A systematic review of sport-based life skills programs for young people: The quality of design and evaluation methods

Authors: Charlotte Williams[1], Rich Neil*[1], Brendan Cropley[2], Tim Woodman[3], Ross Roberts[3].

[1]Cardiff School of Sport and Health Sciences, Cardiff Metropolitan University
[2]School of Health, Sport and Professional Practice, University of South Wales
[3]School of Sport, Health and Exercise Sciences, Bangor University

* Corresponding Author: Contact – rneil@cardiffmet.ac.uk

Abstract

Over the past two decades, researchers have reported positive life skills outcomes for young people participating in sport-based life-skills programs. However, to date, there has been a lack of consideration in the literature regarding the quality of the programs designed and the evaluation methods adopted. Therefore, we conducted a systematic review of the life skills literature to: (a) assess the quality of sport-based life skills program design and evaluation methods; and (b) identify characteristics relating to the quality of sport-based life skills programs where authors had evidenced life skills development *and* transfer. Using the PRISMA guidelines, we searched six databases for relevant research papers and applied inclusion and exclusion criteria to the papers returned, of which 13 papers met the criteria. We conducted two quality assessment exercises (design and evaluation methods) and found two moderate-high quality life skills programs, ten moderate quality programs, and one low quality program. We present the characteristics (regarding quality) of intervention designs and methods, conclude with recommendations for designing quality sport-based life skills programs, and provide guidelines for researchers to evaluate sport-based life skills programs.

Lay Summary: Through engaging in sport-based life skills programs, young people can develop transferable skills. However, the quality of these life skills programs is unclear. We assess the quality of the design and evaluation methods of sport-based life skills programs, present the characteristics of moderate-high and moderate quality programs, and offer recommendations for future research and practice.

Practical Implications:

- The characteristics identified can be used to aid the development of the content, delivery and evaluation methods within future sport-based life skills programs.
- The quality assessment tool (QATID) that is embedded within this paper can be used by applied researchers to ensure that the design of their life skills interventions is of high quality.
- By using the QATID and the Mixed Method Appraisal Tool (MMAT) when designing and evaluating sport-based life skills programs, applied researchers can validate better subsequent claims of program effectiveness.

A Systematic Review of Sport-based Life Skills Programs for Young People: The Quality of Design and Evaluation Methods

Sport is a context in which young people can learn to develop functional skills that could be used in most aspects of life (Fraser-Thomas & Côté, 2009). These functional skills are often referred to by researchers in the field of Sport Psychology as *life skills*, and can be categorized as *behavioral*, *cognitive*, *interpersonal*, or *intrapersonal* skills (Danish, Forneris, Hodge, and Heke, 2004). Over the past three decades numerous researchers have developed, implemented, and evaluated programs within sport and physical activity contexts to promote the development of life skills in young people (under the age of 18). Indeed, programs such as Going for Goal (GOAL; Danish, 1992), Sports United to Promote Education and Research (SUPER; Danish, 2002), and The First Tee (Weiss, Stuntz, Bhalla, Bolter, & Price, 2013) have been used as mechanisms to evidence the positive relationship between sport participation and life skills development (e.g., Bean, Kendellen, & Forneris, 2016; Papacharisis, Goudas, Danish, & Theodorakis, 2005; Weiss et al., 2013). As a result of taking part in these life skills programs, researchers have proposed that young people can develop skills such as goal setting, emotional regulation, and communication.

Whilst young people appear to glean life skills via participation in sport, the pathway via which they do so remains unclear. To this end, Mahoney, Eccles, and Larson (2004) proposed that the structure and delivery of youth-based activities can determine whether young people experience positive or negative outcomes. Specifically, Mahoney (2000) noted that intentionally structured programs with clear program outcomes tend to lead to more favourable developmental results than non-structured programs. Advancing this perspective, researchers introduced the notion of implicit and explicit life skills development and transfer (Bean, Kramers, Forneris, & Camiré, 2018; Turnnidge, Côté, & Hancock, 2014). Specifically, an implicit approach denotes the conditions coaches put in place to facilitate life skills development and transfer, without those delivering the program having to discuss life skills

development or transfer (Turnnidge et al., 2014). In comparison, an explicit approach consists of those delivering life skills programs drawing upon specific pedagogical strategies to facilitate life skills development and transfer.

Researchers have claimed 'effectiveness' of these life skills programs through illustrating that participants developed and/or transferred (to a different context from sport) life skills. Each of these programs varies in relation to the design and evaluation methods adopted by researchers. Due to the variations across programs, it is often difficult to synthesize knowledge and, thus, compare life skills programs (Hodge, Danish, & Martin, 2012). In the broader field of positive youth development, researchers have attempted to synthesise knowledge through publishing an array of critical reviews. These include: a qualitative meta-study of positive youth development through sport (Holt et al., 2017); a systematic review on the impact of sport on the positive youth development of high performance athletes (Rigoni, Beleem, & Vieira, 2017); an integrative review of sport-based youth development literature (Jones, Edwards, Bocarro, Bunds, & Smith, 2017); a systematic review of life skills devlopment through sports programs serving socially vulnerable youth (Hermens, Super, Verkooijen, & Koelen, 2017); a systematic review of sport-based youth development programs in the United States (Whitley, Massey, Camiré, Boutet, & Borbee, 2019a); and a systematic review of sport for development interventions across six cities (Whitley et al., 2019b). Each of these reviews has enhanced our knowledge and understanding of positive youth development within a sport context. However, an important stage within a systematic review is establishing the quality of the papers included within the review and the quality of methods adopted by the reviewer. In doing so, this helps to increase a reader's level of confidence in the results presented by the researchers who conducted the systematic review, and minimises risk of bias. In reviewing the quality of the design of youth development programs and/or the quality of evaluation methods adopted, those conducting systematic reviews can assess the strength of researchers' claims of intervention effectiveness. That is, through assessing quality we can start

to identify if the outcomes presented by researchers can be believed (Higgins, 2008). Whilst evaluating the quality of papers within a review has been noted as an integral stage within the systematic review process, few researchers in the domains of life skills development through sport and positive youth development have focused their reviews entirely on assessing *quality*. Rather, researchers have attempted to assess quality as a secondary aim within their review (e.g., Holt et al., 2017) or have assessed the quality of papers as a means to determine which papers to include/exclude within their review (e.g., Hermans et al., 2017). Indeed, only two groups of authors have focused their review primarily on assessing the methodological quality of youth development programs (e.g.,Whitley et al., 2019a; 2019b). As such, only two of the above review papers examined methodological quality in sufficient breadth and depth.

In 2017, both Holt and colleagues, and Hermans and colleagues attempted to assess the methodological quality of the papers included within their review. To assess quality, Holt et al. conducted a meta-method analysis whereby they appraised the strengths and weaknesses of the methods employed by researchers. From this, Holt et al. concluded that the strengths of the studies were attributable to "multiple data collection and validity techniques, which facilitated the production of high-quality data" (Holt et al., 2017, p. 38). Whilst Holt et al. have attempted to explore quality, the main purpose of their review was not to evaluate quality; rather, their focus was on creating a model of positive youth development. Additionally, they drew conclusions relating to 'high-quality data' without engaging in a formal analysis of 'quality'. Indeed, their conclusions are based on two aspects of methodological quality (i.e., data collection methods and validation techniques, such as member checking). As such, it is important for researchers to adopt explicit, validated strategies to assess a broad and comprehensive range of methodological quality indicators in order to make valid claims in relation to the quality of studies. It is important to note, that Holt et al. (2017) may not have disclosed the specific protocols followed to evauate quality due to publication restrictions (e.g., an 8000 word limit).

Hermens et al. (2017) adopted a different approach to assessing quality by evaluating the 'rigour' of studies using the results as an inclusion criteria for their review. Specifically, they utilised the TAPUPAS (Transparance, Accuracy, Purposivity, Utility, Propriety, Accessibility, Specificity) framework (Pawson, Boaz, Grayson, Long, & Barnes, 2003) and postulated that only papers with medium, or high rigour would be included within their review. In adopting this approach, Hermens et al. made attempts to infer that the results of the papers included within the review were valid. Recently, Whitley and colleagues (2019a; 2019b) conducted two reviews of sport-based youth development programs and explicitly focused on assessing the methodological quality of research. Specifically, Whitley et al. (2019a) conducted a review of sport-based youth development programs, assessing the methodological quality of evaluations of sport-based youth development programs in the USA, with the aim of identifying characteristis of intervention efficacy. Their results, in relation to quality, reflected "weak" and "incoherent" interventions. Due to the low quality of interventions, they were unable to identify the characteristics of effective programs. Consequently, whilst researchers have claimed that sport-based youth programs can enhance the development of life skills, the quality of the evaluation methods used by researchers to evaluate the programs is weak. Whilst Whitley and colleagues (2019a; 2019b) enhanced our understanding of quality and the relationship between methodological quality and youth development program outcomes, they did not consider the quality of the design of such programs. A lack of consideration for the quality of program design is also evident within the broader context of the youth development literature. Indeed, there has been no focus on whether the design of life skills interventions are of high quality. Thus, researchers' reports of intervention effectiveness is questionable. Therefore, it is imperative to explore the quality of design and the quality of evaluations of sport-based life skills programs.

Purpose and Aim

Given the aforementioned variations across studies, and the lack of research assessing both the evaluative and design quality of life skills interventions, we sought to conduct a systematic review. Through adopting a systematic process of identifying, appraising, and synthesizing the results of all relevant individual research papers, we can begin to determine the quality of sport-based life skills interventions. Through conducting the systematic review, we, therefore, aimed to assess the quality of *design* and *evaluation* methods of sport-based life skills programs. By assessing the quality of existing life skills research, we hope to encourage researchers and practitioners to consider and/or improve the quality of life skills program design, and the methodological quality of the evaluations they conduct. In doing so, they may be able to evidence more reliably that life skills were developed and transferred (Higgins & Green, 2011). As a result of conducting a systematic review, we may also uncover areas where knowledge may be limited (Higgins & Green, 2011).

Method

Definitions

For the purpose of this paper, we are concerned with reviewing sport-based life skills programs as opposed to life skills development efforts within traditional youth sport programming. The distinguishing feature of sport-based life skills programs being that sport-based life skills programs have been developed by researchers and/or practitioners to explicitly focus on facilitating the development of life skills in young people through sport.

In order to conduct the review, it was important to define *life skills*. Currently, within the sport psychology domain, a number of definitions exist that have been developed to describe the term life skills. For example, Danish et al. (2004) defined life skills as, “Skills that enable individulas to suceed in the different environments in which they live, such as school, home, and in their neighborhoods” (p. 40). Further, Danish et al. (2004) considered life skills as *behavioral* (e.g., communicating effectively with peers/adults) or *cognitive* (e.g., making effective decisions), and *interpersonal* (e.g., being assertive), or *intrapersonal* (e.g., setting

goals) in nature. Building on this definition, Gould and Carson (2008) proposed that life skills are, "Those internal personal assets, charactersitics and skills such as goal setting, emotional control, self-esteem, and hard work ethic that can be facilitated or developed in sport and are transferred for us in non-sport settings" (p. 60). Whilst the definitions offered by Danish, Gould and associates provided a foundation for life skills research, no acknowledgement of the life skills transfer process was included within their work. Consequently, we provide our own definition of life skills to guide this review:

> "[life skills] are functional skills that individuals develop and use effectively in one context to manage demands (such as the home, school, sport, community, workplace) and that are also used effectively in other contexts beyond that in which they were learnt."

Search Strategy

Prior to developing the search strategy, we consulted with the lead author's institution librarian who supported the identification of the databases listed below and the development of the search terms used within this review. We employed an electronic search strategy for published studies using the following databases: (i) EBSCOhost; (ii) SPORTDiscus; (iii) Education Research Complete; (iv) PsycInfo; (v) PsychArticles; and (vi) Psych Source. We chose these databases as they were deemed the most suitable databases for the topic and would ensure that all relevant studies were detected. Keyword combinations used in the search strategies included the following Boolean search terms: Life skills *OR* Life skills Development *OR* Life skills Intervention *OR* Life skills Program *OR* (Positive Youth Dev* OR PYD) *AND* Sport *OR* Physical activ*. Further, we also searched these databases for known authors in the field (e.g., Danish). We also conducted a hand search of available literature to ensure that eligible papers were not missed. To action this, we scanned the reference pages of all of the included papers and published review papers in the field of life skills development through sport (e.g., Gould & Carson, 2008; Holt et al., 2017) for further relevant research articles.

Eligibility Criteria

The criteria for inclusion in the systematic review were: (1) peer-reviewed journal articles; (2) articles published in English between 1985 to the last search conducted in November 2019; (3) young people under the age of 18 years old were reported as participants; (4) sport-based life skills programs were the primary interventions reported. That is, sport programs that were developed to specifically facilitate life skills development and/or transfer; (5) life skills development and/or transfer was identified as the primary aim of the program; and (6) life skills outcomes were assessed or described. That is, there was evidence (qualitative or quantitative) of participants developing and/or transferring life skills.

We applied the following exclusion criteria: (1) adults over the age of 18 years old were reported as participants; (2) abstracts, book chapters, conference proceedings, dissertation abstracts, editorials, forewords, or review papers; (3) articles with life skills in the title, but where no reference to life skills is provided in the full body of text; (4) sport-based programs where the main aim was to develop outcomes such as well-being, academic improvement, or drug prevention; (5) programs that solely claim implicit development of life skills; and (6) life skills outcomes were not assessed or described (i.e., there was no qualitative or quantitative evidence of participants of developing and/or transferring life skills).

Procedure

Systematic review team. Our review team consisted of the lead author, and the second and third authors. At each stage of the process (search, screening, and data analysis), we met to discuss and challenge key decisions. In total, we met four times, with the lead and second author meeting a further four times.

Search and reporting process. Initial team discussions centered around the inclusion of individual life skills, such as (but not limited to) ‘team work’ and ‘communication’. Due to the vast array of individual life skills that there could potentially be, we (the review team) decided to use the search term ‘life skills’ as an umbrella term to encapsulate all potential life

skills. The lead author conducted the electronic search exercise. Following this, all returned articles were stored in an electronic folder in Mendeley, a reference management tool. Manual search procedures were also conducted whereby the lead author searched peer-reviewed journals and the reference lists of life skills review papers.

We followed the guidelines provided within the 27-item Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) to conduct the systematic review and report the findings of the review (Moher, Liberati, Tetzlaff, & Altman, 2009). In line with the PRISMA guidelines, the lead author identified the studies and removed all duplicate papers. Following this, the lead author screened all titles and abstracts. During the screening process, discussions between the lead and second author took place, and centered on issues with one particular criterion, that 'life skills are the main aim of the program'. Specifically, within some papers we found it difficult to decipher the primary aim of the research. As a result of these discussions, we (lead and second author) agreed to advance any ambiguous papers to the full text stage. At full text stage we made the decision to remove any papers in which life skills development as the primary aim could not be identified, and where it was unclear if life skills outcomes were assessed or described. At this stage, the lead author applied the inclusion and exclusion criteria to full texts (n = 79) to assess each paper's eligibility for inclusion. The lead author then presented the eligibility of each of the full texts (n = 79) to the second and third authors. Here, we discussed all papers and their eligibility for inclusion. It was at this point that we reached consensus, which resulted in the inclusion of 15 papers (see Figure 1).

Quality assessment. We conducted two quality assessment exercises: (a) to assess the quality of the design of each life skills program; and (b) to assess the quality of the evaluation methods adopted by each research team.

Design Quality. Despite the existence of a body of research devoted to enhancing program evaluation (cf. Wholey, Hatry, & Newcomer, 2010), there appears to be no formal assessment tool that can be used to assess the quality of an intervention design. Therefore, we

used two existing quality assessment guides: The Consolidated Standards of Reporting Trials (CONSORT) statement (Schulz, Altman, & Moher, 2010) and The *QualSyst* (Kmet, Lee, & Cook, 2004) to develop our criteria for intervention design quality. These tools were developed by researchers predominantly to assess the methodological quality of interventions. However, the authors of the protocols identified the following indicators of good intervention design: *theoretical underpinning* – intervention designs are informed by theory; *intervention description* – interventions are described clearly and in depth; *duration of intervention* – intervention duration is justified and appropriate for behavior change to occur; and, *implementation fidelity* – the intervention is delivered as intended. These indicators have also been identified as appropriate markers of intervention design quality by other authors (e.g., Davies, Walker, & Grimshaw, 2010; Jackson & Waters, 2005). We also searched the wider literature base (i.e., sport & exercise psychology, health, health psychology, and education journals) and found that some researchers had identified other criteria to assess intervention design quality. These included: *individualization within program* – the intervention is bespoke for each participant's needs*; ongoing feedback* – each participant receives ongoing and tailored feedback; *intervention piloted* – the intervention is piloted, reflected upon and, where required, revised*;* and *intervention directed at intended outcomes (intervention focus)* – the intervention is designed to improve the variables measured (Mullen, Green, & Persinger, 1985). Collectively, these indicators formed the criteria for our quality assessment tool for intervention designs (QATID): (a) *theoretically underpinned*; (b) *intervention description*; (c) *duration of intervention*; (d) *implementation fidelity*; (e) *individualization within program*; (f) *ongoing feedback*; (g) *pilot intervention*; and (h) *intervention focus*. The QATID was developed specifically for use within this study, however, there is potential for the QATID to be used on a wider scale to evaluate the quality of intervention designs.

As a review team, we reviewed existing quality assessment scoring systems and calibrated scoring system according to the question, "*Does the intervention adhere to the*

specific quality criteria?" (Yes = 2, Partial = 1, No = 0; Kmet et al., 2004). To reduce bias, and increase the reliability of the quality assessment, the lead author independently assessed each paper (n = 15) against the quality of intervention design criteria, and the second author followed the same process for all papers (n = 15). Mutual agreement was made between the two reviewers. We recorded 93% agreement prior to discussion, and 100% post discussion. Discussions here centered on determining the classification of a life skills theory (e.g., BNT/LDI, Hodge et al., 2012) and a framework or model (e.g., Positive Youth Development; Petitpas, Cornelius, Van Raalte, & Jones, 2005). We assigned papers that used a life skills theory to underpin the program with a score of 2 as a theory can be used to explain relationships, and we assigned papers that incorporated a life skills framework or model with a score of 1, as these are used to describe relationships. Once we had rated each criterion, we attributed an overall score to each paper. Quartile cut-off points have been used by researchers to categorize levels of quality (e.g., Robertson et al., 2018). Thus, we used the following cut-off points to categorize levels of quality: overall scores from 13 to 16 were high quality, overall scores of 9-12 were moderate-high quality, overall scores of 5-8 were moderate quality, and overall scores of 1-4 were low quality.

Evaluation Quality. To assess the methods adopted by researchers to evaluate life skills programs within each study, we used the *Mixed Methods Appraisal Tool* (MMAT; Pluye et al., 2011). As the researchers of the included studies used a variety of evaluation methods, we determined that the MMAT was the most appropriate quality assessment tool to use. The MMAT was designed to evaluate the methodological quality for three domains of research: (1) qualitative research; (2) quantitative research (divided into three sub-domains; descriptive, randomized control, and non-randomized control); and (3) mixed-methods research. The qualitative criteria outlined in the MMAT includes: appropriateness of sampling procedure; appropriateness of data analysis processes; consideration of context on data collection procedures; and consideration of researchers' philosophy (i.e., ontological and epistemological

beliefs). The quantitative descriptive criteria consisted of: relevance of sampling strategy; appropriate representation of sample; appropriateness of measures; and acceptable response rate. Finally, the quantitative non-randomized criteria within the MMAT includes: minimization of selection bias; appropriateness of measures; comparable groups; and acceptable response rate (for specific criterion requirements see Mixed Methods Appraisal Tool guidelines; Pluye et al., 2011).

Each quality indicator is rated on a categorical scale (yes, no, and cannot tell), and the number of yes answers are added together to create an overall score. The overall score (reflected as an overall percentage) was calculated by adding the total number of 'yes' items, dividing this by four, and multiplying this by 100. So, if two out of four were scored as 'yes' we divided two by four, which gave 0.5 and multiplied this by 100 to get the percentage of 50%. Therefore, scores varied from 25% (one criteria met) to 100% (all four criteria met). In line with Robertson et al. (2018), we categorized papers with overall scores of 100% as high quality, overall scores of 75% as moderate-high quality, overall scores of 50% as moderate quality and, overall scores of 25% as low quality. When testing the reliability and efficiency of MMAT, researchers have reported that the consistency of the global score between reviewers (ICC) is between 0.72 and 0.94 (Pace et al., 2012). In line with the design quality assessment exercise, the lead author independently assessed each paper (n = 15) against the MMAT, and the second author followed the same process for all papers (n = 15). We recorded an agreement score of 87% agreement prior to discussion, and 100% post-discussion. Discussions centered on the ambiguity of information (e.g., there were times when we assumed information within the paper). Thus, we agreed to score the paper only if the information was present.

Overall Quality. To determine the overall quality of each paper, we converted the QATID scores into percentages so that they were in line with the MMAT scores. We took the total number scores, divided it by 16 (the total score available) and then multiplied this by 100. We then calculated the mean percentage for the two quality scores for each paper (see Table 4).

Data extraction and synthesis. After determining scores for each paper for the quality of design and evaluation, we (the first and second authors) used the quality criteria from both assessment tools to inform the development of a paper-based data extraction form. The data extraction form included generic information such as the author(s) and year of publication. In addition, the data extraction form included the following information relating to the quality criteria derived from the QATID: theoretical underpinning; intervention focus (i.e., clear program goals, clear session descriptions, life skills embedded into program content and delivery); program description (i.e., context, location, structure, life skills); duration; individualization; ongoing feedback (i.e., strategies used); pilot implementation; and program fidelity. Further, the data extraction form also included the following information from the MMAT with regards to the quality of program evaluation: sampling procedure (i.e., sample size; participant demographics); data analysis process (i.e., domain; qualitative/quantitative, methods, frequency of evaluation); measures (i.e., type of measure, validity of measure); comparable groups (i.e., control group); and researcher philosophy. In the following section, we present the data extracted through descriptive narrative.

Results

Quality Assessment Result

We assessed 15 studies against the QATID, a total score of 16 represented the highest score that any paper could achieve. Scores ranged from 3 to 10, with Huysmans, Clement, Whitley, Gonzalez, and Sheehy (2019) the only one to achieve a score of 10 (see Table 1).

Of the 15 studies we assessed against the MMAT (Pluye et al., 2011), eight were assessed against the qualitative criteria, four studies against the quantitative non-randomized criteria, one study against the quantitative descriptive criteria, and two against the mixed method criteria (see Table 2). For the eight qualitative studies, overall scores ranged between 25%-100% with three studies scoring above 50% (i.e., Bean et al., 2016; Holt et al., 2013;

Jacobs & Wright, 2019). The quantitative non-randomized studies and the quantitative descriptive scores were all 50%. The scores for mixed method studies were 50% (see Table 2).

We combined the results from both of the quality assessment exercises (see Table 2) and categorized papers into the following quartiles: high, moderate-high, moderate, or low quality (Robertson et al., 2018). We categorized three papers as moderate-high quality (62.5%-75%), 11 papers as moderate quality (34.5%-56.5%), and one paper as low quality (28%). Furthermore, in line with our working definition of life skills whereby transfer is highlighted as a crucial aspect within the life skills domain, we have provided overall quality scores for the papers in which life skills development *and* transfer is evidenced (see Tables 2 and 4). These included the three papers identified as moderate-high quality (i.e., Bean et al., 2016; Huysmans et al., 2019; Jacobs & Wright, 2019); six of the 11 papers identified as moderate quality (i.e., Goudas & Giannoudis, 2010; Hodge Kanters, Forneris, Bocarro, & Sayre-McCord, 2017; Holt et al., 2013; Waldron, 2009; Weiss et al., 2013; Weiss, Bolter, & Kipp, 2016); and the one paper identified as low quality (i.e., Lee, Park, Jang, & Park, 2017).

Summary of Studies

In this section, we provide a descriptive overview of the design and evaluation quality characteristics that we obtained through the data extraction process. Split into two sections, we first provide an overview of the *quality of the design of sport-based life skills programs* and then we offer insight into the *quality of evaluation of sport-based life skills programs*. Each section is split further into sub-sections that illustrate the characteristics of either the design or evaluation of life skills programs. Within each sub-section, we first provide an overview for all of the papers that met the inclusion criteria (n = 15). Then, and in coherence with our working definition that highlights transfer as an important factor within the life skills domain, we illustrate the design and evaluation quality characteristics in relation to *only* the three moderate-high and six moderate papers (n = 9) that we identified in the previous section, in which the authors evidenced life skills development *and* transfer (see Table 3 & Table 4).

Quality of the design of sport-based life skills programs.

Underpinned by theory. We extracted data in relation to the *theoretical underpinning* of the programs. Of the 15 papers included within this review, the authors of only five papers referred to the program being underpinned by a 'theoretical' youth development framework. The authors of the remaining ten papers did not make reference to the program being underpinned by any theoretical framework. Of the nine moderate-high and moderate quality papers in which the authors evidenced life skills development *and* transfer, two group of authors made reference to using Hellison's (1995) Teaching Personal and Social Responsibility framework (TPSR; Bean et al., 2016; Huysmans et al., 2019) and two groups of authors made reference to using Petitpas et al.'s (2005) Positive Youth Development framework as an underpinning theoretical approach (Weiss et al., 2013; Weiss et al., 2016). The authors of the remaining five moderate and moderate-high quality papers in which the authors evidenced life skills development *and* transfer did not report a theoretical underpinning.

Intentional focus. The authors of each of the 15 papers included within this review demonstrated an intentional focus on life skills development and/or transfer. We refer to intentional focus as the designing of life skills programs to promote life skills development. We identified three factors that contribute toward program focus: clear program goals; clear session descriptions; and life skills embedded into the program content and delivery. Authors of the nine moderate-high and moderate quality papers who evidenced life skills development and transfer provided clear program goals, and embedded life skills into their program. However, four of these authors also provided clear session descriptions that would permit a practitioner to replicate the intervention (Bean et al., 2016; Goudas & Giannoudis, 2010; Hodge et al., 2017; Huysmans et al., 2019).

Program description. We identified that there were a range of sport-based life skills programs that have been developed, implemented, and evaluated in different parts of the world: Canada (3); Eswatini (1); Greece (4); Korea (1); and, USA (6). Of the nine moderate-high and

moderate quality papers in which the authors evidenced life skills development *and* transfer, five originated from the USA (Hodge et al., 2017; Jacobs & Wright, 2019; Waldron, 2009; Weiss et al., 2013; Weiss et al., 2016); two originated from Canada (Bean et al., 2016; Holt et al., 2013); one originated from Greece (Goudas & Giannoudis, 2010); and one originated from Eswatini (Huysmans et al., 2019). Overall, the 15 programs were delivered across two contexts, *Education* (9) and within the *Community* (6). The education context consisted of physical education (3) and extra-curricular activities (sport; 6). Five of the nine moderate-high and moderate quality programs in which the authors evidenced life skills development *and* transfer were delivered within the Community (Bean et al., 2016; Hodge et al., 2017; Huysmans et al., 2019; Jacobs & Wright, 2019; Waldron, 2009), two within Sport (Weiss et al., 2013; Weiss et al., 2016), and two within Education (Goudas & Giannoudis, 2010; Holt et al., 2017). Overall, the life skills included within the 15 programs were: goal setting (12), positive thinking (4), problem solving (5), communication (5), teamwork (7), health skills (3), leadership (5), social support (2), self-management (3), media skills (1), reflection (1), planning (2), seeking help (2), self-talk (3), social skills (1), relaxation (3), and values (1). Of the nine moderate-high and moderate quality programs in which the authors evidenced life skills development *and* transfer, the life skills included were: goal setting (7), positive thinking (1), problem solving (3), communication (4), teamwork (6), health skills (3), leadership (5), social support (2), self-management (4), media skills (1), reflection (1), planning (2), seeking help (2), self-talk (3), social skills (1), relaxation (3), and values (1).

Program duration. Within the 15 included papers, programs ranged from one week to two years, of which the number of sessions ranged between 3-57 sessions, and the duration of the sessions ranged between 10-100 minutes. With regards to the nine moderate-high and moderate quality papers in which the authors evidenced life skills development *and* transfer, the life skills programs were delivered: over three weeks (Huysmans et al., 2019); four weeks (Waldron, 2009); between 11 and 17 sessions (Goudas & Giannoudis, 2010; Hodge et al.,

2017); over two years (Bean et al., 2016); and, over three months (Holt et al., 2013). The authors of three of the moderate-high and moderate quality papers in which life skills development *and* transfer were evidenced did not state the precise duration of the program (Jacobs & Wright, 2019; Weiss et al., 2013; Weiss et al., 2016). This may reflect the type of community programs that they are, with no definitive start or end point. Catalano, Berglund, Ryan, and Hawkins, (2004) suggested that for youth development programs to foster change, they should run for a minimum of nine months or 10 sessions. Of the 15 papers included in the review, only one program ran for the minimum duration of nine months (Bean et al., 2016). Further, seven programs ran for the minimum duration of 10 sessions (e.g., Bean et al., 2015; Goudas & Giannoudis, 2008; 2010; Hodge et al., 2017; Holt et al., 2013; Huysmans et al., 2019; Lee et al., 2017). Of the nine moderate and moderate-high quality papers in which the authors evidenced life skills development *and* transfer, one program did not meet the duration criteria (Waldron, 2009); three groups of authors did not state the duration of their programs (Jacobs & Wright, 2019; Weiss et al., 2013; Weiss et al., 2016), and five programs met the minimum duration of 10 sessions and/or 9 months (Bean et al., 2016; Goudas & Giannoudis, 2010; Hodge et al., 2017; Holt et al., 2013; Huysmans et al., 2019).

Individualization. Ten out of the 15 teams of authors referred to individualization of the sport-based life skills program. Of the nine moderate and moderate-high quality papers in which authors evidenced both life skills development *and* transfer, six made reference to individualization within the program (i.e., Bean et al., 2016; Goudas & Giannoudis, 2010; Hodge et al., 2017; Holt et al., 2013; Huysmans et al., 2019; Jacobs & Wright, 2019). Individualization was illustrated through authors providing insight to the adaptations made to SUPER for respective contexts (Goudas & Giannoudis, 2010; Hodge et al., 2017); providing participants with the choice of activities that would be carried out as part of the program (Bean et al., 2016); creating bespoke core values (Jacobs & Wright, 2019); or adapting the life skills program based on the needs of the participants (Holt et al., 2013; Huysmans et al., 2019).

Ongoing feedback. Six out of the 15 teams of authors indicated opportunities for ongoing feedback within the life skills programs. These six papers were also moderate-high and moderate quality papers in which life skills development *and* transfer was evidenced. The feedback strategies that these six adopted to evidence life skills development *and* transfer included: debriefs to support progress (e.g., Bean et al., 2016; Jacobs & Wright, 2019); coach reinforcement to remind young people of the life skills that they were developing (e.g., Hodge et al., 2017; Jacobs & Wright, 2019; Weiss et al., 2013); ongoing feedback from parents through reinforcement through setting homework (e.g., Goudas & Giannoudis, 2010); and reflective practice (Huysmans et al., 2019).

Pilot. Of the 15 included papers, only one group of authors provided information pertaining to implementing pilot versions of the sport-based life skills program. This paper was of moderate quality and one in which the authors evidenced life skills development *and* transfer. Specifically, Holt et al. (2013) conducted an action research based study whereby they used data collected from the first phase of the study to influence changes made to the second phase of the study.

Intervention fidelity. Due to the integral role coaches play in the development of life skills in young people and the lack of formal training provided to those delivering youth sport programs (Petitpas et al., 2005), we specifically focused on coach training as an indicator of intervention fidelity. Only four out of the 15 groups of authors referred to any form of coach training. Of the nine moderate-high and moderate quality papers in which the authors evidenced life skills development *and* transfer, three groups of authors referred to coach training. Specifically, Weiss et al. (2013) outlined that coaches attended a two-day workshop, whereby they were taught four ‘deliberate teaching methods’. In addition, Jacobs and Wright (2019) made reference to ‘facilitators’ engaging in annual coach training through a national youth development sport organization. Huysmans et al. (2019) highlighted that coaches attended three days of training through a train-the-trainer approach. Whilst insight is given into

the teaching methods covered in the workshop, little insight is given into the detailed content and delivery of the workshop.

Quality of evaluation of sport-based life skills programs.

Sampling process

Sample size. Of the 15 studies, sample sizes ranged between six and 564. Samples ranged between six and 145 for the qualitative studies, between 72 to 564 for the quantitative studies, and between 15 and 36 for the mixed method studies. For the nine moderate-high and moderate quality papers in which the authors evidenced life skills development *and* transfer sample sizes ranged from 8-145 for the qualitative papers, 192-564 in the quantitative paper, and 36 within the mixed method study.

Participants. We extracted data related to the participants included within each of the 15 life skills programs. Participants were both male and female, aged between seven and 18 years old. In the nine moderate-high and moderate quality papers in which authors evidenced both life skills development and transfer, two groups of authors used only female participants between 11 and 16 years old (Bean et al., 2015; Waldron, 2009), and in the remaining seven papers, male and female participants between 11 and 17 years old were included.

Data Analysis

Domain and methods. We extracted the *domains* (i.e., qualitative or quantitative) and the *methods* that authors used to evaluate each of the 15 life skills programs. Authors adopted a qualitative approach and used methods such as interviews or focus groups in ten papers and adopted a quantitative approach and used methods such as questionnaires in the seven papers (figures inclusive of both mixed method studies). Of the nine moderate-high and moderate quality papers in which the authors evidenced life skills development *and* transfer, one team of authors (Weiss et al., 2016) used quantitative methods, seven teams of authors used qualitative methods (Bean et al., 2016; Goudas & Giannoudis, 2010; Hodge et al., 2017; Holt et al., 2013;

Jacobs & Wright, 2019; Waldron, 2009; Weiss et al., 2013), and one team of authors used mixed methods (Huysmans et al., 2019).

Frequency of evaluation. Of the 15 papers, authors conducted evaluations across four time points: pre-intervention (6); during the intervention (4); post-intervention (14); and during a follow-up period (5). In four of the nine moderate-high and moderate quality papers in which the authors evidenced life skills development *and* transfer, data was collated post-intervention (Bean et al., 2016; Holt et al., 2013; Waldron, 2009; Weiss et al., 2013). Two sets of authors of the nine moderate-high and moderate quality papers collated data during the intervention and post-intervention (Goudas & Giannoudis, 2010; Jacobs & Wright, 2019); one set of authors collated data during the intervention and after a follow up period (Hodge et al., 2017); one set of authors collected data pre intervention, post intervention and after a follow up period (Huysmans et al., 2019); and another collated data during the intervention, post-intervention, and after a follow-up period (Weiss et al., 2016).

Appropriate measures. We extracted data relating to the tools that authors used to measure life skills development and/or transfer. Specifically, of the 15 papers included in the review there were 15 different scales used to assess program effectiveness and, therefore, life skills development and/or transfer (see Table 4). Of the nine moderate-high and moderate quality papers in which the authors evidenced life skills development *and* transfer, one group of authors used a valid measure to evaluate life skills development (Huysmans et al., 2019). Further, of the nine moderate-high and moderate quality papers in which authors evidenced life skills development *and* transfer, only one group of authors used a valid measure specifically designed to evaluate life skills transfer, the Life Skills Transfer Scale (LSTS; Weiss, Bolter, & Kipp, 2014). Of the 15 papers included in the review, the authors of nine papers relied solely on self-report data, whilst the authors of six papers also included parent, coach, and/or facilitator perspectives of life skills development. Each of the authors of the nine moderate-high and moderate quality papers in which life skills development *and* transfer collected self-

report data, with the authors of five papers relying solely on self-report data. Authors of two of the moderate quality papers also considered parent responses (Hodge et al., 2017; Weiss et al., 2013), one group of authors considered coach responses to assess life skills development and transfer (Weiss et al., 2013), one group of authors also considered teacher responses to life skills development and transfer (Holt et al., 2013), and one group of authors considered both coach and teacher responses to life skills development and transfer (Huysmans et al., 2019).

Comparable groups

Control group. Authors of only five of the 15 papers within this review included a control group. Of these, only two were of moderate quality and evidenced life skills development and transfer (Waldron, 2009; Weiss et al., 2016).

Discussion

Through this systematic review, our aim was to assess the quality of *design* and *evaluation methods* of sport-based life skills programs in order to better understand the quality of existing life skills research, and to influence the quality of future research in the area of life skills development through sport. In doing this, we hoped to encourage researchers and practitioners to consider and/or improve the quality of the life skills program design and the methodological quality of the evaluations they conduct. As a result of considering quality, researchers and practitioners can assess the strength of their claims of intervention effectiveness and, thus, provide more credible findings. Fifteen papers met the inclusion criteria on which we conducted two quality assessment exercises. As a result of combining both design and evaluation quality, we determined three papers as moderate-high quality, 11 papers as moderate quality, and one paper as low quality. We then considered which of these studies evidenced life skills development *and* transfer. Those included in this analysis were the three moderate-high quality, six of the 11 moderate quality, and the one low quality paper. Whilst the number of papers in the moderate quality and above category reflects an increase in quality compared to the “weak” quality inferred by Whitley et al. (2019a; 2019b), it is

important to note that the quality score within our study is an indication of the quality of both design and evaluation methods. Further, of the 15 papers included in this review, only three were of moderate-high quality. As such, claims of effectiveness for those papers whereby quality is lacking should be interpreted with caution. Through a rigorous process of data extraction and analysis, synthesized within a narrative description, we have provided a descriptive overview of the characteristics of design and evaluation quality for moderate-high and moderate quality papers in which life skills were proposed to be developed *and* transferred. In this section we provide a discussion of these characteristics and then offer recommendations on how researchers and practitioners can increase the design quality of life skills program and the evaluation methods used.

Characteristics of Design Quality

From our synthesis, we observed that for the moderate-high and moderate quality programs whereby life skills were developed and transferred, the program goals were clear and life skills activities were embedded within the programs, illustrating a *focused intervention*. Further, supporting the work of Jones and Lavallee (2009), young people believed that the life skills included in the programs were important, and relatable to external contexts. Researchers of the moderate and moderate-high quality programs in which authors evidenced life skills development *and* transfer provided *program descriptions*, in which the following life skills were included: *communication, goal setting, teamwork, relaxation, self-talk, seeking help, leadership, planning, self-management, health, social support, reflection, media skills, social skills, positive thinking, problem solving, and values.* By collating these life skills, we have provided researchers and practitioners with a list of skills (derived from the moderate and moderate-high quality peer reviewed papers included within this review in which authors evidenced life skills development *and* life skills transfer) that are deemed valuable to the functional development of young people. We have also found support for Catalano et al.'s (2004) findings that effective youth development programs should be delivered for a minimum

of 10 sessions *or* 9 months to facilitate behavior change. Indeed, of the nine moderate-high and moderate quality papers whereby life skills development and transfer were evidenced, four groups of authors specifically stated that their respective programs ran for at least 10 sessions. Further, one group of authors stated their program ran for a minimum of nine months and specifically used Catalano's recommendations as a guide.

There were a number of design characteristics that were not demonstrated, which affected the quality of the research. Indeed, by failing to evidence these design characteristics, the overall quality score for program design was lower than what it would have been should these characteristics have been evidenced. First, only one of the nine teams of authors of the moderate-high or moderate life skills programs in which the authors evidenced life skills development *and* transfer presented information pertaining to a pilot intervention (i.e., Holt et al., 2013). Pilot interventions are important indicators of quality as they encourage researchers to review program content, evaluation methods against the desired program aims, and make necessary changes to ensure that the program is addressing the reported aims (McBride, 2016). Second, with the exception of three teams of researchers, authors provided no insight into the training offered to those delivering the life skills programs. Whilst we acknowledge that facilitator training is not the only indicator of program fidelity, insight into facilitator training is essential to support claims of effectiveness. Without this information it is unclear as to whether the facilitator delivering the program had the appropriate knowledge and understanding to deliver the program effectively. Researchers have also noted that facilitator training is important as it can help coaches to develop an awareness of their role in facilitating life skills in young people, and the strategies they can use to deliver life skills development and transfer (Camiré, Kendellen, Rathwell, & Charbonneau, 2018). Third, and similar to the low numbers reported within Holt et al. (2017) and Whitley et al.'s (2019b) reviews, only five life skills programs were theoretically underpinned. By utilizing theory to underpin their work, researchers may identify and then test hypotheses that help improve understanding on the

mechanisms that influence behavior. Such improved understanding may also help inform what content and strategies researchers/practitioners should include in programs to better facilitate life skills development and transfer (Prestwich et al., 2015). Our findings support Whitley et al.'s (2019b) call for researchers to begin to test intervention theories as opposed to intervention outcomes in order to identify the conditions and mechanisms that explain life skills development outcomes. Last, we also observed a lack of involvement from parents in the delivery and evaluation of life skills programs. Whilst researchers within the field of life skills development have focused on the role of the coach in delivering sport-based life skills programs (Camiré et al., 2012), very few researchers have explored the role of parents in sport-based life skills programs (e.g., Hodge et al., 2017). The role of parents is important as researchers have highlighted that the facilitation of life skills development in young people is the collaborative role of parents, coaches, and significant others such as teachers (Bowley, Cropley, Neil, Hanton, & Mitchell, 2018), as all these collaborators may have an influence on young people across many contexts.

Characteristics of Evaluation Quality

In the moderate-high and moderate quality papers in which authors evidenced development and transfer of life skills, there were aspects of the evaluation methods that were relatively consistent (e.g., sampling strategy [sample size and participants]; data analysis [e.g., domain, methods]; and appropriate measure [e.g., self-report data]). However, there were also notable differences, and missing information in relation to the evaluation methods (e.g., data analysis [frequency of evaluation]; appropriate measures [e.g., measuring life skills; multiple sources]; and comparable groups [e.g., use of control groups]; researcher philosophy). Specifically, researchers predominantly collated data post intervention, most likely due to the nature of the research (i.e., qualitative). In that sense, there appears to be an over reliance on qualitative research, potentially due to the lack of validated sport-specific measures that were available to researchers at the time of publishing, an issue first identified by Gould and Carson

in their review in 2008. In addition, there also appears to be an over-reliance on self-report data, which may be contaminated by such reliability related issues as memory recall and social desirability. Only two of the nine moderate-high and moderate quality papers in which the authors evidenced life skills development *and* transfer included a control group, which enabled them to infer that young people developed and transferred life skills as a specific result of participating in the life skills program. Thus, these researchers were able to evidence a causal relationship between life skills development and the life skills program. Further, researchers failed to describe the demographics of groups, and illustrate how they accounted for any differences between groups.

Future Research Recommendations

We recommend that those designing and evaluating life skills programs take steps towards ensuring their research is of high quality. We reported only three papers as moderate-high quality. From a design perspective, researchers wishing to publish a high-quality paper should ensure that they provide a detailed description of the life skills intervention alongside providing a clear program focus. Such a description would include providing insight into the theory used to underpin the program, the duration of the program, and the structure of the life skills program. Further, authors and/or practitioners should present the steps they took to individualize the life skills program, and if and how ongoing feedback was integrated within the program. Researchers should look to pilot their programs and share with readers the effects of the pilot intervention. Lastly, it is important that researchers who want to evidence high quality papers provide the reader with insight into life skills program fidelity. Whilst there are a number of indicators of program fidelity, we have identified facilitator training as an important avenue to enhance quality. That is, researchers should provide information about the training that facilitators of life skills programs undergo prior to delivering the programs.

From an evaluation perspective, researchers wishing to evidence high quality evaluations should look to include a control group which would enable researchers to compare

program effects. Furthermore, researchers should utilize the quality assessment criteria (as reported in this paper) to guide the development and evaluation of life skills programs. Specifically, those adopting qualitative approaches should, at minimum, "identify their disciplinary affiliation, what brought them to the question, and the assumptions they make about the topic of interest" (Caelli, Ray, & Mill, 2003, p. 5). This supports Whitley et al.'s (2019a; 2019b) suggestion that researchers must consider philosophical, theoretical, methodological, and analytical perspectives. In addition, those adopting quantitative approaches should consider presenting information relating to complete outcome data, including withdrawal/dropout. By illustrating this information, researchers may reinforce the strength of the methodological design and administration they have engaged with, and as such, the results of the study can be better generalized.

Limitations

When assessing the quality of studies, we acknowledge that there is a risk of evaluating too harshly, as the researchers whose studies we are evaluating may not have had the *space* (page limit) to disclose all relevant information. Further, when designing sport-based life skills programs, they may have been governed by external organizations who can restrict the scope of what is implemented. Given that the assessment of the quality of the design and evaluation methods adopted by researchers relies on sufficient information being provided, the results of our study should be considered with potential restrictions in mind. In addition, whilst we identified six databases that we felt were relevant and would provide the best opportunity for detecting relevant studies, it is important to note that it is difficult to identify all relevant studies. In order to minimize the risk of not including appropriate studies, we enlisted the support of the institution librarian to support with the development of the search strategy, included a search for known authors in the field, and hand-searched the reference list of relevant review papers and all included papers. Whilst these strategies were put in place to ensure that we did not miss any papers, there is always a risk of eligible papers being missed.

Indeed, we did not hand search the reference list of every single published review paper related to positive youth development and/or life skills. As such, relevant research papers may have been missed. Finally, we acknowledge that when designing and evaluating any intervention program there are numerous difficulties that researchers may face that may impact the evaluation methods adopted. For example, researchers' access to participants and/or the willingness of participants to engage fully in the research may influence how the intervention is designed and then delivered, and how it is evaluated throughout the program. In line with our recommendations, it is, therefore, important that researchers provide sufficient information related to the design and evaluation methods adopted, along with any restrictions and issues faced, to help readers make their own decisions about researcher statements of quality and intervention effectiveness.

Applied Implications

Schinke et al. (2020) recently identified a lack of rigor in relation to intervention design within the Sport Psychology domain. Schinke et al. also made reference to a lack of specific, high quality interventions within the field, highlighting that often high-quality research is perceived to be difficult to conduct. Throughout this review, we have made practice-related recommendations to researchers and practitioners proposing ways in which they can develop high quality program designs and evaluations. Indeed, researchers and practitioners can draw upon the quality guidelines outlined in the QATID to design high quality life skills interventions. Specifically, researchers/practitioners should use frameworks, such as the BNT/LDI (Hodge et al., 2012), to underpin the design of life skills programs in order to identify how and why specific outcomes of life skills programs arise (Whitley et al., 2019b). In addition, researchers/practitioners should ensure that life skills programs meet the minimum duration recommendations for behavior change of 9 months or 10 sessions (Catalano et al., 2004). Researchers/practitioners should conduct and evaluate pilot interventions whereby they assess the content and delivery methods of the program and make changes if necessary.

Finally, researchers/practitioners should consider how they will attempt to ensure programs are delivered as intended. We recommend that researchers/practitioners provide training for those delivering life skills programs to facilitate program fidelity and increase the quality of the life skills program. By using an intervention design-related quality assessment tool when designing sport-based life skills programs, researchers may be better able to validate subsequent claims of program effectiveness. Additionally, researchers and those practitioners wishing to assess the effectiveness of sport-based life skills programs must ensure they conduct quality evaluations, considering: (a) sampling procedures; (b) data analysis processes; (c) appropriateness of measures to evaluate life skills development; (d) inclusion of control groups; and (e) the role of the researcher in the context. These indicators are important to increase the quality in design and evaluation of sport-based life skills programs and should be used by researchers, practitioners, academics and reviewers in their development and assessment of papers to ensure that the issues relating to quality recently outlined by Schinke et al. (2020) are addressed.

Conclusion

We have conducted the first systematic review that has explored the quality of sport-based life skills development programs. Whilst it was difficult to compare these sport-based programs due to the significant variations in program design and evaluation methods, the present review provides insight into some important characteristics that influence the quality of sport-based life skills programs. From a practical application perspective, we highlight that these quality guidelines should be used by researchers and practitioners when designing and evaluating future sport-based life skills programs.

References

Bean, C., Forneris, T., & Fortier, M. (2015). Girls just wanna have fun: Understanding perceptions of effective strategies and outcomes in a female youth-driven physical activity-based life skills programme. *Journal of Sport for Development, 3,* 28-40.

Bean, C., Kendellen, K., & Forneris, T. (2016). Moving beyond the gym: Exploring life skill transfer within a female physical activity-based life skills program. *Journal of Applied Sport Psychology, 28,* 274-290.

Bean, C., Krammers, S., Forneris, T., & Camiré, M. (2018). The implicit/explicit continuum of life skills development and transfer. *Quest, 70,* 1-15.

Bowley, C., Cropley, B., Neil, R., Hanton, S., & Mitchell, I. (2018). A life skills development programme for youth football coaches: Programme development and preliminary evaluation. *Sport & Exercise Psychology Review, 14,* 3-22.

Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development.* Thousand Oaks, London: SAGE Publications

Braun, V. & Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology, 3,* 77-101.

Brunelle, J., Danish, S., & Forneris, T. (2007). The impact of a sport-based life skills program on adolescent prosocial values. *Applied Developmental Science, 11,* 43–55.

Caelli, K., Ray, L., & Mill, J. (2003). 'Clear as mud': Toward greater clarity in generic qualitative research. *International Institute for Qualitative Methodology, 2,* 1-13.

Camiré, M., Trudel, P., & Forneris, T. (2012). Coaching and transferring life skills: Philosophies and strategies used by model high school coaches. *The Sport Psychologist, 26,* 243-260.

Camiré, M., Kendellen, K., Rathwell, S., & Charbonneau, E. (2018). Evaluation of the pilot implementation of the coaching for life skills program. *International Sport Coaching Journal, 5,* 227-236.

783 Catalano, R., Berglund, M., Ryan, J., & Hawkins, J. (2004). Positive youth development in the
784 United States: Research findings on evaluations of positive youth development programs.
785 *The Annals of the American Academy, 591,* 98-124.
786 Danish, S. J. (2002). Teaching life skills through sport. In M. Gatz, M. Messner & S. Ball-
787 Rokeach (Eds.), *Paradoxes of youth and sport* (pp. 49-60). Albany, NY: State University of
788 New York Press.
789 Danish, S. J., Petitpas, A., & Hale, B. (1993). Life development intervention for athlete's life
790 skills through sports. *The Counseling Psychologist, 21,* 352-385.
791 Danish, S., Forneris, T., Hodge, K., & Heke, I. (2004). Enhancing youth development through
792 sport. *World Leisure Journal*, *46*, 38-49.
793 Danish, S. J., Mash, J., Howard, C., Curl, S., Meyer, A., & Owens, S. (1992a). *Going for the*
794 *goal leader manual*. Richmond, VA: Virginia Commonwealth University.
795 Davies, P., Walker, A., & Grimshaw, J. (2010). A systematic review of the use of theory in the
796 design of guideline dissemination and implementation strategies and interpretations of the
797 results of rigorous evaluations. *Implementation Science, 5,* 1-6.
798 Fraser-Thomas, J., & Côté, J. (2009). Understanding adolescents' positive and negative
799 developmental experiences in sport. *The Sport Psychologist, 23*, 3–23.
800 Goudas, M., & Giannoudis, G. (2008). A team-sports-based life-skills program in a physical
801 education context. *Learning and Instruction*, *18*, 528-536.
802 Goudas, M., & Giannoudis, G. (2010). A qualitative evaluation of a life-skills program in a
803 physical education context. *Hellenic Journal of Psychology*, *7,* 315-334.
804 Goudas, M., Dermitzaki, I., Leondari, A., & Danish, S. (2006). The effectiveness of teaching a
805 life skills program in a physical education context. *European Journal of Psychology of*
806 *Education, 21,* 429-438.
807 Gould, D., & Carson, S. (2008). Life skills development through sport: Current status and
808 future directions. *International Review of Sport and Exercise Psychology, 1*, 58-78.

809 Hellison, D. (1995). *Teaching responsibility through physical activity.* Champaign, IL: Human
810 Kinetics.
811 Hermens, N., Super, S., Verkooijen, K., & Koelen, M. (2017). A systematic review of life kill
812 development through sports programs serving socially vulnerable youth. *Research*
813 *Quarterly for Exercise & Sport*, *88*, 408-424.
814 Higgins, J., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions,*
815 Version 5.1.0. The Cochrane Collaboration, available from: www.cochrane-handbook.org
816 Hodge, K., Danish, S. J., & Martin, J. (2012). Developing a conceptual framework for life
817 skills interventions. *Counselling Psychologist*, *41*, 1-28.
818 Hodge, C., Kanters, M., Forneris, T., Bocarro, J., & Sayre-McCord, R. (2017). A family thing:
819 Positive youth development outcomes of a sport-based life skills program. *Journal of Park*
820 *and Recreation*, *35*, 34-50.
821 Holt, N. L., McHugh, T., Tink, L., Kingsley, B., Coppola, A., Neely, K., & McDonald, R.
822 (2013). Developing sport-based after-school programmes using a participatory action
823 research approach. *Qualitative Research in Sport, Exercise & Health*, *5*, 332-355.
824 Holt, N. L., Neely, K., Slater, L., Camiré, M., Côté, J., & Fraser-Thomas, J. (2017). A
825 grounded theory of positive youth development through sport based on results from a
826 qualitative meta-study. *International Review of Sport & Exercise Psychology*, *10*, 1-49.
827 Huysmans, Z., Clement, D., Whitley, M., Gonzalez, M., & Sheehy, T. (2019). "Putting kids
828 first": An exploration of the teaching personal and social responsibility model to youth
829 development in Eswatini. *Journal of Sport for Development, 7*, 15-32.
830 Jackson, N., & Waters, E. (2005). Criteria for the systematic review of health promotion and
831 public health interventions. *Health Promotion International*, *20*, 367-374.
832 Jacobs, J., & Wright, P. (2019, online). Thinking about the transfer of life skills: Reflections
833 from youth in a community-based sport programme in an underserved urban setting.
834 *International Journal of Sport & Exercise Psychology*.

Jones, M. I., & Lavelle, D. (2009). Exploring the life skills needs of British adolescent athletes. *Psychology of Sport and Exercise, 10*, 159-167.

Jones, G., Edwards, M., Bocarro, J., Bunds, W., & Smith, J. (2017). An integrative review of sport-based youth development literature. *Sport in Society*, *20*, 161-179.

Kendellen, K., & Camiré, M. (2017) Examining the life skill development and transfer experiences of former high school athletes. *International Journal of Sport and Exercise Psychology*, *15*, 395-408.

Kmet, L., Lee, R., & Cook, L. (2004). Standard quality assessment criteria for evaluating primary research papers from a variety of fields. *HTA Initiative, 13*, 1-20.

Lee, O., Park, M., Jang, K., & Park, Y. (2017). Life lessons after classes: Investigating the influence of an afterschool sport on adolescents' life skills development. *International Journal of Qualitative Studies on Health and Well-Being*, *12*, 1-10.

Mahoney, J. (2000). School extracurricular activity participation as a moderator in the development of antisocial patterns. *Child Development*, *71*, 502-516.

Mahoney, J., Eccles, J., & Larson, R. (2004). Processes of adjustment in organized out of school activities: Opportunities and risks. *New Directions for Youth Development*, *101*, 115–144.

McBride, N. (2016). *Intervention research: A practical guide for developing evidence-based school prevention programmes.* Singapore: Springer.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D., & Group, T. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annual of Internal Medicine*, *151*, 264-269.

Mullen, P., Green, L., & Persinger, G. (1985). Clinical trials of patient education for chronic conditions: A comparative meta-analysis of intervention types. *Preventative Medicine, 14*, 753-781.

Pace, R., Pluye, P., Bartlett, G, Macaulay, A., Salsberg, J., Jagosh, J., & Seller, R. (2012). Testing the reliability and efficiency of the pilot mixed methods appraisal tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies, 49,* 47-53.

Papacharisis, V., Goudas, M., Danish, S. J., & Theodorakis, Y. (2005). The effectiveness of teaching a life skills program in a sport context. *Journal of Applied Sport Psychology, 17*, 247-254.

Pawson, R., Boaz, A., Grayson, L., Long, A., & Barnes, C. (2003). *Types and quality of knowledge in social care knowledge review 3,* SCIE: London UK.

Petitpas, A., Cornelius, A., Van Raalte, J., & Jones, T. (2005). A framework for planning youth sport programs that foster psychosocial development. *The Sport Psychologist, 19*, 63-80.

Pierce, S., Gould, D., & Camiré, M. (2016). Definition and model of life skills transfer. *International Review of Sport and Exercise Psychology, 10*, 186-211.

Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathian, A., Griffiths, F., Boardman, F., Gagnon, M., & Rousseau, M. (2011). *Proposal: A mixed methods tool for systematic mixed studies reviews.* Available online at: http://mixedmethodsappraisaltoolpublic.pbworks.com (Accessed August 2017)

Prestwich, A., Webb, T., & Conner, M. (2015). Using theory to develop and test interventions to promote changes in health behavior: Evidence, issues, and recommendations. *Current Opinion in Psychology, 5*, 1-5.

Rigoni, P., Belem, I., & Viera, L. (2017). Systematic review on the impact of sport on the positive youth development of high performance athletes. *Journal of Physical Education, 28*, 1-13.

Robertson, E., Wakefield, C., Signorelli, C., Cohn, R., Patenaude, A., Foster, C., Pettit, T., & Fardell, J. (2018). Strategies to facilitate shared decision-making about pediatric oncology clinical trial enrolment: A systematic review. *Patient Education and Counseling, 101*, 1157-1174.

Schinke, R., Mellalieu, S. D., Ntoumanis, N., Kavussanu, M., Standage, M., Strauss, B., & Papaioannou, A. (2020, online). Getting published: Suggestions and strategies from editors of sport and exercise psychology journals. *Journal of Applied Sport Psychology*.

Schulz, K., Altman, D., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal*, *340*, 697-702.

Turnnidge, J., Côté, J., & Hancock, D. (2014). Positive youth development form sport to life: Explicit or implicit transfer? *QUEST*, *66*, 203-217.

Waldron, J. (2009). Development of life skills and involvement in the girls on track program. *Women in Sport and Physical Activity*, *18*, 60-73.

Weiss, M., Bolter, N., & Kipp, L. (2016). Evaluation of the first tee in promoting positive youth development: Group comparisons and longitudinal trends. *Research Quarterly for Exercise and Sport*, *87*, 271-283.

Weiss, M., Stuntz, C., Bhalla, J., Bolter, N., & Price, M. (2013). More than a game: Impact of the first tee life skills program on positive youth development: Project introduction and year 1 findings. *Qualitative Research in Sport, Exercise and Health*, *5*, 214-24.

Whitley, M., Massey, W., Camiré, M., Boutet, M., & Borbee, A. (2019a). Sport-based youth development interventions in the united states: A systematic review. *BMC Public Health*, 19, 1-20.

Whitley, M., Massey, W., Camiré, M., Blom, L.C., Chawansky, M., Forde, S. … Darnell, S. (2019b). A systematic review of sport for development interventions across six global cities. *Sport Management Review*, 22, 181-193.

Wholey, J., Hatry, H., & Newcomer, K. (2015). *Handbook of practical program evaluation*. Jossey-Bass: San Francisco

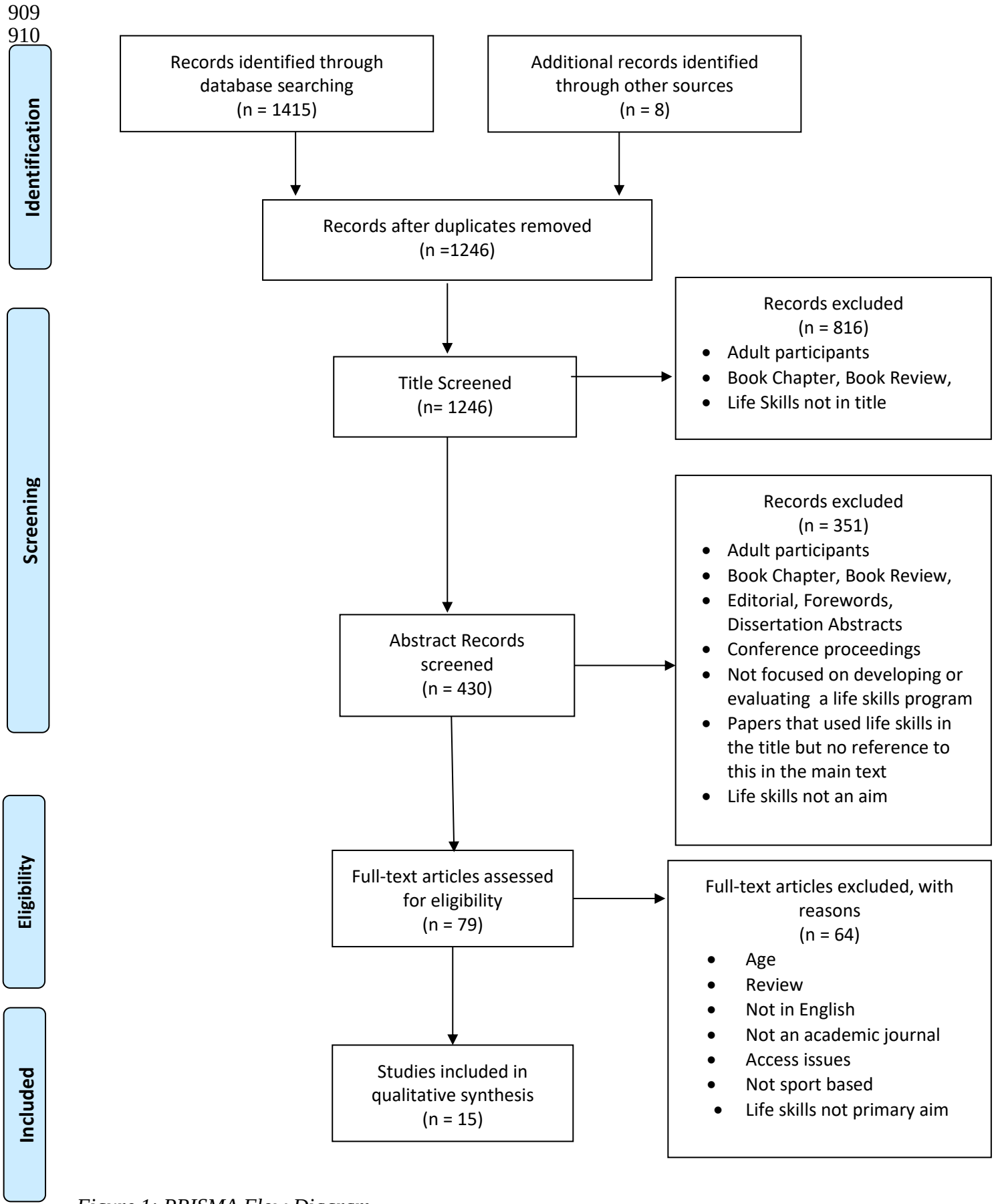

Figure 1: PRISMA Flow Diagram

Table 1: Intervention Design Quality Assessment Scores

Name	THEO	INTFOC	DET	DUR	INDIV	ONFEED	PILOT	FID	Overall Score	%
Bean et al., (2015)	1	2	1	2	1	0	0	1	8	50%
Bean et al. (2016)	1	2	1	2	1	2	0	0	9	56%
Brunelle et al. (2007)	0	2	1	0	0	0	0	0	3	19%
Goudas et al. (2006)	0	2	1	0	1	0	0	0	4	25%
Goudas & Giannoudis, (2008)	0	2	1	2	1	1	0	0	7	44%
Goudas & Giannoudis, (2010)	0	2	1	2	1	1	0	0	6	38%
Hodge et al. (2017)	0	2	1	2	1	1	0	0	7	44%
Holt et al. (2013)	0	1	1	2	1	1	0	0	7	44%
Huysmans et al. (2019)	2	2	1	2	2	1	0	1	11	69%
Jacobs & Wright (2019)	0	2	1	1	1	2	0	1	8	50%
Lee et al. (2017)	0	2	0	2	0	0	0	1	5	31%
Papacharisis et al. (2005)	0	2	0	0	1	0	0	0	3	19%
Waldron, (2009)	0	2	1	0	0	0	0	0	3	19%
Weiss et al. (2013)	1	2	0	0	0	1	0	1	5	31%
Weiss et al. (2016)	1	2	0	0	0	0	0	0	3	19%

THEO = *theoretically informed*; INTFOC = *intervention focus*; DET = *detail described*; DUR = *duration justified*; INDIV = *individualization*; ONFEED = *ongoing feedback*; PILOT = *pilot implementation*; FID = *fidelity*.

Table 2: Mixed Methods Appraisal Tool (MMAT; Pluye et al., 2011)

	Qualitative				Quantitative Non-Randomized				Quantitative Descriptive				Mixed Methods			Results	Combined %
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3		
Bean et al. (2015)	✓	✓	×	×					✓	×	✓	✓	✓	✓	×	50%	50%
Bean et al. (2016)	✓	✓	✓	×												75%	65.5%
Brunelle et al. (2007)									×	×	✓	✓				50%	34.5%
Goudas et al. (2006)					✓	✓	×	×								50%	37.5%
Goudas & Giannoudis, (2008)					✓	✓	×	×								50%	47%
Goudas & Giannoudis, (2010)	✓	✓	✓	×												75%	56.5%
Hodge et al. (2017)	×	✓	×	×												25%	34.5%
Holt et al. (2013)	✓	✓	✓	×												75%	56.5%
Huysmans et al. (2019)	✓	✓	✓	✓					×	✓	✓	×	×	×	×	50%	62.5%
Jacobs & Wright (2019)	✓	✓	✓	✓												100%	75%
Lee et al. (2017)	×	✓	×	×												25%	28%
Papacharisis et al. (2005)					✓	✓	-	-								50%	34.5%
Waldron, (2009)	✓	✓	-	×												50%	34.5%
Weiss et al. (2013)	✓	✓	×	×												50%	40.5%
Weiss et al. (2016)					✓	✓	×	-								50%	34.5%

Table 3: Characteristics of Design

	Theory	Intentional Focus			Program Description						Dur.	Indiv.	Feedback	Pilot	Program Fidelity	
		CPG	CSD	LSE	Prog.	Loc.	Cont.	Struct.	Life skills						Delivery	Train.
Bean et al., (2015)	PYD	Y	N	Y	None	Canada	Com.	30 Session 75mins	Communication Confidence Respect Goal Setting Seeking Help Relaxation	Teamwork Self-Talk Focus Leadership Planning Responsibility	8 months	Y	N	N	Program Staff	Not Stated
Bean et al., (2016)	TPSR	Y	Y	Y	SUPER	Canada	Com.	57 sessions Once per week 75mins & 90mins	Communication Goal Setting Leadership Seeking Help	Teamwork Relaxation Planning Self-Talk	2 years	Y	Debrief	N	Researcher as leader	Not Stated
Brunelle et al., (2007)	None	Y	Y	Y	SUPER	USA	Com.	5 Sessions 45mins	Goal Setting		One week	N	N	N	Life skills Staff	Not Stated
Goudas et al., (2006)	None	N	Y	Y	SUPER	Greece	Edu.	8 Sessions Twice per week 10-15mins	Goal Setting	Positive Thinking	4 weeks	Y	N	N	Research Assistant	Not Stated
Goudas & Giannoudis, (2008)	None	N	Y	Y	SUPER	Greece	Edu.	17 Sessions	Goal Setting Problem Solving	Positive Thinking	Not Stated	Y	N	N	PE Teacher	Not Stated
Goudas & Giannoudis, (2010)	None	N	Y	Y	SUPER	Greece	Edu.	17 Sessions 3 times per week	Goal Setting Problem Solving	Positive Thinking	2 months	Y	Parent reinforce Homework	N	PE Teacher	Not Stated
Hodge et al., (2017)	None	Y	Y	Y	SUPER	USA	Com.	Once per week 15-25mins	Communication Emotional Man.	Teamwork Goal Setting	11 weeks	Y	Coach reinforce	N	Peer Students	Not Stated
Holt et al., 2013	None	Y	N	Y	TRY-Sport	Canada	Sch.	Twice per week	Teamwork Confidence	Leadership	3 months	Y	Debrief Reflection	Y	Researcher Fieldworker	Not Stated
Huysmans et al., (2019)	TPSR	Y	Y	Y		Eswatini	Com.	Every weekday 75-100mins	Social skills Emotional reg. Teamwork Leadership Relaxation	Self-Talk Decision making Communication Goal Setting Multiple Values	3 weeks	Y	Reflection	N	Local Coaches	3 days Train-the-trainer
Jacobs & Wright (2019)	None	Y	N	Y	None	USA	Com.	Not stated	Perseverance Leadership Responsibility	Community Respect Teamwork	Season bi-weekly	Y	Debrief Coach reinforce	N	Program Coaches	Annual training
Lee et al., (2017)	None	Y	N	Y	None	Korea	Sport	Once per week	Communication Goal Setting	Teamwork	12 weeks	N	N	N	Coaches	3 sessions
Papacharisis et al., (2005)	None	Y	N	Y	SUPER	Greece	Sport	Once per week 15mins	Goal Setting Problem Solving	Positive Thinking Self-Regulation	8 weeks	Y	N	N	Researcher & Coach	Not Stated
Waldron, (2009)	None	Y	N	Y	None	USA	Com.	4 weeks 3 sessions	Problem Solving Health	Reflection Media	4 weeks	N	N	N	Volunteer Coaches	Not Stated
Weiss et al., (2013)	PYD	Y	N	Y	First Tee	USA	Sport	Not stated	Self-Mgmt. Social Support	Health Goal Setting	Not stated	N	Coach reinforce	N	Trained Coaches	2-day workshop
Weiss et al., (2016)	PYD	Y	N	Y	First Tee	USA	Sport	Not stated	Self-Mgmt. Social support	Health Goal Setting	Not stated	N	N	N	Trained Coaches	Not Stated

Transf. Exp. = *Transformative* experience; CPG = *Clear program goals*; CSD = *Clear session descriptions*; LSE = *Life skills embedded*; Prog. = *Program*; Loc. = *Location*; Cont. = *Context*; (Com. = *community*; Edu. = *education*; Sch. = *school*); Struct. = *Structure*; Dur. = *Duration*; Indiv. = *Individualization*; Feedback = *Ongoing feedback*; Train. = *Training for those delivering intervention*.

Table 4: Characteristics of Evaluation Methods

	Sampling		Data Analysis			Appropriate Measure			Comp. groups	Outcome
	Sample Size	**Program Participants**	**Domain**	**Method**	**Freq. of Eval.**	**Instrument**	**Valid Meas.**	**Self-Report**	**Cont. Group**	
Bean et al., (2015)	10 Youth 5 Leaders	Girls Avg. Age 11.75	Mixed Method	Interviews Questionnaire	Post	Semi-structured Interviews Youth Experience Survey (YES) 2.0	Y	Youth Leaders	No	Development
Bean et al. (2016)	8	Girls Age 11-16	Qualitative	Interview	During Post	Semi-Structured	N/A	Youth	No	Development & Transfer
Brunelle et al. (2007)	100	Boys & Girls Age 13–17	Quantitative	Questionnaire	Pre/Post Follow up	SPRS, IRI, SIS, Goal Knowledge, Goal Self Efficacy, Comm service	Y	Youth	No	Partial Development
Goudas & Giannoudis (2010)	86	Boys & Girls Age 11-14	Qualitative	Interviews, Field Notes Workbooks	Post	Semi-Structured	N/A	Youth	No	Development & Transfer
Goudas & Giannoudis (2008)	130	Age 12-14	Quantitative	Questionnaire	Pre/Post	Knowledge Test, Self-belief of ability & Sport Skills test	Y	Youth	Yes	Development
Goudas et al. (2006)	73	Boys & Girls Avg. age 12	Quantitative	Questionnaire	Pre/Post Follow up	Physical Fitness Test, Knowledge Test, Self- Belief of ability	Y	Youth	Yes	Partial Development
Hodge et al. (2017)	36	Boys & Girls Age 7–15	Qualitative	Interviews	During Follow up	Semi-Structured	N/A	Youth Parents	No	Development & Transfer
Holt et al. 2013	Study 1: 28 +2 Study 2: 14 +3	Boys & Girls Age 7-9	Qualitative	Interviews	Post	Semi-structured	N/A	Youth Teachers	No	Development & Transfer
Huysmans et al., 2019	33 youth 2 coaches 1 Teacher	Boys & Girls Age 11-15	Mixed Method	Focus Groups Questionnaire Interview	Pre/Post Follow up	Focus Groups, Interview PSRQ, MSPSE Learning quiz	Y	Youth Coaches Teachers	No	Development & Transfer
Jacobs & Wright (2019)	11	Boys & Girls Age 12-18	Qualitative	Focus Group Interviews Observations Field journal	During Post	N/S	N/A	Youth	No	Development & Transfer
Lee et al. (2017)	6 2 instructors	Boys & Girls Avg. age 11.5	Qualitative	Interviews	Pre Post	N/S	N/A	Youth Facilitator	No	Development & Transfer
Papacharisis et al. (2005)	72	Boys & Girls Age 10-12	Quantitative	Questionnaire	Pre Post	Knowledge Test, Self-belief of ability & Sport Skills test	Y	Youth	Yes	Development
Waldron (2009)	19	Girls Age 11-13	Qualitative	Interviews	Post	Semi-Structured	N/A	Youth	Yes	Development & Transfer
Weiss et al. (2013)	95 youth 26 coaches 24 parents	Boys & Girls Age 11- 17	Qualitative	Interviews Focus Groups	Post	N.S	N/A	Youth Parents Coaches	No	Development & Transfer
Weiss et al. (2016)	Study 1: 564 Study 2: 192	Boys & Girls Age 10-17	Quantitative	Questionnaire	During Post Follow up	LSTS, SPP, SPPa, Character Dev. scale, SSE; Preference for challenge scale	Y	Youth	Yes	Transfer

SPRS = *social personal responsibility scale;* IRI = *Interpersonal Reactivity Index;* SIS= *Social Interest Scale;* PSRQ = *Personal and Social Responsibility Questionnaire*; MSPSE = *Multidimensional Scales of Perceived Self-Efficacy*; LSTS = *Life skills Transfer Scale;* SPP = *Self Perception Profile;* SPPa = *Self Perception Profile for Adolescent; SSE = Scale for Self-Efficacy.*