# SPORT FOR DEVELOPMENT IN THE UNITED STATES: A SYSTEMATIC REVIEW AND COMPARATIVE ANALYSIS

## FINAL REPORT

**Submitted by:**

Meredith A. Whitley, PhD, Adelphi University

In collaboration with:

William Massey, PhD, Oregon State University

Lindsey Blom, EdD, Ball State University

Martin Camiré, PhD, University of Ottawa

Laura Hayden, EdD, University of Massachusetts Boston

Simon Darnell, PhD, University of Toronto

November 30, 2017

## EXECUTIVE SUMMARY

This review examined and compared the reported evidence of youth-focused Sport for Development (SfD) interventions and non-sport youth development interventions from the United States. The quality of evidence was assessed, with a particular focus on four thematic areas: (a) community development, social cohesion, and peacebuilding; (b) education; (c) employment; and (d) mental health and well-being. This allowed for an informed analysis of the gaps that exist, as well as a basis from which to suggest 'next steps'.

For both academic and grey literature in SfD, the quality of methods and evidence was largely appraised as weak, suggesting the need to interpret the findings with caution. This matched the findings from the reviews of non-sport youth development research, in which overall quality was appraised as largely weak. In both fields, the most frequently assessed theme was mental health and well-being, followed by community development, social cohesion, and peacebuilding. Employment was rarely assessed. As for the data which emerged from the studies with rigorous assessments (quantitative) and/or methodological coherence (qualitative), there was no clear pattern, given the low number of studies that met these standards. Of the 77 SfD studies critically appraised, 53 represented interventions/models with multiple research studies. This allowed for an assessment of the quality of evidence within each group, along with identification of common outcomes and moderating variables. The critical factors that impacted the efficacy of SfD interventions were inconsistently reported and studied, resulting in more questions than answers as to 'what works' and 'what influences' impactful (and non-impactful) SfD interventions, and 'why'. The non-sport youth development literature examined programmatic and contextual factors in more depth, identifying approaches which could be implemented more systematically in the SfD field (e.g., stakeholder engagement, parental engagement, whole school/community approach). Overall, these findings informed a set of recommendations for organizations, researchers, funders, and policy makers, such as: (a) designing locally-focused, culturally sensitive interventions; (b) integrating measurement of critical factors into research and evaluation efforts; (c) reporting negative and null outcomes; (d) using validated measures and assessments of actual behavior; (e) prioritizing funding for rigorous research and evaluation; and (f) considering contextual factors when attempting to 'scale up' from individual interventions.

## REVIEW OF METHODOLOGY

A detailed outline of the methodology is described in Annex 4, with a summary presented herein. Through a systematic search of both the academic and grey literature, **10,074 SfD articles and documents** were identified. Through a systematic search of the academic literature, **829 non-sport youth development reviews** were identified. A rigorous screening process of both sets of literature was then conducted by two independent reviewers, with inclusion criteria for SfD including: (a) full text articles/evaluations with enough methodological data to critically appraise the study; (b) data collected completely/partly in the United States; (c) empirical studies reporting primary data; (d) average age of participants between 10 and 24 years old; and (e) evidence of a plus-sport or sport-plus intervention for SfD studies. The inclusion criteria for non-sport youth development were: (a) full text articles with enough methodological data to critically appraise the review; (b) reviews of research on interventions conducted completely/partly in the United States; (c) full text articles reporting complete data from a review; (d) average age of participants between 10 and 24 years old; and (e) reviews of research examining interventions that intentionally foster youth development outcomes. This resulted in **77 SfD academic studies**[a] and **47 non-sport youth development reviews** (see Annexes 2 and 3 for reference lists).

The methods were then assessed and critically appraised by two independent reviewers, with the detailed tables of the methodological quality of these studies presented in Annexes 7 – 14. Data extraction for all studies included detailed information about each intervention and study/review (e.g., intervention participants, instrumentation, data analyses, results, critical factors). The Quality Tool for Quantitative Studies was used for all SfD studies containing quantitative data, and a level of evidence rating was given for interventions with multiple studies conducted using the Early Intervention Foundation (EIF) criteria. Overall, there were high levels of heterogeneity in the designs, methods, interventions, and outcomes reported across the included SfD studies. Within different intervention groups, some studies possessed enough similarity to conduct a meta-analysis, with these steps outlined in Annex 10. However, the majority of studies were unable to be considered for meta-analysis, resulting in a narrative

[a] While there are 82 SfD references listed in Annex 2 (and, therefore, 82 articles included in this systematic review), 5 of these studies were presented in multiple publications (e.g., dissertation and peer-reviewed article, full report and brief report); thus, these were considered duplicate documents, with 77 SfD studies total.

description of the evidence. To examine the methodological quality of qualitative SfD studies, a meta-theory and meta-method approach was taken. As for the non-sport youth development reviews, the AMSTAR 2 checklist (see Annex 13) was used to examine the methodological quality of the included reviews, with a narrative description of the evidence provided below.

## REVIEW OF FINDINGS

### QUALITY OF EVIDENCE FOR SPORT FOR DEVELOPMENT RESEARCH

The methods presented in 77 SfD studies were independently assessed and critically appraised (see Annexes 7 and 8), with **40 quantitative studies**, **29 qualitative studies**, and **8 mixed methods studies**. The quantitative/mixed methods studies were assessed in 7 key areas (i.e., selection bias, study design, confounding variables, blinding, validity of data collection instruments, withdraws and dropouts), with each section rated as strong, moderate, or weak along with an overall assessment. **Out of the 48 studies with a quantitative component, 41 were classified as weak evidence, 4 were classified as moderate evidence, and 3 were classified as strong evidence.** The studies classified as weak evidence included 30 single group designs, 3 retrospective designs, 7 quasi-experimental, two-group pre-post designs, and 1 randomized control trial. Common problems were self-report measures in which the participants knew the purpose of the study, selection bias in sampling procedures, the lack of control for confounding variables, and an over-reliance on single group studies. As for the studies which were rated as moderate or strong evidence, 5 were quasi-experimental designs and 2 were randomized control trials. Only 3 of these studies [41,44,51] showed positive results, with 2 studies [30,35] reporting mixed results and 2 studies [28,46] reporting no effect. Overall, studies rated as moderate and strong evidence were more likely to report a null finding than studies classified as weak evidence (29% vs 2%, $p = .052$).

As for the 37 studies with a qualitative component, the interpretations made from the findings were limited by a lack of philosophical, methodological, and/or theoretical underpinnings to the studies. **Only 1 study [78] made explicit and coherent connections between ontology, epistemology, methodological paradigm, sampling, data collection methods, and data analysis methods**, with 15 studies demonstrating some level of methodological coherence. Data analysis was often the missing link, as many authors used “grounded theory analysis” without doing grounded theory. Sampling procedures were not addressed in 6 studies, data

saturation was only mentioned in 1 study, and triangulation and member checking were the most often cited forms of rigor/validity (mentioned in 16 studies).

## QUALITY OF EVIDENCE FOR NON-SPORT YOUTH DEVELOPMENT REVIEWS AND ANALYSES

The methods presented in 46 non-sport youth development reviews were independently assessed and critically appraised using the AMSTAR 2 checklist (see Annexes 12 and 14), with meta-analyses conducted for 18 reviews and narrative analysis conducted for 1 review. Methodological strength for the AMSTAR 2 is determined by the number of flaws in seven critical domains (i.e., protocol registered before commencement of review, adequacy of the literature search, justification for excluding individual studies, risk of bias from individual studies, appropriateness of meta-analytic methods, consideration of risk of bias when interpreting results of a review, assessment of presence and likely impact of publication bias). **Out of the 47 reviews, 3 were rated as having moderate confidence in the reported effects,** [108,120,122] **9 were rated as having low confidence in the reported effects, and 34 were rated as having critically low confidence in the reported effects. There were no reviews rated as strong.** Strengths for these reviews included comprehensive search processes and the use of targeted PICO questions (i.e., population, intervention, control group, outcome). Common problems for these reviews included justification for excluding individual studies, risk of bias from individual studies, appropriateness of meta-analytic methods, consideration of risk of bias when interpreting results of a review, and assessment of presence and likely impact of publication bias.

Methodological strengths of the individual studies (presenting research on non-sport youth development interventions) examined in these reviews included fairly large samples with both genders, along with strong exploration of youth outcomes and effects. Overall, methodological weaknesses of the individual studies included widespread inconsistency with the type of experimental design used for intervention assessment (i.e., randomized control trials vs. other designs). Regardless of the design, the review authors frequently cited concerns with the methods or lack of randomization as well as inappropriate control or non-matching control groups. Additionally, few studies described intervention fidelity, while there was also a general gap in follow-up assessment post intervention, with only 3 studies referring to a follow-up procedure. Finally, there was a concern about the high level of bias within these intervention studies. **Overall, for the majority of the studies examined in these reviews, the**

**methodological quality was considered to be (largely) weak or moderate, and there were gaps in the methods reported by many of the studies.**

## SUMMARY OF REPORTED SPORT FOR DEVELOPMENT INTERVENTION OUTCOMES

Of the 77 SfD studies critically appraised, 24 represented stand-alone papers (see Annex 11) and 53 represented interventions/models with multiple research studies published (see Annex 9). Interventions grouped together (i.e., more than 1 study) will be discussed briefly below, to provide a sense of the quality of evidence within each group.

### NYSP/SUMMER CAMPS

**Summer sport and life skills camps sponsored by the National Youth Sport Program (NYSP) or universities were assessed in 11 studies (9 quantitative, 2 qualitative), which garnered an NL2 EIF rating.** Eight quantitative studies were classified as weak evidence, with the vast majority using single group, pre-test post-test designs, while the qualitative studies lacked methodological coherence (1) and consistency between data analysis procedures and design (1). Pooled data from these studies showed no short term (pre-post) effect on attraction to physical activity, self-control, hope, or self-worth (see Annex 10). Small, but statistically reliable effects were reported for social and athletic competence. **Overall, data suggests a limited effect of summer sport and life skills camps on youth outcomes. A major limitation in this work is the overall short assessment period and lack of follow-up. Given that the outcomes assessed could be considered more stable (e.g., perceptions of competence, hope, self-worth), it is possible that the short intervention and observation phases were not sensitive enough to show any change that may have occurred.**

### TEACHING PERSONAL AND SOCIAL RESPONSIBILITY

**The Teaching Personal and Social Responsibility (TPSR) model was used in 17 studies (2 quantitative, 13 qualitative, 2 mixed methods) that met the criteria for inclusion in this group, with an EIF rating of NL2.** The quantitative components of 3 of the studies were judged as a weak level of evidence, with 1 study[51] rated as strong evidence. Of the 15 studies with a qualitative component, 6 studies showed no methodological coherence, 8 studies showed partial methodological coherence, and 1 study had full methodological coherence. Common weaknesses noted were a lack of philosophical or theoretical basis to the study, along with an over-reliance on deductive/confirmatory analyses. Moreover, studies had an over-reliance on participant perspectives, without triangulation with more objective outcomes. Strengths of the qualitative studies included multiple sources of data. Overall, low rated qualitative studies

confirmed that participants experienced some of the TPSR levels (e.g., respect, effort, leadership). While studies often mention transferring these skills to other domains, actual data on the concept of transfer was scarce. Perhaps encouraging for proponents of the TPSR model is a study conducted by Miller.[51] This was the only study to receive a strong methodological rating and examined a TPSR intervention within the context of self-determination theory and goal perspective theory, while targeting moral reasoning. Results showed positive effects on distributive justice reasoning and perceived competence for those in the intervention condition. **Thus, it appears the TPSR model can be useful; however, more theoretically-based interventions may need to be used in conjunction with this professional practice model to elicit positive changes in youth development outcomes.**

## GIRLS ON THE RUN

**Girls on the Run (GotR) was examined through 5 published studies and 3 evaluation reports (6 quantitative, 2 mixed methods), with an NL2 EIF rating.** Of the quantitative studies, 7 were rated as weak evidence and 1 as moderate evidence, while there was no methodological coherence in the qualitative components. **While pre-post data within GOtR provide consistent evidence of small, but positive effects of self-esteem and body image satisfaction, these findings are not upheld by multiple group and more rigorous designs (see Annex 10).** The effect seen in pre-post studies is likely a spurious result explained by maturation bias, selection bias, and/or the combination of not blinding participants from the study purpose and the concurrent use of self-report measures. More rigorous research should be conducted to examine how changes in physical activity might mediate mental health and wellness related outcomes. **Overall, data suggests that initial exposure to GOtR likely has the greatest effect and that there may be a ceiling effect for intervention benefit.**[19]

## THE FIRST TEE

**The First Tee was assessed in 3 studies (2 quantitative, 1 qualitative), which garnered an NL2 EIF rating.** The quantitative designs were rated as weak (1) and moderate (1) evidence, while there was partial methodological coherence for the qualitative design. The qualitative study[75] identified various life skills that youth perceive to have developed through their experience with The First Tee (e.g., the ability to meet and greet others, show respect, control emotions), with these data corroborated by longitudinal, quantitative data collected by Weiss et al.[74] **Overall, while limited, the data suggests a positive effect on life skill development for those in The First Tee.** This data is strengthened by the inclusion of multiple group data, longitudinal data, qualitative data, and first-person reports, as well as the identification of various critical features (e.g., baseline risk, gender, race, engagement outside of the intervention) that need to be considered to effect positive change over time.

## PLAY IT SMART

**Play It Smart was assessed in 2 studies (1 qualitative, 1 quantitative), with an NL2 EIF rating.** The quantitative study was rated as weak evidence, while the qualitative study lacked methodological coherence despite using data from exit interview forms from 1,361 youth. **Limited data support positive youth development and academic outcomes for participants in Play It Smart, although an overall lack of methodological processes and procedures limits the findings.**

## SPORT SQUASH

**Sport Squash had 3 studies conducted (2 quantitative, 1 qualitative), with an NL2 IEF rating overall.** Doctoral dissertations represented the quantitative studies, which were rated as moderate evidence and limited by possible selection bias and high attrition rates; there was no methodological coherence for the qualitative study. While qualitative data indicates participants have positive perceptions of the intervention, there seem to be no effect for overall academic functioning, behavioral academic engagement, or psychological academic engagement (see Annex 10). **Overall there is inconclusive evidence to suggest Sport Squash has a positive impact on academic skills.** Data from the Green [30] paper did show a significant increase in GPA among participants, but given the null results reported elsewhere, the mechanisms behind this result and/or the ability to replicate this finding are limited.

## UP2US SPORTS COACH ACROSS AMERICA

**Up2Us Sports Coach Across America was assessed with 2 quantitative evaluation reports, garnering an NL2 EIF rating.** Both studies were rated as weak level evidence due to missing data, unclear data collection procedures (e.g., reliability and validity of instruments, data collection processes, blinding), and lack of control for confounding variables. Data showed improvements in self-reported physical activity and physical fitness tests for intervention participants. **Data examining youth development outcomes were mixed, with better results for younger participants and those with higher levels of organizational contact. However, these findings should be examined cautiously, given the methodological concerns of the studies.**

## DOC WAYNE

**Doc Wayne had 2 studies conducted (1 quantitative article, 1 quantitative evaluation report), with an EIF rating of NL2. Both studies were rated as weak evidence, although results demonstrated positive trends for those participating in a trauma-informed sports league on participant behavior and levels of emotional regulation.**

### SPORT HARTFORD

**Sport Hartford was assessed by 3 qualitative studies, although no EIF rating was given due to the lack of overall data and lack of experimental studies.** One study lacked methodological coherence, while the other 2 had partial coherence. All studies were limited by low participant numbers. **Data suggests Sport Hartford shows promise but currently lacks efficacy data.**

### SPORT FOR PEACE

**The Sport for Peace curriculum was assessed by 2 qualitative studies which lacked methodological coherence, and no EIF rating was given due to the lack of overall data and lack of experimental studies. Data suggests Sport for Peace shows promise, but there is a lack of data to support its efficacy.**

Of the 24 represented stand-alone papers, two demonstrated moderate or strong evidence, which are outlined below.

### PLAYWORKS

**A quantitative study by Madsen et al.**[41] **was rated as a strong level of evidence. With each additional year of exposure to Playworks, students reported significantly higher scores in physical activity, meaningful participation in school, problem solving skills, and goal aspirations.** While the effects reported are small, they are clinically meaningful when considered across time and within the context of percentile rank. Other Playworks studies exist, but were excluded in this systematic review due to the average age of the participants.

### TRUANCY INTERVENTION

**A quantitative study by Marvul**[44] **was rated as a strong level of evidence.** This was a holistic intervention that included phone calls home, school sports programming (a school only for those with truancy problems), and participation in a morality class. **Research demonstrated significant positive effects for absences, educational attitudes, educational expectations, and academic engagement.**

## COMPARISON WITH REPORTED NON-SPORT YOUTH DEVELOPMENT INTERVENTION OUTCOMES

The quality of evidence and related outcomes from non-sport youth development will now be outlined and compared with the SfD findings, organized by four thematic areas: (a) mental health and well-being; (b) community development, social cohesion, and peacebuilding;

(c) education; and (d) employment. **Across all themes, there was mixed evidence on a wide range of outcomes in the non-sport youth development literature, with 20 reviews reporting only positive results among the interventions assessed, 23 reviews reporting mixed results (the majority of which were positive and null findings), and 3 reviews reporting only null findings (2 of which were rated as moderate).**

## MENTAL HEALTH AND WELL-BEING

Out of the 46 non-sport youth development reviews, 35 included outcomes related to mental health and well-being. **Of these, 2 were rated as having moderate confidence in the reported effects,[108,122] while 7 were rated as low confidence and 26 as critically low confidence based on the AMSTAR 2 scoring.** The outcomes assessed included psychological (e.g., post-traumatic stress disorder, overall mental health), behavioral (e.g., substance use, sexual abuse information, quality of life), and motivational (e.g., coping, self-management, confidence, empowerment) components of mental health and well-being. Focusing specifically on the reviews rated with moderate confidence, Kremer et al. [108] examined the effects of after-school interventions with at-risk youth, with null findings reported on externalizing behaviors (e.g., substance use). Tolan et al. [122] reviewed mentoring interventions for high-risk youth, with one of the four outcome areas (i.e., drug use) found to have significant positive moderate effect sizes.

Within the 77 SfD studies, 54 included outcomes related to mental health and well-being. There was a larger focus on mental health and well-being in the NYSP/summer camps (e.g., emotional self-regulation, self-worth, hope, attraction to physical activity), TPSR interventions (e.g., confidence, emotional regulation, health risk-taking), The First Tee (e.g., self-management skills), Up2Us Sports Coach Across America (e.g., physical activity, nutrition, fitness), Sport Hartford (e.g., views of self, views of health), and Playworks (problem solving, physical activity), with this theme exclusively measured in GotR (e.g., self-esteem, body image satisfaction, self-efficacy, empowerment, attitudes towards physical activity) and Doc Wayne (e.g., mental health, emotional regulation, stress). Overall, 42 studies reported only positive results within this theme, with findings including: (a) positive effects for perceived competence; (b) improved emotional regulation; (c) small, but statistically reliable effects for athletic competence; and (d) improvements in self-reported physical activity and physical fitness. There were mixed findings on some outcomes (e.g., body image satisfaction, self-esteem, physical activity, physical activity commitment). For example, GotR showed consistent evidence of small, but positive effects of self-esteem and body image satisfaction, although these findings are not upheld by multiple group and more rigorous designs.

## COMMUNITY DEVELOPMENT, SOCIAL COHESION, AND PEACEBUILDING

Out of the 46 non-sport youth development reviews, 27 included outcomes related to community development, social cohesion, and peacebuilding. **Of these, 3 were rated as having moderate confidence in the reported effects,[108,120,122] while 5 were rated as low and 19 as critically low confidence.** The outcomes assessed included anti-social behaviors, delinquency behaviors, resilience, prosocial behaviors, disruptive behaviors, use of violence, social competence, and bullying. Focusing specifically on the reviews rated with moderate confidence, Taheri and Welsh[120] examined after-school interventions focused on addressing delinquency behaviors. The overall findings indicate that after-school interventions are not effective for preventing delinquency, but may assist parents and provide important social opportunities for youth. Tolan et al.[122] examined mentoring interventions for high-risk youth to decrease delinquency behaviors, with significant positive effect sizes found for delinquency and aggression, although both were modest in size and heterogeneity was noted for the outcomes. Kremer et al.[108] examined the effects of after-school interventions with at-risk youth, with null findings reported on externalizing behaviors (e.g., disruptive behaviors such as delinquency, aggression, stealing, fights, and gang activities).

Within the SfD literature, 44 studies included outcomes related to community development, social cohesion, and peacebuilding. There was a larger focus on community development, social cohesion, and peacebuilding in the NYSP/summer camps (e.g., belonging, teamwork, social skills, empathic concern, prosocial behaviors), TPSR interventions (e.g., distributive justice reasoning, social responsibility, prosocial norms), The First Tee (e.g., interpersonal skills, social responsibility), and Sport Hartford (e.g., connection to community, contribution, sense of belonging). Overall, 35 studies reported only positive results within this theme, with findings including: (a) small, but statistically reliable effects for social competence; (b) positive effects on distributive justice reasoning; and (c) improved ability to meet and greet others. There were also mixed findings on some outcomes (e.g., belonging, teamwork, social responsibility, respect).

## EDUCATION

Out of the 46 non-sport youth development reviews, 10 included outcomes related to education. **Of these, 2 were rated as having moderate confidence in the reported effects,[108,122] while 3 were rated as low confidence and 5 were rated as critically low confidence.** The outcomes assessed included school attendance, bonding to school, school attitudes, and academic performance. Focusing specifically on the reviews rated with moderate confidence, Kremer et al.[108] studied after-school intervention effects on school attendance, with no significant effects found. If school attendance is an outcome of interest, the review authors suggest that this must be explicitly stated with a behavior

change theory integrated into the intervention. Tolan et al. [122] studied the impact of mentoring interventions on high-risk youth, finding a significant moderate positive effect size on academic achievement, with heterogeneity noted.

Within the 77 SfD studies, 19 included outcomes related to education. There was a larger focus on education in TPSR interventions (e.g., grades, school tardiness, school absences), Play It Smart (e.g., ACT/SAT scores, GPA), Sport Squash (e.g., intellectual functioning, academic functioning, academic achievement), Playworks (e.g., participation in school), and a truancy-prevention intervention (e.g., truancy, attitudes towards education, educational expectations, educational engagement). Overall, 16 studies reported only positive results within this theme, with findings including: (a) significantly higher scores in meaningful participation in school and (b) significant positive effects for absences, educational attitudes, educational expectations, and academic engagement. There were with mixed findings on some outcomes (e.g., intellectual functioning, academic functioning, academic achievement). The positive results within this theme were often in the realm of attitudes towards education or more process measures, while minimal positive effects were seen on academic performance. Given that academic performance (e.g., GPA, test scores) can be difficult to influence and often requires systemic and longitudinal programming and research, it may behoove researchers to select outcomes more appropriate to measure at group levels.

### EMPLOYMENT

Out of the 46 non-sport youth development reviews, only 2 included outcomes related to employment. [94,98] Both of these were rated as critically low in confidence about the results of the reviews. The outcomes assessed included career recognition, career attitudes, and career success.

Within the SfD literature, the only studies which focused on employment examined one TPSR intervention, [68,70,71] the Kinesiology Career Club, which explicitly explored participants' future selves. Outcomes assessed in this intervention included clarity about the future and the ability to see their paths, with improvement reported on all employment-related outcomes.

## CRITICAL FACTORS

The SfD evidence base mainly presents specific intervention outcomes and not the critical factors that may impact these outcomes, although there was some exploration of programmatic factors. Those presented herein are from the grouped and/or more rigorous studies, although some of these findings came from studies which showed minimal impact on the intervention. Thus, these findings should be interpreted with caution. This begins with the NYSP/summer camps, with many of these studies using mediational and moderation analyses to

identify key programmatic factors. Data suggests that the use of a caring climate has the potential to increase empathic concern [52] and that affective self-regulatory efficacy and empathic self-efficacy help mediate the link between a caring climate and social behaviors. [29] Moreover, **leader behavior and perceived leader support** are important predictors of youth outcomes and ongoing intervention attendance. [58,63] Program engagement was also shown to be influenced by lower BMI and higher self-worth. [63] These findings have large implications, as sustained program engagement increases the chance of beneficial outcomes. Another critical factor is developing a sense of belonging, [5] which also has been shown to influence youth outcomes. Finally, data suggests that camps are most effective for older youth [64] and those who come in with higher levels of baseline risk. [5] Qualitative data support these findings, with summer camps reported as safe places where relationships can be developed. [55] As for TPSR interventions, some authors in the higher rated qualitative studies identified critical elements that may help enhance youth outcomes. This included the importance of **teacher-student relationships**, [15,61,78] along with the need for a **relevant curriculum** with opportunities to practice life skills both within and outside of the intervention context. [80] **Engaging youth as leaders** was another key programmatic factor for TPSR interventions, including active roles in programming, decision-making, and evaluation. [33,43,50,61,69-71,78,80,81] Additionally, both Schilling et al. [61] and Whitley et al. [78] reported that the **intervention structure and environment** are likely antecedents as to whether or not youth engage in SfD interventions, and that intervention logistics can be a barrier to participation. Overall, there was a focus on the importance of **sustained engagement** in TPSR interventions, which supports enhanced development and more effective transfer into other domains. [33,37,68,69] Girls on the Run research suggested that social support and the development of self-worth through the program were important factors in program success, [67] while data on The First Tee suggest several potential moderating variables. First, those who entered the intervention with the lowest scores for life skills showed the highest levels of improvement over time, [74] suggesting this intervention is effective when **targeting the right youth**. Additionally, Brunelle and colleagues [13] reported that those engaging in high levels of community service had more favorable outcomes, suggesting a critical need to engage youth outside of the sport program itself. Findings from Marvul [44] indicate the importance of **providing respect and support for students, cultivating hope within students, and outlining positive reasons for attending school, along with guidance from caring adults and a supportive, engaging curriculum**.

Shifting focus to the non-sport youth development reviews, there were numerous programmatic factors which may translate into the SfD field. Those presented herein were mentioned in multiple reviews and/or cited in the more rigorous reviews. This begins with a need for interventions to explicitly focus on **skill building, foster youth leadership development, and cultivate strong, positive, sustainable, caring relationships with adults**. The authors of several reviews cited all three elements as being critical to the most comprehensive, effective interventions on psychosocial change.[89,109,113] Positive youth development indicators and other study findings routinely emphasize the importance of building and sustaining long-term human relations and support systems to effect change,[e.g.,89,105,109] beginning with, but not limited to, relationships between staff and youth.[102,105,109,128] A meta-analysis on mentoring interventions[122] outlined stronger interventions effects when advocacy and emotional support were part of the mentoring process, once again highlighting the importance of relationships. **Parental involvement** in interventions was also cited as a critical factor, as sustainable change is most likely to occur when both parents and peers are engaged in the programming with youth, providing support and learning together.[85,88,106,107] For example, Kenny et al.[106] found that when youth engage in programming activities with their parents, they receive repeated exposure to prevention information in their natural environment (i.e., at home). Another meaningful programmatic factor was identified by Hahn et al.,[100,b] who suggested that **school attendance and monitoring interventions with negative contingencies may adversely affect youth** by reducing access to requisite support and resources which help change targeted behaviors. Similarly, there is a need to emphasize **strength-based development**, rather than taking punitive approaches.[83,84,89] Findings also suggested the use of a **cognitive-behavioral approach** to youth development interventions, which is especially beneficial for supporting transferability over time.[85,114] Interestingly, the analyses conducted by Kremer et al.[108] and Taheri and Welsh[120] indicated that none of the moderating variables studied (e.g., participant age, intervention duration/type/context/focus) explained the intervention outcomes (i.e., externalizing behavior, school attendance, delinquency). These findings (from reviews rated as moderate) seem in direct contrast to some of the SfD findings, as well as findings in the non-sport youth development

[b] While based on a sole systematic review rated as critically low, it is particularly unique in the potential to transform our thinking in the SfD about what guidelines which be set to foster youth development; thus, it was included here to stimulate thinking and discussion.

literature which cited the benefits of repeated exposure to life skills along with opportunities to practice these skills.[106,109]

Moving beyond programming, there were a set of contextual factors from the SfD qualitative studies which should be thoughtfully considered. This begins with taking a **systems approach** to the development and implementation of SfD programming, with strong connections and formal partnerships with schools, families, and the larger community. This aids in the transfer of life skills, with teachers, parents, and mentors able to reinforce and support youth in their use of these skills in new domains.[12,34,80] These individuals and larger groups can become **key stakeholders** who can actively work to support and promote this transference, while also ensuring the intervention is culturally relevant to the youth and the community and providing resources critical for intervention success.[12,34,61] Finally, there was also a focus on **coach and leader training and education** in literature from **The First Tee**, **TPSR**, and **Up2Us Sports Coach Across America**, with a need for initial and ongoing support to ensure intervention success.[39,50,75,78] This included, but was not limited to, intervention-specific training, awareness of potential impact on youth, supervised practice of techniques and approaches with youth (rather than adults), and ongoing individual, peer, and group reflection.

Similar themes arose in the non-sport youth development reviews, along with new approaches which could be implemented more systematically in the SfD field. This begins with the need for intervention facilitators to **understand the community or neighborhood** in which they work,[109,120] along with **developing local partnerships** that meaningfully engage families, community members, counselors/psychologists, educators, schools, local organizations, private foundations, federal agencies, researchers, and/or policy makers who can provide expertise and help support/cultivate a comprehensive, sustainable, and environmentally appropriate intervention.[83,94,98,102,109,115,117,126,128] These individuals/groups were identified as **key stakeholders** who could either promote or support the intervention over time, so it is critical to engage with these individuals and groups, remaining open to feedback and willing to adapt. One review (rated as moderate) identified the need to reexamine the purpose of after-school interventions and the ways in which these interventions are designed, implemented, and evaluated, because **many of the interventions were not specific in nature to addressing the needs of the targeted population**.[120] This matches the call for interventions which are culturally relevant, even if the same base curricula is offered in different settings.[101,104] There was also a focus on **multi-modal,**

**whole school/community approaches**, with the goal of maximizing the infrastructure and systematic aspect.[e.g.,101,119] The connection to schools was noted in the SfD research as well; of the 7 studies rated as moderate to high quality evidence, the 3 studies with positive results were all school-based and the 2 studies with mixed results had a heavy connection to schools with after-school mentoring. While the dominant focus in these non-sport reviews was on youth development outcomes, there was also a focus on **addressing and influencing systemic change**, which was most likely when this was the target of these interventions. Durlak et al.[92] assessed the potential for community coalitions and task forces to achieve this, while Hahn et al.[99] explored the impact of engaging the schools, parents, and community while also modifying the broader environment. This includes potentially moving to alternative environments outside of schools (particularly in urban settings), as schools often have rigid environmental expectations that may dictate the design and implementation of programming.[96,126] There was also a call for more **tertiary-level interventions** (over primary- or secondary-level interventions), which manage the impact of ongoing conditions and challenges, creating greater potential to positively impact a target population.[85,c]

These findings have implications for policy in SfD as well. In particular, the findings illustrate the significance and importance of both organizational partnerships and social relationships within SfD, and the need to consider sport as but one component of successful youth development policy. The findings also highlight the importance of context, and suggest that the successful 'scaling up' of SfD is likely to require an appreciation of the needs of targeted populations and of the particular development challenges to which sport is asked to contribute.

## POTENTIAL FOR SCALING IMPACT

The potential to scale these SfD interventions and/or an assessment of the critical factors necessary for scaling these approaches was rarely examined. While some interventions had expanded or shown the capacity to scale, there was little discussion of what enabled this. However, the three systematic reviews of non-sport youth development rated as having moderate confidence in the reported effects[i.e., 108,120,122] all found limited (if any) moderating effects for outcome variables, possibly suggesting the potential to scale interventions across a diverse range

[c] While based on a sole systematic review rated as critically low, it is particularly unique in the potential to transform our thinking in the SfD about what guidelines which be set to foster youth development; thus, it was included here to stimulate thinking and discussion.

of settings. These findings contrast with some of the SfD findings, as well as findings in the less rigorous non-sport youth development literature, which suggested the need for more targeted interventions. A question then emerges about which moderating variables are critical for scaling SfD interventions? An assumption could be made that the contextual factors identified above (e.g., systems thinking, stakeholder engagement, local partnerships, leader training/education, adaptation, tertiary-level interventions) could be facilitators to scaling impact for SfD interventions, although this assumption should be rigorously tested in the field. It is also possible that these contextual factors may limit scaling, such as the need to create culturally/contextually relevant programming and develop strong local partnerships and meaningful stakeholder engagement within each community. Additionally, barriers to programming identified in the SfD and non-sport youth development literature (e.g., funding gaps, parent disengagement, conflict with community, lack of community understanding) may also be barriers to scaling impact, but again, this assumption must be examined in depth before such conclusions can be made.

## RECOMMENDATIONS

Given the concerns with the quality of methods and evidence, it is especially difficult to determine which thematic areas SfD interventions seem capable of making the most effective and impactful contributions in the future. One could argue that the thematic areas most represented in the results (i.e., mental health and well-being, community development, social cohesion, and peacebuilding) offer the greatest potential for impactful contribution, but it is possible these results were featured for a variety of reasons, such as their ease of measurement, the funding climate, critical factors, or policy. Additionally, it is difficult to identify when sport-based approaches may offer a ‘best buy’ option to contribute to youth development and broader social policy objectives, and when sport-based approaches are likely to have less of an impact. Outlined below are recommendations that may allow these questions, and others, to be addressed in the detail they deserve. However, the overarching recommendation for all organizations, researchers, funders, and policy makers, guided by the Bioecological Systems Theory, is:

> **Consider multiple systems (e.g., microsystem, mesosystem, exosystem, macrosystem), levels of influence (e.g., individual, school, community, policy), and influencers (e.g., parents, peers, youth workers, teachers, funders), and the interaction of these factors, in the development, funding, implementation, evaluation, and policy of/for SfD interventions, over time. While there may be limits**

**to the scaleability of SfD interventions, as organizations, funders, researchers, and policy makers must incorporate specific adaptations based on historical, social, cultural, and political considerations, this requires knowledge of these factors at the local level, and beyond. Thus, systems thinking and systems change should guide the efforts of organizations, researchers, funders, and policy makers as they operate within/outside of the SfD field.**

## ORGANIZATIONS

A number of recommendations emerge from this review that are relevant to SfD organizations specifically, along with broader youth development organizations. Overall, many of the programmatic and contextual factors are analyzed and discussed in the context of an SfD intervention that isn't reaching significant outcomes, so these should be carefully considered.

1. Rather than viewing sport as a tool for development or as possessive of a particular developmental instrumentality, it may be more accurate and effective to think of sport as but one mechanism within a larger context.
2. Consider the most appropriate target population for an intervention, as those with higher baseline risk levels and/or lower life skill scores may have the greatest improvement over time.
3. Eschewing the deficit discourse, interventions are most effective when they pursue strength over punishment and when the skills, abilities, and resiliency of youth are nurtured, rather than punishment meted out for failing to succeed or behave as expected.
4. Develop contextually-relevant theories of change for interventions that identify appropriate outcomes for which to strive and assess.
5. Design curricula around predictors of ongoing engagement.
6. In the broadest terms, interventions are likely to do better or achieve the greatest success and positive impact when the focus is on people, more than interventions or practices. Recognizing stakeholders, providing training, and pursuing relationships are all important, which suggests the need to focus on the people, more so than the structures. The structures might be the reason that an intervention will fail, but they are unlikely to be the reason that an intervention thrives.
7. Certain features of interventions can be scaled up, but these features must be considered and delivered in culturally sensitive manners. For example, making efforts to develop quality relationships is a feature that must be scaled up and implemented across interventions. However, the exact manner in which these relationships are nourished over time must be culturally sensitive.
8. Integrate measurement of intervention quality and fidelity into measurement, evaluation, and research efforts.
9. Integrate measurement of critical factors (e.g., contextual factors, stakeholders, policy) into measurement, evaluation, and research efforts.

10. Report negative and null outcomes from interventions. This will enhance the transparency of research and evaluation efforts in the field, which can lead to the identification of best research and evaluation practices, the continuing development of the field's knowledge base with a stronger understanding of how this knowledge was produced, and identification of gaps and/or common barriers which must be explored in more depth.

## RESEARCHERS

While the review team certainly recognizes (and has experienced) the challenges inherent in conducting research in the SfD field, this does not change the need for more rigorous studies. In other words, simply because it is challenging to conduct rigorous research does not mean the standards should be relaxed. As a field, we must respond to this challenge with a renewed focus on ways to enhance rigor such that we can learn, improve, and ultimately advance the field of SfD. Recommendations which may lead the way include:

1. Utilize multiple groups, as there is an over-reliance on single group studies.
2. Increase the number of studies that are multi-site, comparison, and/or longitudinal.
3. Account for confounding explanations of results within study designs.
4. Decrease reliance on attitudes, knowledge, or perceptions, as there is little support that changes in knowledge and attitudes lead to behavior change.
5. Clearly report sampling procedures.
6. Fully integrate studies across philosophical, theoretical, methodological, and analytical perspectives.
7. Contextualize research within the social, cultural, and political climate of the community.
8. Ground research in methodological and/or theoretical basis to ensure critical analysis.
9. Assess intervention quality and fidelity, enabling the differentiation of outcomes resulting from ineffective intervention design and implementation vs. outcomes resulting from effective intervention design and implementation (leading to valid learning for SfD as a field).
10. Examine enrollment and retention rates for SfD interventions, compared with non-sport youth development interventions, which will enable the evaluation of sport as a 'best buy' approach that can reach (and thus impact) more youth.
11. Provide quality training and education for academics and measurement & evaluation (M&E) personnel, such that methodological rigor can be met at a high standard.
12. Consider new research directions that would inform intervention design and practice, as well as the overall development of the field, based on the gaps identified in this review related to thematic areas (e.g., employment) and critical factors (e.g., contextual factors, role of stakeholders, policy).

## FUNDERS

Recommendations for funders begin with the need for more meaningful support for research and evaluation. Research lines and methodologies in SfD have not advanced as they should over the past 20 years, which could be partially attributed to a lack of funding which would provide the training, resources, and time required for rigorous research and evaluation. Without this support, researchers and M&E personnel are pushed towards 'low cost' work, which often excludes experimental, longitudinal, and multi-site designs. Additionally, the overwhelming focus on intervention outcomes within the SfD field is likely reinforced by funders who may only require evidence on intervention outcomes. This outcomes-focused environment leaves many gaps in the SfD field, such as developing a better understanding of 'what works' and 'what influences' SfD interventions, and 'why'. Recommendations to address these concerns, among others, include:

1. Prioritize funding for research and evaluation in SfD that is rigorous and (frequently) resource-intensive.
2. Create diverse funding streams which support: (a) research examining questions about intervention efficacy (e.g., rigorous experimental research designs) and (b) research examining questions about beneficiaries, moderating variables, etc. (e.g., alternative/flexible research designs that are still rigorous).
3. Support efforts to take a systems approach to research and evaluation, with stakeholder and community engagement that frequently requires time, resources, and expertise.
4. Set expectations (with associated funding and support) for organizations to assess intervention quality and fidelity, along with the critical factors that may impact intervention efficacy.
5. Consult with organizations to determine what assessments are useful and appropriate.
6. Consistently communicate and support organizations in identifying and reporting null/negative findings. When organizations are able to operate in a funding climate where assessments can legitimately be framed to demonstrate 'what works' but also 'what needs to improve', the entire SfD movement stands to benefit. Not only will this result in more honest scholarship but more authentic partnerships that can truly be designed to address the organization's needs. Additionally, the resultant findings can add to the SfD knowledge base in meaningful ways by being constructively critical rather than evangelical of sport's developmental potential.
7. Support organizations for longer funding cycles to reduce the fear of funding being removed after a short period of time, enabling stronger funding relationships to be developed and allowing time for growth to occur.

## POLICY MAKERS

In several notable respects, the policy implications that emerged from this systematic review largely corroborate key insights from the burgeoning SfD policy literature. In particular, there is a need to situate the role and place of sport within the broader social and cultural contexts, and organizational dynamics, in which SfD takes place. Policy recommendations that proceed from this include:

1. Focus on people and their relationships, particularly with a needs-based approach, while supporting the integration of SfD within the larger youth development sector and policy agenda.
2. Position SfD as a complement to other youth programming and/or a means to fill gaps in the larger youth development sector.
3. Approach the process of ‘scaling up’ in a contextually specific manner, based on the needs of the targeted population or community.
4. Develop intervention theories in SfD that can help to explain how and why sport leads to certain outcomes.
5. Consider lobbying for and/or contributing to an overarching policy and funding framework to guide the SfD field and the research that takes place therein.
6. Broaden the conceptualization of what counts as data and/or evidence in the appraisal of SfD interventions, as well as consider the types of data and research designs that are needed to answer specific questions within the field.