overcoming obstacles and challenges, teamwork, appreciating self, peer pressure, and respect) were not directly aligned to the initial program outcomes. Furthermore, the curriculum did not directly address a sense of community. The program was delivered over a 35-week period, with 35 sessions of 45 min each. Although long, relative to other interventions, the amount of direct contact with participants was approximately 26 hr. In that time, 15 outcomes were excessive, spreading resources thin. At the high end, emotional awareness and diversity awareness and attitudes received 3 hr of intervention, but on the low end, sportspersonship and respect received 45 min of intervention. With such limited contact, it is not surprising that the program had a statistically significant effect on only one variable out of 12.

Theme #3: Lack of program clarity. A third issue within the program was a lack of clarity of the program as being directed at PE, sport, or social and health education. The Explorer program was delivered within the time allocated for PE, but its desired outcomes (e.g., social competence, conflict management, diversity awareness and attitudes, aspirations, health-seeking, psychological quality of life, and emotional wellness) were more aligned with personal, social, and health education (PSHE). As a result, the schools evaluated the program differently in terms of its capability

to meet curriculum standards for both PE and PSHE, which impacted their satisfaction. For example, the head teacher of the second host school stated the following:

> When I initially looked at the program, we talked about the PSHE side of it and the value that the children would get and whether they would actually get enough sport as well as the PSHE, and we have worked very closely with Magic Bus to develop the program and ensure it ticks both of those boxes, and it has worked really well. . . .

> You're supposed to have two hours of PE and thirty minutes of PSHE per week, so the challenge was how to sufficiently cover them. We still do additional sport and additional PSHE, but what I didn't want to do was replace one with the other because, initially, the children were doing a lot of games that weren't covering everything we needed to do in our curriculum, so we had to sit down and say these are the requirements we need to cover for sport and we talked about the programme for Magic Bus and what they wanted to deliver, and we had to tweak the two in order to meet.

Thus, the school was able to tweak the Magic Bus program to cover their PE needs; although these "tweaks" seemed small, it is not immediately clear how much the program integrity was compromised by the variation.

Although School 2 was able to tweak the program to cover their PE needs, School 1 was expecting a more traditional PE lesson with only some elements of PSHE education. As such, they were not satisfied with Explorer's balance between sport-based outcomes and PSHE outcomes. A teacher from the school stated the following:

> We met with Magic Bus at the start, but we didn't understand that it would be as much PSHE as it has been. It's not working for us, and I wouldn't recommend it to another school because we need our PE lesson to be PE and time is so limited. Combining PHSE and PE means they're not getting enough PE. . . . Initially, we thought it was a PE program with some PSHE in it, and I feel like they are trying to make it more PSHE, and the PE isn't done particularly well because they are trying to fit PSHE into it. . . . They're trying to do both things [*PE and PSHE*] and because they're trying to do both things, I think they're doing both things not very well.

This lack of satisfaction can be understood when analyzing national curriculum requirements. The national curriculum for PE aims to ensure that all students "develop competence to excel in a broad range of physical activities, are physically active for sustained periods of time, engage in competitive sports and activities, and lead healthy, active lives" (Department for Education, 2013a, para. 2). These sport-based outcomes are legal requirements that schools must meet, which led to the dissatisfaction of the first school. A teacher from the first school stated, "There is less PE and less sport happening than there should be. . . . It's all with very good intentions, but I would just like a PE session."

Although they did not express dissatisfaction, the second school confirmed that the Explorer program did not meet curriculum sport-based outcomes. The head teacher of the second school stated, "Whether it's improved their sporting ability, I would question." A teacher from the first school was also dissatisfied with the lack of teaching mainstream sports:

> I want to do a block of learning how to play cricket and a block of learning how to play basketball. . . . I don't think you should

leave primary school not knowing the rules for some of these key sports. You should know how to play these games so that you can develop an interest in them. That's what they should be learning, rather than PSHE games.

Another teacher from the first school reported that the lack of sport resulted in dissatisfaction of some parents, which is very problematic for an SFD intervention in terms of sustainable access to a target group:

> All my class mentioned that they don't do any specific sports in the school and parents have asked why are their children not playing a range of different sports and developing a range of skills. They're doing a lot of passing and catching, but other sporting skills are not being developed that much.

Whereas the PE curriculum prescribed by the Department for Education constitutes a legal requirement, greater flexibility exists for PSHE education. The Department for Education describes PSHE education as an important and necessary part of all pupils' education that should be taught by all schools, but which is a nonstatutory subject (Department for Education, 2013b). The Department for Education provides schools with flexibility in their delivery of PSHE because it views them as being best placed to understand the needs of their pupils (Department for Education, 2013b). Despite this flexibility, the first school felt that the Explorer program did not meet all of their PSHE requirements:

> It's not like it ticks off your PSHE. It doesn't cover most of your PSHE. It doesn't cover your sex education and drugs education. It doesn't really cover any of the stuff. It covers a very small part of the PSHE. . . . There's only so much PHSE that can be related to it [Magic Bus]. It's usually teamwork or they've tried relating hygiene a couple of times.

Although a lack of understanding of the context whereby the legal curriculum requirements were not fully met does not necessarily explain the lack of effect against the research variables, this flaw limits the sustainability of the Explorer program.

General Discussion

The influence of the Magic Bus program in the United Kingdom was assessed using several variables. Within-subject analyses established that time significantly influenced goal-setting. However, an improvement in goal-setting was also present in the control group, suggesting that other social/situational factors were influencing this outcome. Qualitative findings revealed similarities between the Explorer program and PSHE education, which was also present in the control group. This structural aspect helps explain the lack of Time × Condition effect because Magic Bus failed to identify a social problem that required an intervention. Rather, the program deliverers were largely duplicating a PSHE education intervention that was already compulsory within the British education system. In addition, the program deliverers were attempting to validate their methodology against outcomes for which the participants were not devoid, as evident in the high preprogram variable measurements. This aspect of program deployment runs counter to the recommendation of applying social marketing techniques or a theory of change to create, communicate, and deliver value to society (Coalter, 2012; Walker & Hills, 2017). This is an especially important consideration given that the qualitative analysis established that no additional value above PSHE

education was achieved through the intervention. That stated, we do not necessarily view this as a bad thing, as an educationally "additive" program is certainly one approach to SFD. However, PSHE bolstering was not one of the stated/intended program goals.

Along these lines, we further established through the qualitative analyses that a misalignment between overambitious curriculum and the perception of the intended social/psychological outcomes was present. This was a direct result of not having a clearly articulated theory of change, as several critical elements of the curriculum were not directly measured or addressed in the curriculum. In addition, we observed a lack of intentional design and rigor in articulating the relationships between actions, processes, and effects (Coalter, 2012). The combination of omissions is a certain contributor to the lack of program influence on the beneficiary group. Finally, through the qualitative analyses, we also established failures to (a) understand and align stakeholder expectations, (b) understand the context, and (c) meet the legal curriculum requirements in the community where Magic Bus was working. Coalter (2015) identified participation in sport as a necessary but not sufficient condition for SFD interventions to achieve social change. He further argued that mechanisms, processes, and experiences are sufficient conditions. The fact that Magic Bus failed in the above-mentioned areas demonstrates a lack of appreciation of the experiences and needs of the participating schools. They were attempting to validate a methodology that proved effective in India without sufficient appreciation of their new context, in particular, different expectations and different legal curriculum requirements.

Theoretical Implications

Previous literature established the importance of a theory of change in SFD (Coalter, 2007, 2012; Bruening et al., 2015; Walker et al., 2017), but prior to this study, there had been limited empirical research to evidence such a necessity. Although Coalter (2012) studied the mechanisms underpinning four SFD interventions from which theory of change weaknesses were established, the author did not map theory of change weaknesses against program effects. Like Coalter (2012), this study established theory of change weaknesses, but went a step further by quantitatively measuring program effects via a quasi-experimental design and linking the theory of change weaknesses to the absence of program effects.

Practical Implications

The absence of quantitative effects can be understood in terms of the absence of a robust theory of change, as evidenced in the qualitative data, thus establishing the importance of a theory of change. Although some of the findings in this study are specific to Magic Bus, this case study does serve to illustrate lessons that can be learned in SFD design generally. First, it is critical that SFD interventions operate within the bounds of a robust theory of change, which should be documented and understood by all stakeholders. The starting point should be a substantial and measurable social problem. An intervention should be targeted at a genuine social/community need, which can be confirmed in pre–post program measurements. For example, consider an education intervention targeted at an underserved population. For this type of intervention, the specific educational outcomes must be clearly defined (e.g., increased knowledge vs. attendance vs. advancement) and appropriate measures determined. For example, for improved academic performance, official exam scores or grades

would be a better measure than a Likert-scale measure of self-perceived academic capabilities.

Second, within the social marketing framework, practitioners should consider competing behaviors (i.e., continuing with the undesirable behavior that contributes to the social problem, rather than adopting the desirable behavior), but also competing products and interventions, such as PSHE education. If competing products or interventions exist, practitioners should be confident that their intervention adds value beyond the structural and educational elements that already exist in the community.

Third, desired outcomes should lead to alleviating a social problem and should be achievable. The relationships between intermediate impacts and broader outcomes should be mapped, documented, and based upon credible evidence where available. As one example, a regression model establishing that intermediate impact X is strongly associated with broader outcome Y when controlling for potential confounding variables is one important method. As another example, a randomized controlled trial would tease out differences between experimental and control conditions when attempting to establish a cause-and-effect relationship between mechanism X and intermediate impact Y. This approach provides internal validity for a theory of change. For external validity, however, contextual factors (e.g., legal requirements) should be considered in the theory of change. SFD stakeholders with an understanding of contextual factors should be engaged to assist with this process, which would also serve as a means to communicate expectations and provide opportunities to make adjustments so all objectives are correctly aligned.

Finally, we also illustrated the importance of a control group in evaluating SFD interventions. Had a pre- and postexperimental design without a control group been conducted, there would be a risk that an invalid causal inference would have been made with regard to the four variables for which time had a significant effect. Only when the Time × Control interaction was tested did a more valid understanding of program effect emerge.

With regard to the practical implications specific to Magic Bus, from the contextual understanding we provided in the study, it is clear the Explorer program should be repositioned in the United Kingdom as an additive in-school intervention. Furthermore, it should be retooled in this context in order to meet each school's curricular requirements. A major challenge is that Explorer is currently delivered in place of PE classes, but is not attempting to meet the sport-based outcomes that a school is legally required to deliver. As such, positioning Explorer as an alternative type of PE will likely lead to dissatisfaction. However, the greater flexibility afforded to schools for PSHE education provides an opportunity for Magic Bus. If Explorer is to continue as an in-school intervention, it should be positioned as a PSHE class that is responding to the specific social problems faced by a specific school's pupils. Although positioning as PSHE provides less time within a typical school schedule in control to PE, the Explorer program could be delivered as a single-hour in-school class, which would promote the program to all. The program could also be delivered as an after-school program for added flexibility and positioning. These latter approaches could serve two purposes: (a) contribute to the school's PSHE curriculum, as it will remain as part of the school day or (b) contribute by moving outside of the school's standard schedule to allow Magic Bus to define their own agenda in terms of social problems for which the Magic Bus methodology can be applied. This also opens up greater and more sustainable funding channels beyond the Department for Education, which is unlikely to "double pay" for the PSHE curriculum.

Limitations and Future Research

Several limitations must be acknowledged and addressed in future research. Three research variables used scales that were equal to or exceeded an alpha of .63 but did not meet the threshold of an alpha equal to or exceeding .70. This is justified on the basis of a limited number of items, which in turn, limit the alpha (Nunnally & Bernstein, 1994; Sijtsma, 2009), but there still remains a question of the internal consistency of these three scales. The quasi-experimental design only made use of a single class for a control group; thus, there was a large discrepancy in sample sizes between conditions, which encumbered statistical analyses. Furthermore, as a quasi-experiment that did not randomly assign study participants to condition, it is not possible to randomize out preprogram differences. As such, a simple independent samples t test on the postprogram measures only was not possible because the assumption that there are no differences between conditions was not present. Therefore, rather than attempting to make causal inferences via a control of experimental and control group postprogram measures, we attempted to make causal inferences via a control of experimental and control group respective differences between the pre- and postmeasures. Therefore, in order to strengthen causal inferences, independent samples t tests were used to test for differences between experimental and control groups on preprogram measures for all variables. However, this falls short of the strength of causal inference that comes with randomized controlled trials, regarded as the "gold standard" of causal inference (Burtless, 1995; Cartwright, 2007; Fisher, 1935; Hakim, 2000). To test the recommendations made in this study, a control could be made between two interventions tackling the same social problem with the same desired outcomes and research variables, one of which diligently designs and implements a theory of change based upon the recommendations in this paper with the support from academic researchers, and one which does not make use of the recommendations and does not receive support from academic researchers.

References

Attride-Stirling, J. (2001). Thematic networks: An analytic tool for qualitative research. *Qualitative Research, 1*(3), 385–405. doi:10.1177/146879410100100307

Bond, L., Wolfe, S., Tollit, M., Butler, H., & Patton, G. (2007). A comparison of the Gatehouse Bullying Scale and the Peer Relations Questionnaire for students in secondary school. *Journal of School Health, 77*(2), 75–79. PubMed ID: 17222158 doi:10.1111/j.1746-1561.2007.00170.x

Brewer, J. (2000). *Ethnography*. London, UK: McGraw-Hill Education.

Bruening, J.E., Welty Peachey, J., Evanovich, J.M., Fuller, R.D., Murty, C.J., Percy, V.E., . . . 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. doi:10.1016/j.smr.2014.07.002

Burtless, G. (1995). The case for randomized field trials in economic and policy research. *The Journal of Economic Perspectives, 9*(2), 63–84. doi:10.1257/jep.9.2.63

Cartwright, N. (2007). Are RCTs the gold standard? *BioSocieties, 2*(1), 11–20. doi:10.1017/S1745855207005029

Coakley, J. (1993). Sport and socialization. *Exercise and Sport Sciences Reviews, 21*(1), 169–200. doi:10.1249/00003677-199301000-00006

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

Coalter, F. (2010). Sport-for-development: Going beyond the boundary? *Sport in Society, 13*(9), 1374–1391. doi:10.1080/17430437.2010.510675

Coalter, F. (2012). 'There is loads of relationships here': Developing a programme theory for sport-for-change programmes. *International Review for the Sociology of Sport, 48*(5), 594–612. doi:10.1177/1012690212446143

Coalter, F. (2013). *Sport for development: What game are we playing?* Routledge.

Coalter, F. (2015). Sport-for-change: Some thoughts from a sceptic. *Social Inclusion, 3*(3), 19–23. doi:10.17645/si.v3i3.222

Creswell, J.W. (2012). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: Sage Publications.

Creswell, J.W., & Miller, D.L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice, 39*(3), 124–130. doi:10.1207/s15430421tip3903_2

Creswell, J.W., Plano Clark, V.L., Gutmann, M.L., & Hanson, W.E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209–240). Thousand Oaks, CA: Sage.

Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature, 48*(2), 424–455. doi:10.1257/jel.48.2.424

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The SAGE handbook of qualitative research*. Thousand Oaks, CA: Sage Publications.

Department for Education. (2013a). National curriculum in England: physical education programmes of study. Retrieved from https://www.gov.uk/government/publications/national-curriculum-in-england-physical-education-programmes-of-study/national-curriculum-in-england-physical-education-programmes-of-study

Department for Education. (2013b). Personal, social, health and economic (PSHE) education. Retrieved from https://www.gov.uk/government/publications/personal-social-health-and-economic-education-pshe/personal-social-health-and-economic-pshe-education

Firestone, W.A. (1993). Alternative arguments for generalizing from data as applied to qualitative research. *Educational Researcher, 22*(4), 16–23. doi:10.3102/0013189X022004016

Fisher, R.A. (1935). *Design of experiments*. New York, NY: Hafner.

Green, B.C. (2008). Sport as an agent for social and personal change. In V. Girginov (Ed.), *Management of sports development* (pp. 129–145). Oxford, UK: Elsevier.

Greene, J.C., & Caracelli, V.J. (1997). Defining and describing the paradigm issue in mixed-method evaluation. *New Directions for Evaluation, 1997*(74), 5–17. doi:10.1002/ev.1068

Hakim, C. (2000). *Research design: Successful designs for social and economic research*. London, UK: Psychology Press.

Hansen, D.M., Larson, R.W., & Dworkin, J.B. (2003). What adolescents learn in organized youth activities: A survey of self-reported developmental experiences. *Journal of Research on Adolescence, 13*(1), 25–55. doi:10.1111/1532-7795.1301006

Harter, S. (1982). The perceived competence scale for children. *Child Development, 53*(1), 87–97. doi:10.2307/1129640

Hartmann, D. (2003). Theorizing sport as social intervention: A view from the grassroots. *Quest, 55*(2), 118–140. doi:10.1080/00336297.2003.10491795

Hills, S., Gomez Velasquez, A., & Walker, M. (2018). Sport as an analogy to teach life skills and redefine moral values: A case study of the 'Seedbeds of Peace' sport-for-development programme in Medellin, Colombia. *Journal of Sport for Development, 6*(10), 25–37.

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. *American Political Science Review, 105*(4), 765–789. doi:10.1017/S0003055411000414

Jorgensen, D.L. (2015). Participant observation. In *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource* (pp. 1–15). https://onlinelibrary.wiley.com/doi/full/10.1002/9781118900772.etrds0247

Kruse, S.E. (2006). *Review of kicking AIDS out: Is sport an effective tool in the fight against HIV/AIDS?* Oslo, Norway: NORAD.

Larke, P.J. (1990). Cultural diversity awareness inventory: Assessing the sensitivity of preservice teachers. *Action in Teacher Education, 12*(3), 23–30. doi:10.1080/01626620.1990.10734396

Lyras, A., & Welty Peachey, J. (2011). Integrating sport-for-development theory and praxis. *Sport Management Review, 14*(4), 311–326. doi:10.1016/j.smr.2011.05.006

Magic Bus. (2016). About Us. Retrieved from https://www.magicbus.org/about-us

Magic Bus UK. (2016). Home. Retrieved from http://www.magicbusuk.org/home/

Maxwell, J.A. (2012). *Qualitative research design: An interactive approach* (Vol. *41*). Thousand Oaks, CA: Sage Publications.

Meuleners, L.B., Lee, A.H., Binns, C.W., & Lower, A. (2003). Quality of life for adolescents: Assessing measurement properties using structural equation modelling. *Quality of Life Research, 12*(3), 283–290. PubMed ID: 12769141 doi:10.1023/A:1023221913292

Morgan, D.L. (1996). Focus groups. *Annual Review of Sociology, 22*(1), 129–152. doi:10.1146/annurev.soc.22.1.129

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric Theory* (McGraw-Hill Series in Psychology) (Vol. *3*). New York, NY: McGraw-Hill.

Rubenstein, J.L., & Feldman, S.S. (1993). Conflict-resolution behavior in adolescent boys: Antecedents and adaptational correlates. *Journal of Research on Adolescence, 3*(1), 41–66. doi:10.1207/s15327795jra0301_3

Ryan, A.M., & Pintrich, P.R. (1997). Should I ask for help?" The role of motivation and attitudes in adolescents' help seeking in math class. *Journal of Educational Psychology, 89*(2), 329–341. doi:10.1037/0022-0663.89.2.329

Schulenkorf, N. (2012). Sustainable community development through sport and events: A conceptual framework for Sport-for-Development projects. *Sport Management Review, 15*(1), 1–12. doi:10.1016/j.smr.2011.06.001

Sefton, T., Byford, S., McDaid, D., Hills, J., & Knapp, K. (2002). *Making the most of it: Economic evaluation in the social welfare field*. York, UK: York Publishing Services.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. PubMed ID: 20037639 doi:10.1007/s11336-008-9101-0

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research*. Thousand Oaks, CA: Sage Publications.

Sugden, J. (2008). Anyone for football for peace? The challenges of using sport in the service of co-existence in Israel. *Soccer & Society, 9*(3), 405–415. doi:10.1080/14660970802009023

Vallerand, R.J., Deshaies, P., Cuerrier, J.P., BriÈre, N.M., & Pelletier, L.G. (1996). Toward a multidimensional definition of sportsmanship. *Journal of Applied Sport Psychology, 8*(1), 89–101. doi:10.1080/10413209608406310

Vieno, A., Santinello, M., Pastore, M., & Perkins, D.D. (2007). Social support, sense of community in school, and self-efficacy as resources during early adolescence: An integrative model. *American Journal of Community Psychology, 39*(1-2), 177–190. PubMed ID: 17437191 doi:10.1007/s10464-007-9095-2

Walker, M., & Hills, S. (2017). Social program evaluations: Strategies and shared value. *Sport and Entertainment Review, 3*, 65–71.

Walker, M., Hills, S., & Heere, B. (2017). Evaluating a socially responsible employment program: Beneficiary impacts and stakeholder perceptions. *Journal of Business Ethics, 143*(1), 53–70. doi:10.1007/s10551-015-2801-3

World Bank. (2004). *Monitoring and evaluation: Some tools, methods and approaches*. Washington, DC: World Bank.