The Importance of Theorizing Social Change in Sport-for-Development: A Case Study

of Magic Bus in London

Abstract

For sport for development (SFD) 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 study of an award winning SFD charity who expanded their operations from India to London to illustrate the 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 participant experience. The quantitative analyses yielded no significant effects. Triangulating the qualitative data revealed that personal, social, health, and economic (PSHE) 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. Additionally, alignment 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: *sport-for-development; theory; theory of change; social change*

The Importance of Theorizing Social Change in Sport-for-Development: A Case Study

of Magic Bus in London

The importance of a theory of change to guide a sport for development (SFD) program towards achieving its desired outcomes is established in the literature (Coalter, 2007, 2012; Bruening, Peachey, Evanovich, Fuller, Murty, Percy, Silverstein, & Chung, 2015; Walker, Hills, & Heere, 2017). For example, Walker et al. (2017) evaluated a soccer-based employability program delivered by an English Premier League club and concluded that the lack of hard employment outcomes achieved was attributable to a lack of synergy between program delivery and goals, and a lack of direct ties between the desired outcomes and factors related to those outcomes. 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 (UK), to demonstrate how the intended beneficiary effects did not materialize because the theory of change was not established and the social problem was misunderstood.

SFD Theory

When the desired 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, traditional and non-traditional sports, and 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 quite 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) labelled the problem of displacement of scope, whereby micro-level (i.e., program-level) mechanisms are incorrectly generalized as macro-level SFD outcomes. Rather than attempt to produce a SFD 'theory of everything' at a macro-level, theorizing should occur at the program-level, to understand 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 causes of socialization outcomes, or may work as a hook to engage a target group before other program societal 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 an analogy to replace other mechanisms/pathways of social development Hills et al. (2018) demonstrated this idea when describing the 'Seedbeds of Peace' program in Medellin, Colombia, that 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 a SFD intervention can result contextually bracketed individual or social change (Coalter, 2015). Without this, according to Coalter (2012), we are left with a 'black box' view of SFD interventions as being able to effectively change a range of values, attitudes and behaviours, with little understanding of the processes behind any change. There is a need to understand the relationship between program conception, delivery, and intended outcomes.

SFD Theory of Change

For a SFD intervention to achieve the desired outcomes, a theory of change is essential. According to Coalter (2012), a theory of behaviour change or program theory is a sequence of causes and presumed effects underpinning interventions, which articulate the 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) refer to intentional design, whereby program processes, mechanisms, and experiences add value to a desired behavior. Intentional design elements help bridge social problems, program mechanisms and processes, and broader outcomes. Through mapping out the causal factors important for success (World Bank, 2004) and understanding the relationship between program conception, delivery, and intended outcomes (Coalter, 2012), a theory of change serves two primary purposes; (1) informing program design and practice so to maximize the possibility of achieving the desired outcomes (Coalter, 2007, 2012), and (2) 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 37 in-depth interviews with participants, delving into their experiences and program elements perceived to have had the greatest impact on their values, attitudes and behaviour. This research identified

limitations with regard to the use of a theory of change in the programs. Coalter (2012) found lack of clarity in defining a target group and flawed assumptions that youths were at-risk. Programs allowed participants to self-select as opposed to targeting individuals for whom there was evidence of being at-risk. Programs recruited participants on the basis of an implicit deficit model based on an environmental fallacy that all youth from high crime areas will themselves be at-risk. Coalter (2012) found that there was little systematic analysis of the presumed deficits as a pre-program measure, creating a paradoxical danger of well-meaning projects being based on negative stereotypes, leading to misconceived provision and inappropriate performance indicators. Coalter (2012) also found that sport was presumed to play an important role in change, but that participants rarely mentioned the sport components when discussing 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) advocates that the design of a theory of change should start by defining desired outcomes and work backwards, identifying the mechanisms that will serve as causal factors, rather than assuming a cause-and-effect relationship between participation and desired outcomes. Similarly, Walker and Hills (2017), in applying the principles of social marketing, advocate that a theory of change should start by identifying a substantial and measurable social need or problem. Working back from this social problem or 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. A weakness 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 overcomes this limitation by focusing on a single case study.

Study Context

Originating in India, Magic Bus is a SFD charity who 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 UK to extend their work from the Low to Middle Income Country (LMIC) of India to the High Income Country (HIC) 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 UK. As a school-based program, it is delivered within physical education 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 physical education classes, it was not intended to deliver a physical activity curriculum. Rather, the program designers attempted to challenge the assumption these classes can only been 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 manifests 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 (or ideas) 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, since it is first experienced and accepted as a game 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 and behaviors.

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, 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 effects (or lack of) illustrates 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 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 on 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 post-program (i.e., 1-week after) questionnaires. Data collection in the qualitative phase consisted of participant focus groups and administrator interviews. The relative strengths of quantitative data were leveraged to answer research questions one through three, which measure program influence. Concurrently, the relative strengths of qualitative data were leveraged to answer research question four, which detail the mechanisms, processes, and experiences.

There are several benefits of utilizing a mixed methods approach. Statistical analysis of the quantitative data allowed for extrapolation and generalization of findings from the sample to the population (Firestone, 1993), whilst the richness, detail, depth, and nuances that are inherent in qualitative data allowed for understanding of the complexities in mechanisms, processes, and experiences (Creswell, 2012). The use of quantitative data provided a perspective based upon quantifiable and measurable 'facts,' whilst qualitative analysis made sense of and interpreted 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 effects (Burtless, 1995). In contrast, the qualitative phase used focus groups to study 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, quantitative data was used as part of a quasi-experimental design, so to answer the research questions seeking to establish the *effect* of the Explorer program, but this does not answer *how* or *why* the treatment variable works (Deaton, 2010), leading to a black box view of causality (Imai, Keele, Tingley, & Yamamoto, 2011). Therefore, qualitative data from participant focus groups and administrator interviews was required to answer the final research question to understand the mechanisms, processes and experiences that explain effects or lack of effects.

Quantitative Technique

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 pre- and post-treatment measurements, as well as between experimental and control groups for all research variables. This way, any change from pre- to post-treatment could be attributed to the program if the change was not present in the control group (Hakim, 2000). Participants were non-randomly assigned to 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 one 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 who was to participate in the program the following year. According to Sefton et al. (2002), constructing a control group from a waiting list limits selection bias and is a valid

means of constructing a non-random control group when a randomised means of construction

is not possible. In this case, both experimental and control groups were targets of the

intervention, but experienced it at different points in time. Furthermore, as an in-school

intervention where recruitment and participation goes through teachers and parents,

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. Pre- and

post-program questionnaire data were analysed using descriptive statistics, paired samples *t*-

tests to measure within-subjects effects independently (i.e., time; pre- to post-program), and a

two-way mixed analysis of variance (ANOVA) to measure the interaction of within-subject

effects (i.e., time; pre- to post-program) and between-subjects effects (i.e., condition;

experimental vs. control). Although the between-subjects analysis provide the requisite

information to assess the influence of the intervention, the within-subjects effects are

presented illustratively to demonstrate the different picture when ignoring contextual factors

which 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. Effect on beneficiary relationships was measured using the constructs

of (1) 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), (2) conflict management, defined as the behavioral and

attitudinal disposition taken when responding to a disagreement (Rubenstein & Feldman,

1993), (3) diversity awareness and attitudes, defined as respect and appreciation for diversity

leading to reduction in stereotypes and discrimination (Larke, 1990), and (4) 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 (5 items, α = .81; Harter, 1982); conflict management (8 items, α = .65; Rubenstein & Feldman, 1993); and diversity awareness and attitudes (3 items, α = .70; Larke, 1990); and bullying (6 items, α = .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 (Hansen, Larson, & Dworkin, 2003) was measured using Hansen et al.'s (2003) goal-setting scale (4 items, α = .63).

Broader outcomes. Broader outcomes was measured using the constructs of (1) sense of community in school, defined as a reflection of needs fulfilment, membership in a group, influence, and emotional connections (Vieno, Santinello, Pastore, & Perkins, 2007), and (2) 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 (6 items, α = .70; Vieno et al., 2007); and emotional wellness (5 items, α = .63; Hansen et al., 2003). All questionnaire items were adapted for age-appropriateness and were pilot tested prior to the evaluation taking place.

Variables lacking internal consistency. Due to a lack of internal consistency, data was collected but not used for the following variables: Pro-social norms (3 items, α = .50; from Hansen et al. 2003), sportspersonship (3 items, α = .52; from Vallerand et al., 1996), aspirations (3 items, α = .48; from Meuleners, et al, 2003), help-seeking (4 items, α = .32; from Ryan & Pintrich, 1997), and psychological quality of life (3 items, α = .40; from Meuleners et al., 2003). Although the standard criteria for sufficient internal consistency is a Cronbach alpha equal to or greater than 0.7, this study accepted scales with alphas as low as 0.63 on the basis that the scales used contained a limited number of items (e.g., goal-setting,

4 items, α = .63) and that alpha depends on the number of items in a scale (Nunnally & Bernstein, 1994; Sijtsma, 2009). Therefore, it is argued that a limited number of items in the scales limited the scales' alphas.

[Insert Table 1 here]

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 six to ten participants by Morgan (1996). 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 effects (e.g., How did you learn this?). Next, questions delved more 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 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 most?).

Six teacher semi-structured interviews, lasting approximately 45 minutes, were also conducted (three per participating school; two with class teachers, one with head teacher). Initial questions focused on 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 delivery protocols in the future. Finally, two administrator semi-structured interviews, lasting approximately one hour each, were

conducted with the program designer and the program coach. These interviews sought to probe the theory of change behind the Explorer program. Initial questions focused on the social problem and desired outcomes (e.g., How would you describe the social problem or problems that the Explorer programme was seeking to tackle; Why was this social problem or problems important to Magic Bus; What were the desired outcomes of the Explorer programme?), before probing on how 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 endeavours?). 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 semi-structured interviews were professionally transcribed. The data was then thematically analysed to unearth salient themes (Attride-Stirling, 2001). An inductive coding strategy was used, whereby themes and quotes were extracted, focusing on the mechanisms, processes and experience 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 data into larger abstract categories (Creswell & Miller, 2000), reassembling or reconstructing the data in a meaningful or comprehensible fashion (Jorgensen, 2015) so 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 coding and categories before reaching consensus on the themes and representative quotes.

Results

Quantitative Results

N=157 individuals completed the pre- and post-program questionnaires ($n = 134$ treatment group participants; $n = 23$ control group individuals). Ages ranged from 6-10 years (Treatment: $M = 8.38$, $SD = 1.00$; Control: $M = 7.65$, $SD = .49$). 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: Boy, 51%, Girl, 49%; Control: Boy, 48%, Girl, 52%). 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). Number of household members ranged from two (i.e., single parent) to eight (Treatment: $M = 4.29$, $SD = 1.44$; Control: $M = 3.61$, $SD = 1.23$).

[Insert Table 2 here]

Independent sample *t*-tests were run to descriptively test for differences between participants and the control group on pre-program measures, in order to characterize the strength of causal inferences. There was no statistically significant difference in pre-program 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 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 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 pre- and post-program 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 report all effects appear in Table 4.

[Insert Table 3 here]

[Insert Table 4 here]

For Research Question 1, the effect of time on the relationships variables was tested. 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 analyses showed that pre-program means for all relationships variables were positive, exceeding the mid-point of three on the five 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 is no mid-point. However, M = 2.16 acts of bullying in a year can be considered low, thus also already positive.

For Research Question 2, effect of time on goal-setting was tested. Time had a significant and moderate effect on goal-setting (t (122) = -5.249, p = .000, r = .270). Descriptive analyses showed that pre-program mean for goal-setting was already positive, exceeding the mid-point on the five-point Likert scale (M = 3.12).

For Research Question 3, effect of time on the broader outcomes variables was tested. Time had no significant effect on sense of community in school (t (128) = -.085, p = .932), or emotional wellness (t (130) = -1.843, p = .068). Descriptive analyses showed that pre-program means for both broader outcomes variables were already positive, exceeding the mid-point of three on the five point Likert scale (Sense of Community in School: M = 4.28;

Emotional Wellness: $M = 3.50$). In sum, only one (goal-setting) out of seven variables changed from pre to post. All six Likert scale variables had pre-program means above the mid-point of three, two of which had pre-program means above four out of five.

    ANOVA. Descriptive statistics were calculated (see Table 2) and a series of repeated measures ANOVA analyses were conducted in order to test the time × condition interaction on the dependent variables (see Table 3). Dependent variable (i.e., relationships, goal-setting, and broader outcomes) means were compared between pre- and post-program measurements (i.e., time; the within-subjects independent variable) and between experimental and control groups (i.e., condition; the between-subjects independent variable). We conceptually grouped the analyses according to the research questions.

[Insert Table 5 here]

    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 and moderate and negative time × interaction for goal-setting ($F(3) = 0.002$, $p = .97$). For Research Question 3, there was not a significant time × interaction for 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 effect on help-seeking, but there was a significant negative time × condition interaction. There was a statistically significant difference between conditions on the pre-program measure of bullying. Therefore, it is possible that the lack of effect on bullying could be due to differences between 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 the goal-setting for which a time effect had previously been found because there was also a positive and significant change from pre- to

post-program 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 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, Magic Bus personnel listed in excess of twenty outcomes they wished to achieve. Following much discussion, the list was reduced to twelve 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 administrators being unable to clearly articulate a social problem 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 problems that Explorer was seeking to tackle was pupil's self-confidence and
> emotional wellbeing 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, pro-social

behaviour.

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 pre-program measurements, which

averaged above the mid-point for all the research variables and above four out of five for

three of the research variables, suggests that participants were not in need of an intervention

focused on these outcomes. Hence, 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 moves 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 twelve, there was evidence of 'outcome creep'. That is, we observed the program

curriculum contained sessions related to fifteen 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 sense of community. The program was delivered over a

35-week period, with 35 sessions of 45 minutes each. Although over a long period of time,

relative to other interventions, the amount of direct contact with participants was approximately 26 hours. In that time, fifteen outcomes are excessive, spreading resources thinly. At the high end, emotional awareness and diversity awareness and attitudes received three hours of intervention, but on the low end, sportspersonship and respect received 45 minutes 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 physical education, sport, or social and health education. The Explorer program was delivered within time allocated for Physical Education (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, Health and Emotional (PSHE) education. As a result, 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:

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 two was able to tweak the

program to cover their PE needs, school one 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:

> 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 analysing 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 lead 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 programme 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 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 a 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 (personal, social, health and economic) 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 non-statutory subject (Department for Education, 2013b). The Department for Education provide schools with flexibility in their delivery of PSHE because they view 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 lack of understanding of context whereby the legal curriculum requirements were not fully met does not necessarily explain lack of effect against the research variables, this flaw limits the sustainability of the Explorer program.

General Discussion

Via the quasi-experimental design aspect of the evaluation, the influence of the Magic Bus program in the UK was assessed using several variables. Within-subjects analyses established that time significantly influenced goal-setting. However, an improvement in goal-setting were also present in the control group, suggesting that other social/situational factors were influencing these outcomes. Qualitative results revealed similarities between the Explorer program and PSHE education, which was also present in the control group. This structural aspect can help explain the lack of time $\times$ condition effect since 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 English education system. Additionally, the program deliverers were attempting to validate their methodology against outcomes for which the participants were not devoid, as evident in the high pre-program variable measurements. This aspect of program deployment runs counter to the recommendation of applying social marketing techniques to create, communicate, and deliver value to society (see 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 Magic Bus intervention. That stated, we do

not necessarily view this is a bad thing as an educationally 'additive' program is certainly one quality 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 over-ambitious 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. Serval critical elements of the curriculum were not being directly measured or directly addressed in the curriculum. And we additionally observed lack of intentional design and rigor in articulating the relationships between actions, processes, and effects. 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 context, and (c) meet 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 abovementioned areas demonstrates a lack of appreciation of the experiences 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 has 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, Coalter did not map theory of change weaknesses against program effects. Like Coalter (2012), this study established theory of change weaknesses, but went further by quantitatively measuring program effects via a quasi-experimental design and linking the theory of change weaknesses to the absence of program effects. Although Walker et al. (2017) also briefly linked theory of change weaknesses to a lack of hard outcomes in a SFD intervention, this study did not have the same focus on theory of change as the current study and Coalter (2012).

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 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. 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 under-served population. For this type of intervention, official exam results would be a better measure than a Likert scale measure of self-perceived academic capabilities. In our case study of Magic Bus, we were not able to access exam and grade data, which yielded a heavy reliance on self-reported perceptions and observations to gauge program impact. 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 adopt the desirable behavior), but also competing products and interventions, such as PSHE education. If competing products or interventions exist, practitioners should be confident their intervention

adds value beyond the structural and educational elements that already exist in the community.

Desired outcomes should lead to assuaging 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 (RCT) 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.

With regard to practical implications specific to Magic Bus, from the contextual understanding we provide in the study, it is clear the Explorer program should be repositioned as an additive in-school intervention. Furthermore, it should be retooled in order to meet each school's curricular requirements. A major challenge is that Explorer is currently delivered in place of a PE class, 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 after-school program for added flexibility and positioning. These latter approaches could serve two purposes: (1) contribute to the school's PSHE curriculum, since it will remain as part of the school day, or (2) 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 who are unlikely to 'double pay' for the PSHE curriculum. Finally, we also illustrated the importance of a control group in evaluating SFD interventions. Had a pre- and post-experimental design without a control group been conducted, there is 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 $\times$ control interaction was tested, did a more valid understanding of program effect emerge.

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 0.63, but did not meet the standard threshold of an alpha equal to or exceeding 0.7. This is justified on the basis of a limited number of items, which in turn limit the alpha (Nunnally & Bernstein, 1994; Sijtsma, 2009), but still there 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 pre-program differences. As such, a simple independent samples *t*-test on the post-program measures only

was not possible because the assumption that there are no differences between conditions not present. Therefore, rather than attempting to make causal inferences via a control of experimental and control group post-program measures, we attempted to make causal inferences via a control of experimental and control group respective differences between the pre- and post-measures. Therefore, in order to strengthen causal inferences, independent samples t-tests were used to test for differences between experimental and control groups on pre-program measures for all variables. However, this falls short of the strength of causal inference that comes with RCTs, 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.

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.

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

Bruening, J. E., Peachey, J. W., Evanovich, J. M., Fuller, R. D., Murty, C. J., Percy, V. E., Silverstien, 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.

Burtless, G. (1995). The case for randomized field trials in economic and policy research. *The Journal of Economic Perspectives*, *9*(2), 63-84.

Cartwright, N. (2007). Are RCTs the gold standard?. *BioSocieties*, *2*(1), 11-20.

Coakley, J. (1993). Sport and socialization. *Exercise and Sport Sciences Reviews*, *21*(1), 169-200.

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

Coalter, F. (2010). Sport-for-development: Going beyond the boundary?. *Sport in society*, *13*(9), 1374-1391.

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.

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.

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

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, *39*(3), 124-130.

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), 424-55.

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The SAGE Handbook of Qualitative Research*. Sage Publications.

Department for Education (2013a). *National curriculum in England: physical education programmes of study.* Retrieved 1 May 2016, 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 1 May 2016, 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.

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

Green, C. B. (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*, *74*, 5-17.

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

Harter, S. (1982). The perceived competence scale for children. *Child Development*, 53(1), 87-97.

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

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.

Jorgensen, D. L. (2015). Participant observation. *Emerging trends in the social and behavioral sciences: An interdisciplinary, searchable, and linkable resource*, 1-15.

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

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

Lyras, A., & Peachey, J. W. (2011). Integrating sport-for-development theory and praxis. *Sport Management Review*, *14*(4), 311-326.

Magic Bus (2016). *About Us*. Retrieved 1 May 2016, from https://www.magicbus.org/about-us

Magic Bus UK (2016). *Home*. Retrieved 1 May 2016, from http://www.magicbusuk.org/home/

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach* (Vol. 41). 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.

Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology*, *22*(1), 129-152.

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

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.

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.

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

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research*. 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.

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.

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.

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.

World Bank (2004). *Monitoring and Evaluation: Some Tools, Methods and Approaches*. Washington, DC: World Bank.

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 & 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
Pro-social 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

Pre-program Variables	Standard Error (*df*) Treatment → Control	95% Confidence Interval		Treatment → Control *t*-Value [a]	Treatment → Control *p*-Value
		Lower	Upper		
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 & Attitudes	0.167 (100)	-0.004	0.660	1.960	.053
Bullying	0.240 (155)	-1.404	-0.457	-3.880	.000 [+]
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. [+] Indicates a significant difference
[a] Absolute value of the *t*-statistic

Table 3

Treatment and Control Group Descriptive Statistics

Constructs	Pre-Program (*treatment*) Mean (SD)	Post-Program (*treatment*) Mean (SD)	Pre-Program (*control*) Mean (SD)	Post-Program (*control*) Mean (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 & 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 4

Paired-Samples T-Test

Paired Variables	Standard Error (*df*) Pre → Post	95% Confidence Interval Lower	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 & 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 [+]	.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. [+] Indicates a positive and significant difference

[a] Absolute value of the *t*-statistic

* *r*=0.1 is a small effect, *r*=0.3 is a moderate effect, *r*=0.5 is a large effect

Table 5

Two-way Mixed ANOVA

Time x Condition	F Value (df)	p Value	d
Social Competence	1.97 (*1*)	.16	--
Conflict Management	.076 (*1*)	.78	--
Diversity Awareness & Attitudes	2.06 (*1*)	.16	--
Bullying	1.18 (*1*)	.28	--
Goal-Setting	.002 (*1*)	.97	--
Sense of Community in School	.183 (*1*)	.67	--
Emotional Wellness	.005 (*1*)	.94	--

Note. ⁻ Indicates a negative and significant interaction
  * *d*=0.2 is a small effect, *d*=0.5 is a moderate effect, *d*=0.8 is a large effect