Development of an observational measure assessing program quality processes in youth sport

Corliss Bean1,2*, Sara Kramers3 Martin Camiré4 Jessica Fraser-Thomas4 and Tanya Forneris1

Abstract: Research has demonstrated that quality sport programs have the potential to foster the physical and psychosocial development of youth. However, there is an absence of observational measures to assess program quality related to psychosocial development within youth sport. The purpose of this paper is to report on two studies conducted to develop a valid and reliable observational measure to assess program quality processes in youth sport. Study one outlines the process of attaining content and face validity using an expert panel approach when developing the Program Quality Assessment in Youth Sport (PQAYS) observational measure through a review of literature and collaboration with expert academics and coaches. Study two outlines further steps taken to test the internal reliability, as well as convergent and predictive validity of the measure. Results from the two studies provide initial evidence that the PQAYS is a valid and reliable measure that can be used in intervention and evaluation research within youth sport.

Subjects: Sport and Leisure Studies; Sports Development; Sports Coaching; Research Methods

Keywords: positive youth development; measurement; psychometrics; coaches; program quality

Youth sport is a context with much potential for fostering positive youth development (PYD; Côté & Fraser-Thomas, 2011; Eccles & Gootman, 2002; Holt, 2016). The guiding principle underlying the PYD framework is the shift from a deficit-reduction paradigm to a proactive, asset-building...
paradigm that sees youth as resources to be developed, rather than problems to be managed (Damon, 2004). This strength-based approach has gained popularity over the past three decades and enforces the importance of teaching youth important life skills (Botvin, 2004). Given that approximately 65% of children between 6 and 17 years of age across Canada and the United States are involved in organized sport programs (Guèvremont, Findlay, & Kohen, 2008; United States Census Bureau, 2014), it is critical that we further our understanding of how PYD can be fostered within this context. For the present study, sport was defined as a social and competitive activity requiring specific physical skills and physical exertion, which occurs within an institutionalized setting (Coakley & Donnelly, 2009).

In past research, participation in youth sport has been shown to lead to both positive and negative outcomes (e.g., Merkel, 2013). On one hand, youth sport participation has been associated with improved physical and psychosocial development (e.g., Eime, Young, Harvey, Charity, & Payne, 2013; Fraser-Thomas, Côté, & Deakin, 2005). On the other hand, participation in sport has been associated with physical and psychosocial problems (injury, depression), as well as economic and cultural concerns (financial burden, ethnicity and gender inequality; e.g., Merkel, 2013). To best understand the outcomes emanating from youth sport, there is a need to examine the mechanisms related to psychosocial development (e.g., Gould & Carson, 2008; Hodge, Danish, Forneris, & Miles, 2016).

0.1. Program quality within youth sport

Program quality is a multi-faceted concept; thus, a universal definition of quality does not exist. Different definitions are necessary to deal with specific concepts under different circumstances (Reeves & Bednar, 1994). Quality within youth programming is dynamic (Larson & Walker, 2010), yet for the purpose of this study refers to the structures and processes within a program that relate to youth outcomes (Baldwin & Wilder, 2014). Specifically, program structures refer to an organization’s capacity to deliver a program to youth (e.g., physical space, staffing, funding, community collaborations). Program processes refer to how the program is delivered (e.g., supportive relationships, opportunities for skill-building, autonomy). Researchers contend that how youth programs, including sport, are structured plays a key role in determining whether positive or negative outcomes occur (Côté & Fraser-Thomas, 2011; Petitpas, Cornelius, Van Raalte, & Jones, 2005; Roth & Brooks-Gunn, 2015). As a result, program quality has been outlined as one of the best predictors of the developmental outcomes resulting from participation in youth programs (e.g., Durlak, Mahoney, Bohnert, & Parente, 2010; Roth & Brooks-Gunn, 2015; Yohalem & Wilson-Ahlstrom, 2010). Although there are many factors that have been identified as contributors to the quality of a program, to date, these factors have yet to be extensively examined in the youth sport context.

One of the most acknowledged and comprehensive classifications of program quality was proposed by Eccles and Gootman (2002) who worked with the National Research Council and Institute of Medicine (NRCIM) to summarize two decades of developmental psychology research. These authors proposed eight program features linked with positive psychosocial development: (a) physical and psychological safety; (b) appropriate structure; (c) supportive relationships; (d) opportunities to belong; (e) positive social norms; (f) support for efficacy and mattering; (g) opportunities for skill building; and (h) integration of family, school, and community efforts. Since their development, these eight setting features have been utilized to guide youth programming at the research and practical levels (HighScope Educational Research Foundation [HSERF], 2005; Yohalem, Wilson-Ahlstrom, Fischer, & Shinn, 2009). Although the usefulness of these features has been recognized by youth sport researchers (e.g., Côté, Strachan, & Fraser-Thomas, 2008; Povilaitis & Tamminen, 2017; Strachan, Côté, & Deakin, 2011), little empirical research has been conducted within sport utilizing these features, despite calls to do so (e.g., Côté & Mallett, 2013). One reason may be due, in part, to the lack of available measures to assess quality within the youth sport context that incorporate the NRCIM’s eight setting features.
0.2. Measuring program quality processes

Program quality can be assessed in several ways, including the use of qualitative methods, quantitative self-report measures, and observational measures. In youth sport research, to date, most studies have relied on self-report measures, with observational research being neglected (Jones, 2015). Researchers have argued for the need to integrate observational measures of program quality as observational assessment allows for the description of behavior in natural environments, leading to greater ecological validity. It can also provide more objective evidence of the behaviors and strategies coaches are using in their programs, rather than relying solely on coaches’ perceptions (Flett, Gould, & Lauer, 2012; Holt & Jones, 2008; Jones, 2015).

There are few observational measures that have been developed to assess quality within youth programming (for full details, see Yohalem et al., 2009), and none that have been developed specifically for the youth sport context. The Out-of-School Time Observation Instrument (Pechman, Russell, & Birmingham, 2008) and the Youth Program Quality Assessment (YPQA; HSERF, 2005) have both been designed for the use in youth programming. The Out-of-School Time Observation Instrument is based on Durlak and Weissberg’s (2007) SAFE (Sequential, Active, Focused, Explicit) features and has not been used within academic literature. The YPQA has been used to assess program quality when conducting process evaluations within out-of-school and community programs (HSERF, 2005; Smith & Hohmann, 2005) and is loosely based on Eccles and Gootman’s (2002) eight program setting features. This measure was used as a starting point for the development of the PQAYS.

The YPQA represents a valid and reliable measure for youth aged 8–18 years (Smith & Hohmann, 2005) and can be used as a self, internal, and external evaluation tool. This measure has four domains: (a) safe environment, (b) supportive environment, (c) interaction, and (d) engagement and is measured on a 3-point Likert scale scored as a 1, 3, or 5 (HSERF, 2005). Given the breadth of contexts (e.g., leadership, arts, mentoring, sport) in which this measure can be used, the YPQA is not designed to decipher the contextual intricacies that define quality in various youth programming contexts.

To date, only a few studies have empirically assessed program quality in youth sport using the YPQA. For example, Flett et al. (2012) utilized the YPQA to assess quality within youth softball and baseball programs. Despite the YPQA’s reported strengths in reliability and validity, the authors outlined problematic issues with distribution and psychometric properties when using the YPQA in sport, particularly related to the low internal consistency of subscales. Thus, several revisions were conducted, reducing the measure from 52 items to 26 items in attempts to improve reliability. After shortening the measure for analyses, the authors found that the sport programs studied tended to yield high scores related to providing a safe and supportive environment, but lower scores related to providing opportunities for interaction and engagement.

In another study that also used the YPQA within sport and non-sport programs (Bean & Forneris, 2016a), a 5-point scale was used to aid in variability of program quality scores. Researchers conducted 184 observations across 33 youth programs and yielded similar results to Flett et al. (2012). Specifically, the highest scores were observed for providing a safe and supportive environment and lower scores for interaction and engagement opportunities. Issues related to low reliability persisted (Bean & Forneris, 2016a). In sum, because the YPQA did not include items important to assess in sport programs (e.g., developmental opportunities for sport/physical skills), the examples were not always relevant. Therefore, in its current form, the YPQA is not optimally suited to measure program quality within the sport context. Previous empirical research has supported the notion of sport being a unique context compared to other extra-curricular activities (e.g., Zarrett et al., 2008). Specifically, sport presents unique features that are not necessarily found in other types of youth programs (e.g., opportunities to develop both physical and psychosocial skills, inherent competitiveness; Danish, Forneris, Hodge, & Heke, 2004; Fraser-Thomas et al., 2005; Pierce, Gould, & Camiré, 2017).
Recently, MacDonald and McIssac (2016) discussed how a missing element in the sport psychology literature is an understanding of the processes through which PYD occurs in sport, but that no measures are currently available to assess such processes. Systematic observation is essential as a procedural step in understanding the mechanisms of psychosocial development within sport (Brewer & Jones, 2002). Given its worth for research and evaluation, but also program improvement, funding securement, and the retention of participants, a measure of program quality is needed for initiatives to be examined in the context of everyday practice.

0.3. The present paper

Based on the existing literature, it appears that the field of sport psychology stands to benefit from the development of an observational measure to assess program quality within youth sport for three reasons. First, observational data have been under-utilized within evaluation research and there has been an over-reliance on self-report measures (Flett et al., 2012). Second, quantitative observational measures are needed to understand the processes in which PYD occurs within youth sport (MacDonald & McIssac, 2016). Third, given that sport is the most popular extra-curricular activity across North America (Guèvremont et al., 2008; United States Census Bureau, 2014), there is a need to develop an observational measure specifically for this context. Therefore, the purpose of this paper is to report on two studies conducted to develop a valid and reliable observational measure to assess program quality processes in youth sport. The use of an observational measure will help address several aforementioned limitations within the current literature. The term program is used as this is the term most widely utilized within the positive youth development and program quality literature. However, it should be recognized that this term, in sport, also refers to “sport team” or “sport club.” The first study summarizes the process of developing the PQAYS and presents an overview of the final measure. The second study outlines the steps taken to assess the reliability and validity of the measure. The procedures utilized in both studies followed the development and validation processes used in other observational measurement development related to coach development in the sport literature (e.g., Allan, Turnnidge, Vierimaa, Davis, & Côté, 2016; Brewer & Jones, 2002; Erickson & Côté, 2015).

1. Study one: measure development

The purpose of study one was to establish content and face validity for the observational measure. A series of steps were followed in its creation: (a) conducting a review of literature; (b) developing the initial measure, instructions, response format, and scoring; (c) involving expert academics and coaches to gather feedback on the measure; (d) piloting the measure; and (e) finalizing the measure.

2. Method

2.1. Step one: conducting a review of literature

The first step was to review the English-language sport psychology literature, with the goal of locating and reviewing empirical studies, meta-analyses, position papers, literature reviews, book chapters, and doctoral dissertations that touched on program quality and PYD in youth sport. This was done in October 2016. The following procedures were used to locate sources: (a) computer searches of 10 electronic databases (e.g., SPORTDiscus, SCOPUS, PsycINFO) using different combinations of the following keyword search terms: youth (child, adolescent), organized sport (sport participation, physical activity), program quality (setting features, characteristics), youth development (positive youth development); (b) full scan of the reference lists of all relevant articles; and (c) manual searches of key peer-reviewed journals to find any additional relevant articles that did not arise during the database searches. The literature search yielded 195 articles. When reviewing the literature, many youth sport researchers (e.g., Côté & Abernethy, 2012; Weiss, 2008) advocated for the use of Eccles and Gootman’s (2002) eight program features. Such features were utilized to frame the measure and the processes surrounding measurement development are outlined in step 2 below.
2.2. Step two: developing the initial measure, instructions, and scoring

2.2.1. Developing the initial measure

The literature review helped the researchers understand current best practices for program delivery within sport, which informed the development of each subscale. As noted, it was evident that the eight setting features (Eccles & Gootman, 2002) were the most acknowledged features of program quality in general youth programming, as well as sport-based programming. Thus, these features were used to ground the instrument’s initial organizing constructs, as there was sufficient evidence to support their worth in helping guide item development and the creation of additional subscales described below. Empirical and theoretical work from the youth programming and youth sport literatures, particularly from 2002 onwards, were also used to aid in measure development (see Table 1). For example, other measures, such as the YPQA (HSERF, 2005), were used to guide specific item development. One strength of both the eight setting features and the YPQA is that foundational elements of program quality (e.g., safe environment) are used to build upon to higher-order elements of program quality (e.g., supporting efficacy, providing opportunities for interaction and engagement). This strength was maintained in the development of the new measure. Further, although not specifically assessing program quality, existing observational measures used within sport, including the Coach-Athlete Interaction Coding System (Erickson, Côté, Hollenstein, & Deakin, 2011) and the Assessment of Coach Emotions (Allan et al., 2016) were reviewed to aid as a starting point in measure development. See Table 1 for a breakdown of each subscale, with specific references provided at the end of each item to demonstrate how it was informed by research. Throughout this initial item development process, Smith, Quested, Appleton, and Duda’s (2016) review of observational instruments within sport and physical education was consulted. Based on the review of literature, two subscales were added (in addition to the subscales based on the eight setting features) to the measure and all the items were adapted to be sport-specific. These changes are detailed below. Finally, to ensure this was a measure of program quality and not solely a measure of coach competence, the researchers sought to achieve a balance of items that assessed the program, the coaches, and the youth. For example, in Supportive Relationships, there are two items focusing on coach behaviour, two items focusing on youth behaviour, and one item on the activities occurring within the program.

The initial version of the PQAYS was comprised of 64 items across 10 subscales. More specifically, from the eight setting features, two features were further divided into two separate subscales. First, based on the literature review and previous recommendations (Bean & Forneris, 2016a), safety was broken into two subscales of (a) physical safety and (b) psychological safety (see 1.1 and 1.2 in Table 1) to help ensure internal consistency as each element measures distinct program characteristics. Second, past research outlines the importance of intentionally structuring the sport context to deliberately teach life skills in combination with sport-specific skills (Camiré, Trudel, & Forneris, 2012; Fraser-Thomas et al., 2005; Gould & Carson, 2008; Weiss, Stuntz, Bhalla, Bolter, & Price, 2013). Therefore, opportunities for skill-building was divided into two subscales to measure opportunities for (a) sport and physical skill-building and (b) life skill-building (see 7.1 and 7.2 in Table 1).

Individual items for the PQAYS were developed one at a time, with at least one academic reference associated to each item. For example, the fourth item within the seventh subscale, Opportunities for Skill-Building -Life Skills (7.2), outlined that “Coach(es) debrief how life skills can be applied and transferred outside of a specific sport context” is supported theoretically and empirically in the literature (e.g., Allen, Rhind, & Koshy, 2015; Pierce et al., 2017; Turnnidge, Côté, & Hancock, 2014; Weiss et al., 2013). This process ensured that all items were thoroughly developed based on research. Moreover, sport-specific examples and explanations are provided within each item to contextualize the measure specifically to sport.

2.2.2. Instructions

The following section outlines the instructions of the measure. The measure commences with an introduction page, explaining the purpose of the measure. Comprehensive instructions were
Table 1. Program setting features proposed to foster youth development and supporting literature and number of items for subscales

| Subscale                                           | # of items | Description                                                                                     | Sample References from the Literature to Support each Subscale |
|----------------------------------------------------|------------|----------------------------------------------------------------------------------------------|---------------------------------------------------------------|
| 1.1 Physical Safety                                | 8          | An environment that allows youth to feel free from physical harm                               | Camiré et al. (2012); Eccles and Gootman (2002); HSERF (2005); Yohalem and Wilson-Ahlstrom (2009) |
| 1.2 Psychological Safety                           | 3          | An environment that allows youth to feel free from psychological harm; accepted, and respected | Camiré et al. (2012); Côté et al. (2008); Duda (2013); Eccles and Gootman (2002); HSERF (2005); Larson, Walker, Rusk, and Diaz (2015); Petitpas et al. (2005); Strachan et al. (2011); Yohalem and Wilson-Ahlstrom (2009) |
| 2.0 Appropriate Structure                          | 7          | Clear and consistent rules and expectations, including behavioral guidelines and age-appropriate monitoring | Camiré et al. (2012); Côté et al. (2008); Dryfoos (1990); Eccles and Gootman (2002); HSERF (2005); McLaughlin (2000); Strachan et al. (2011) |
| 3.0 Supportive Relationships                       | 5          | The presence of adults and peers who demonstrate concern and support for youth                | Beauchamp, Barling, and Morton (2011); Catalano, Berglund, Ryan, Lonczak, and Hawkins (2004); Côté et al. (2008); Eccles and Gootman (2002); Gould and Carson (2008); HSERF (2005); Merry (2000); Yohalem and Wilson-Ahlstrom (2009) |
| 4.0 Opportunities to Belong                        | 3          | Providing experiences that allow youth to develop a sense of belonging; feelings of value as an individual and part of a group | Côté et al. (2008); Eccles and Gootman (2002); HSERF (2005); Merry (2000); Yohalem and Wilson-Ahlstrom (2009) |
| 5.0 Positive Social Norms                          | 3          | Fostering clear, healthy, ethical standards, beliefs, and behavioral guidelines that promote prosocial behavior and minimize health risks | Camiré et al. (2012); Eccles and Gootman (2002); Larson et al. (2015); Yohalem and Wilson-Ahlstrom (2009) |
| 6.0 Support for Efficacy and Mattering             | 8          | Providing opportunities for youth to develop leadership, efficacy, autonomy, mattering, and responsibility | Camiré, Forneris, Trudel, and Bernard (2011); Catalano et al. (2004); Côté and Hancock (2014); Côté et al. (2008); Dryfoos (1990); Duda (1989, 2013); Eccles and Gootman (2002); McCarthy, Jones, and Clark-Carter (2008); Petitpas et al. (2005); Standage and Vallerand (2014); Stein, Bloom, and Sabiston (2012); Strachan et al. (2011); Yohalem and Wilson-Ahlstrom (2009) |
| 7.1 Opportunities for Skill-building—Sport and Physical Skills | 5          | Opportunities for youth to develop sport and physical skills that will prepare them for the future | Camiré et al. (2012); Côté et al. (2008); Danish, Petitpas, and Hale (1993); HSERF (2005); McLaughlin (2000); Papacharissi, Goudas, Danish, and Theodorakis (2005) |
| 7.2 Opportunities for Skill-building—Life Skills    | 4          | Opportunities for youth to develop intellectual, psychological, emotional, and social skills that will prepare them for the future | Allen et al. (2015); Bean and Forneris (2016a); Danish et al., (1993); Eccles and Gootman (2002); Gould and Carson (2008); Jones and Lavaliee (2009); Kendellen, Camiré, Bean, Forneris, and Thompson (2017); McLaughlin (2000); Papacharissi et al. (2005); Petitpas et al. (2005) |
| 8.0 Integration of Family, School, and Community Efforts | 5          | Incorporating family, school, and community to increase opportunities for synergy and positive relationships | Camiré et al. (2011); Eccles and Gootman (2002); HSERF (2005); Steinberg (2000); Yohalem and Wilson-Ahlstrom (2009) |

Note. Descriptions of setting features adapted from Community Programs to Promote Youth Development, by the National Research Council and Institute of Medicine, 2002, Washington, DC: National Academy Press.
developed to provide information regarding what is to be done before, during, and after a program observation session. For example, prior to conducting observations, an interview is to be conducted with the coach(es) to acquire an in-depth understanding of their philosophy and team objectives (see Appendix A for sample interview guide questions). In addition, it is necessary for the coach(es) to complete a Program Demographic Form before the observations begin to collect additional information (e.g., frequency and duration of sessions, types of sessions, parental involvement; see Appendix A for Program Demographic Form). Having the coach(es) complete this form is specifically designed to help the observer with scoring the eighth setting feature—integration of family, school, and community efforts. The interview, Program Demographic Form, and observations should all be used to inform the scoring of the items in the measure. Both the pre-interview with coach(es) and the Program Demographic Form can help provide an understanding of contextual features of the youth sport environment prior to observation, as research emphasizes the importance of understanding a sport program's context (Strachan et al., 2011).

A minimum of two observers are to be present at each program session to allow for the assessment of inter-rater reliability. This has been extensively supported within the literature on observational research (e.g., Brewer & Jones, 2002; Hallgren, 2012). Instructions on how to accurately score the program are outlined for observers (see Appendix A). Moreover, for credible conclusions on program quality to be drawn, a minimum of three observation sessions are required to occur throughout the duration of the program. This recommendation has been supported in other naturalistic observation research (Smith & Hohmann, 2005). During observation sessions, the observers should take field notes to use as supporting evidence when conducting the objective scoring of the items. Taking field notes is a common method of documenting observations (Patton, 2002). Immediately after an observation session, observers complete the PQAYS (i.e., not during the session). Excerpts from field notes are to be included within the comment section of each subscale to provide justification of scoring.

Lastly, a second interview with the coach(es) is to be conducted at program end to follow-up on the observations conducted. The purpose of this interview is to further understand elements of the program that were observed during observations, which may clarify some aspects related to the program's quality. Such an approach helps to increase the quality of the interpretations that can be made by the researchers (Tracy, 2010).

The process outlined above is considered the optimal procedure that researchers should strive to follow to assess quality within a youth sport program. However, from a practical standpoint, we acknowledge that not all components may be feasible within a given context or situation (e.g., some coaches may not agree to be interviewed). In the present study, the measure was validated using this optimal procedure, which allows for the use of multiple methods and sources to create a comprehensive account (Yin, 2009).

2.2.3. Scoring
The measure uses a 5-point Likert scale ranging from 1 (never) to 5 (very often). This scale was chosen to improve issues of minimal variability experienced in the original YPQA, where items were scored using a 3-point scale. A 5-point scoring system is commonly used in youth programming (e.g., Search Institute, 2015) and observational measures (e.g., Nakaha, Grimes, Nadler, & Roberts, 2016). For some items, there is also a “Not Applicable (N/A)” option, which is combined with a footnote that provides justification as to when and why scoring an item may not be applicable. If an item is deemed not applicable, no score is given, and the item is not calculated within the subscale’s mean score. For example, if program quality is assessed in an individual sport environment, items for peer interactions may not be applicable.

The final scoring of the PQAYS is calculated by computing averages for each of the 10 subscales. Specifically, the subscale of Supportive Relationships has 4 items; therefore, the total scores of
these items would be summed and divided by 4 to attain a mean score for the subscale. A total score of program quality is calculated by computing a mean score of the 10 subscales' means. This is done so as not to weight certain subscales as more important than others based on the number of items within a given subscale.

2.3. Step three: involving academics and coaches to gather feedback

Schutz and Park (2004) outlined that content validity is supported if individuals who are knowledgeable about the intended construct agree that items reasonably represent the construct and are assigned to the appropriate category. After the first two steps were completed, an expert panel approach was used, which involved identifying knowledgeable individuals to provide feedback on the measure (Zamanzadeh et al., 2014). Based on previous recommendations, the experts emanated from a homogenous population within the same discipline (Clayton, 1997). For this study, two expert panels were created and involved in reviewing the measure: 19 academics (six faculty members and 13 graduate students) considered experts in youth development through sport and 34 youth sport coaches, considered applied experts in the field.

2.3.1. Academic experts

A three-phased review process occurred, whereby researchers and practitioners were involved based on their expertise in the field of youth development through sport. In the first phase, researchers were selected in the review process, in which five research team meetings were conducted between November 2015 and January 2016. These meetings involved six graduate students and two faculty members. The meetings included discussions around content and formatting and provided opportunities to discuss any issues with the measure. Based on the discussions, modifications were made to the PQAYS, which included moving items from one subscale to another and removing overlapping items. In the second phase, the revised version of the measure was emailed to a group of experts at another academic institution (one faculty member and six graduate students). These individuals were asked to review the measure for: (a) appropriateness and clarity of the instructions, (b) potential overlapping of items and concepts across subscales, (c) appropriateness of the examples provided within each item, (d) order of items within each subscale, and (e) additional comments that would aid in refining the measure. The exact instructions sent to this group of experts are available upon request. The experts were given 1 week to complete their review. Once all of their feedback was gathered, a meeting was held between members of the research team and the group of experts to debrief the measure. Once the phase two feedback was integrated, the measure was sent out for review to a third expert panel. Eight additional academics were contacted via email and asked to follow the same instructions provided to the group of experts in phase two. Of these eight, four individuals (three faculty members and one graduate student) provided written feedback on the measure.

2.3.2. Youth sport coaches

Face validity was further established through an online questionnaire hosted on FluidSurveys. Youth sport coaches (N = 34) completed a questionnaire to test the relevance of items within the measure. Coaches who had agreed to participate in the observational component of study two were asked to also complete the online questionnaire. Additionally, a convenience sample was used where the first and second authors sent out the questionnaire link to some of their coach contacts. Snowball sampling was also used, where coaches who agreed to complete the questionnaire were asked to pass on the questionnaire link to additional coaches. In total, 121 coaches completed a portion of the questionnaire; however, despite informing participants about anticipated questionnaire length and having a progress bar on screen, only 34 participants completed the questionnaire in its entirety. Thus, these individuals were included in this portion of the face validity assessment.

2.3.2.1. Coach survey and data collection. Of the 34 coaches who completed the questionnaire, 16 were male, 17 were female, and one did not disclose their gender (M_{age} = 32.01, SD = 12.05). In all,
19 individuals held a bachelor’s degree; nine held a high school diploma, three held a Master's degree, and three individuals held a college diploma or a professional degree. Years of coaching experience ranged from 1 to 40 ($M = 8.08$, $SD = 7.51$). At the beginning of the questionnaire, coaches were provided with a definition of each of the 10 subscales proposed within the PQAYS. For each item, participants were asked three questions related to: (a) the element of program quality they believe the item corresponded to from the 10 subscales (participants were given the option of “more than one subscale” and “none of the above”); (b) the relevance of the item to their current sport practice; and (c) if they believed the item was clearly worded. The second and third questions were measured on a 10-point Likert scale from 1 (not relevant at all/not clear at all) to 10 (very relevant/very clear). Participants could also provide open-ended comments for each item.

2.4. Integrating expert feedback

After all the expert feedback was gathered, the research team held additional meetings to review the feedback and revise the measure. Feedback from academics and coaches resulted in measure improvements by enhancing clarity surrounding the instructions and procedures, ensuring the congruency of items within each subscale, minimizing overlap of elements across subscales, providing appropriate examples for each item, and outlining missing items.

Specifically, academic experts provided valuable feedback to adjust some of the questions. For example, within the Physical Safety subscale, an item was changed from “Coach(es) respond appropriately” to “Program staff respond appropriately” to not discredit the coach if another individual (e.g., trainer) was tasked with directly attending to injured youth. The academic experts also suggested further questions. For example, within the Program Demographic Form, two questions were added to address the parents’ level of involvement within the program (i.e., “Are parents welcome at practices?” “What is the level of parental involvement?”). Based on academic feedback, some items were also moved from one subscale to another. For example, “Coach(es) mediate exclusive/conflict behaviour from youth appropriately” was moved from Positive Social Norms to Psychological Safety (item 3). Finally, as overlap between the eight setting features exists (Eccles & Gootman, 2002), academic experts provided suggestions on how to minimize such overlap. For example, the items “Coach(es) promote empathetic behaviours amongst youth” and “Coach(es) encourage all youth to participate in the activities” within the Opportunities to Belong subscale were removed as they were deemed by academics and coaches as being redundant.

In the questionnaire, coaches rated the items as relevant to their current sport practice ($M = 8.89$, $SD = .52$; Range = 6.56–9.47) and clearly worded ($M = 8.91$, $SD = .22$; Range = 8.21–9.40). Four items were outlined as either passive or detracting from their sport practice, with mean scores below 8 out of 10. These items fell within the Opportunities for Skill-Building–Life Skills (e.g., “Coach(es) provide opportunities for youth to improve life skills through practice”) and Integration of Family, School, and Community Efforts (e.g., “Program provides youth opportunities to work with their community and practice their learned skills”) subscales.

The four items deemed less relevant by coaches were nonetheless retained for two reasons. First, the subscales in which the items fell have been recognized within the academic literature as elements of high-quality programs (e.g., Bean & Forneris, 2016a; Gould & Carson, 2008). Second, academics were further consulted to provide their judgement on whether to retain or remove these four items and all advised to retain them. Changes made based on the academic and coach feedback led to item reduction from 64 to 54. The 54 items were used in step four.

2.5. Step four: piloting the measure

To further assess face and content validity, seven researchers (two individuals on the research team and five research assistants) piloted the measure. To increase the quality of pilot testing, a preliminary meeting was held surrounding proper use of the PQAYS. This included outlining the purpose of the measure, how to use the measure, and how to score the PQAYS items. Scenario-
based questions and case study examples were used during the meeting to test researcher comprehension prior to the commencement of pilot data collection.

The seven researchers piloted the measure within three community sport settings using the optimal procedure (i.e., completion of Program Demographic Form and pre-post interviews, minimum of two observers present). After each of the seven researchers had completed a minimum of three observation sessions, they met to discuss their overall experiences using the measure and the associated tools, including any concerns or difficulties related to clarity or scoring. This meeting was audio-recorded, transcribed, and reviewed. Following a specific recommendation from the academic panel of experts, the researchers documented the length of time it took them to complete the measure after their observation sessions. Completion of the measure ranged between 30 and 60 min, yet the length of time tended to decrease as researchers became more familiar with the process. For example, researchers came to recognize certain situations, behaviors, or interactions that fit within specific PQAYS items, and thus, they referenced these in their field notes (e.g., this behavior supports item 2.3), which made the completion of the measure more efficient. Minor wording changes were made to the measure at this time. Slight modifications were also made to the interview guides by adding and rearranging some questions to improve the flow of the interviews (Maxwell, Chmiel, & Rogers, 2015).

2.6. Step five: finalizing the measure
The piloting process presented in step four resulted in further modifications being made, with the PQAYS reduced from 54 to 51 items representative of the 10 elements of program quality (see Appendix A for full measure). Specifically, the item “There is physical evidence of positive social norms within the environment (e.g., motto/slogan present, youth wear team clothing or have team bags)” was removed from the Positive Social Norms subscale. The researchers concluded from their pilot work that wearing team clothing or having a physical motto present within the sport context did not necessarily influence the quality of social norms that were fostered within a program. Further, after piloting the measure, the importance of the Program Demographic Form was recognized to attain a comprehensive baseline of what to expect from the program prior to observation (e.g., number of participants).

3. Study two: measurement testing
The purpose of the second study was to further test the reliability and validity of the PQAYS developed in study one. Descriptive statistics of the 10 PQAYS subscales, including: (a) Physical Safety (n = 8 items); (b) Psychological Safety (n = 3 items); (c) Appropriate Structure (n = 7 items); (d) Supportive Relationships (n = 5 items); (e) Opportunities to Belong (n = 3 items); (f) Positive Social Norms (n = 3 items); (g) Support for Efficacy and Mattering (n = 8 items); (h) Opportunities for Skill-Building—Sport and Physical Skills (n = 5 items); (i) Opportunities for Skill-Building—Life Skills (n = 4 items); and (j) Integration of Family, School, and Community Efforts (n = 5 items) and the total measure can be found in Table 2. Two forms of reliability were examined: internal consistency and inter-rater reliability. Preliminary testing of convergent validity was also conducted by correlating scores of the PQAYS with scores from a questionnaire that assesses youth perceptions of program quality and predictive validity was assessed using a measure of youth perceived developmental experiences.

4. Method

4.1. Context and procedure
Following ethical approval from the research team’s university Research Ethics Boards, the lead researcher contacted various youth sport programs across Southeastern Ontario in Canada. Study information, including the overall purpose and procedures, was communicated to program leaders and coaches who were interested. Coaches agreed to participate at varying capacities (e.g., solely the observational portion of the study or the completion of both the observations and the questionnaire). Coaches from 52 programs agreed to engage in the observational portion of the study;
17 sport programs run by non-profit organizations that serve youth from low-income neighborhoods and 35 community club sport programs were involved. Within these 52 programs, there was a range of developmental (n = 20), recreational (n = 12), and competitive (n = 20) sport programs across a variety of sport (e.g., football, golf, basketball, dance, baseball, soccer, ball hockey, and ice hockey). Programs involved girls only (n = 12), boys only (n = 9), and mixed (girls and boys; n = 31) teams. Program sessions ran between 60 and 240 min in length (M = 115.48 min) and were offered between one and five times per week. Youth involved in these programs ranged from 5 to 18 years of age. Enrolment within a given program ranged from 6 to 32 youth. Coaches from 24 of the 52 programs agreed to have youth (n = 322) complete self-report questionnaires in addition to the program observations. Prior to conducting observations, consent and assent forms were distributed to and completed by coaches, parents, and youth involved in the programs.

In all, 307 observation sessions were conducted across the 52 programs, with an average of 4.89 (SD = 1.53, range = 3 to 10) sessions observed per program over the course of 24 months. Steps were taken to reduce social desirability during observations by: (a) reiterating to coaches that the objective of the study was to understand program quality as a whole, not solely coaches’ performance, (b) reminding coaches that participation in the study was voluntary, (c) assuring coaches that observation scores would remain confidential, and (d) ensuring that researchers sat in unobtrusive areas while observing the program sessions.

4.2. Measures

4.2.1. PQAYS
This measure was outlined in study one.

4.2.2. Youth program quality survey (YPQS)
An adapted version of the Youth Program Quality Survey (YPQS) was used to assess convergent validity to attain youth’s perceptions of program quality (Bean & Forneris, 2016b; Silliman & Schumm, 2013). Such a measure was selected because it is also based on the NRCIM’s eight setting features (Eccles & Gootman, 2002). It was important to understand if the observed assessment of program quality (measured by the PQAYS) was congruent with youth’s perceptions of their program experiences. As the YPQS is relatively new, few studies have utilized this measure; however, past findings have revealed moderate to high instrument reliability (α = .60-.96; Silliman, 2008; Silliman & Schumm, 2013; Silliman & Shutt, 2010).

One study (Bean & Forneris, 2016b) outlined a poor model fit and as such, modifications were made based on the results of an exploratory factor analysis that showed good model fit (CFI = .932, TLI = .920, SRMR = .0456, RMSEA = .037). The modifications included reducing the measure from a 24-item measure to a 19-item measure for youth between 10 and 18 years of age. The adapted version of the YPQS was used in the present study. The 19 items are comprised within four subscales: (a) Appropriate Adult Support and Structure (five items; e.g., “Rules and expectations were clear” and “Adults listened to what I had to say”), (b) Empowered Skill-building (seven items; e.g., “I was challenged to think and build new skills”), (c) Expanding Horizons (four items; e.g., “I gained a broader view of the world beyond my community”), and (d) Negative Experiences (three items; e.g., “I felt like I didn’t belong”) in which all eight program setting features are represented (Bean & Forneris, 2016b). The YPQS is measured on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree). With the current sample, the YPQS showed good internal consistency (αsubscale range = .71-.89; αtotal measure = .90).

4.2.3. Short-form youth experience survey for sport (YES-S)
To further test the validity of the PQAYS, the short form of the YES-S was used as a measure of youth’s developmental experiences in sport. This scale is comprised of 23 items that assess youth’s perceptions of their personal and interpersonal developmental experiences as well as...
negative experiences in youth sport (Sullivan, LaForge-MacKenzie, & Marini, 2015). Adapted from MacDonald, Côté, Eys, and Deakin (2012), four subscales of this measure were used: Personal and Social Skills (four items; e.g., “Learned about controlling my temper”), Goal Setting (four items; e.g., “Learned to find ways to reach my goals”), Initiative (four items; e.g., “Put all my energy into this activity”), and Negative Experiences (five items; e.g., “Adult leaders intimidate me”). The questionnaire was measured on a 4-point Likert scale from 1 (yes, definitely) to 4 (not at all). With the current sample, the internal consistency for this scale was good ($\alpha_{\text{subscale}}$ range $=.73-.90$; $\alpha_{\text{total measure}} = .87$).

4.3. Internal consistency reliability

Internal consistency was tested using Cronbach’s alpha. Nunnally’s (1978) criteria of .7 is widely used as the acceptable standard for scale reliability, but alpha’s as low as .5 or .6 have also been identified as acceptable (e.g., Nunnally, 1967; Peterson, 1994). Within eight of the 10 subscales of the PQAYS, Cronbach’s alpha statistics demonstrated high levels of internal consistency (i.e., >.7; see Table 2). Two subscales (Physical Safety, Opportunities to Belong) fell just below .7 (Nunnally, 1978). This can happen when subscales have a small number of items or when subscales measure a wide range of constructs (Cortina, 1993; Tavakol & Dennick, 2011). For example, Opportunities to Belong may have a lower Cronbach’s alpha because it is comprised of only three items. Physical Safety measures many different constructs (i.e., if the program space is free of hazards, accessibility to first aid supplies, if proper sporting equipment is worn), which may have contributed to lower reliability. What is most important to note is that the internal consistency of the overall PQAYS yielded good internal consistency ($\alpha = .84$), reinforcing the importance of assessing program quality as a holistic construct. Eccles and Gootman (2002) argued that in order to achieve a high-quality program, programmers must incorporate all eight setting features. Many researchers have identified challenges surrounding the use of Cronbach’s alpha for internal consistency, whereby a low alpha is not necessarily associated with low reliability (e.g., Dunn, Baguley, & Brunsden, 2014; Henson, 2001). Further, researchers have argued that inter-rater reliability should be considered of greater importance when assessing observational measure reliability (e.g., McHugh, 2012).

| Subscale                                      | $M$  | $SD$  | Range         | Kappa statistic | Cronbach’s alpha |
|-----------------------------------------------|------|-------|---------------|-----------------|------------------|
| 1.1 Physical Safety                           | 4.69 | .31   | 3.92-5.00     | .83             | .63              |
| 1.2 Psychological Safety                      | 4.39 | .69   | 2.00-5.00     | .87             | .74              |
| 2.0 Appropriate Structure                     | 4.35 | .53   | 3.00-5.00     | .67             | .79              |
| 2.1 Supportive Relationships                  | 4.23 | .59   | 2.25-5.00     | .66             | .71              |
| 4.0 Opportunities to Belong                   | 4.22 | .54   | 3.00-5.00     | .64             | .61              |
| 5.0 Positive Social Norms                     | 4.23 | .66   | 2.00-5.00     | .71             | .71              |
| 6.0 Support for Efficacy and Mattering        | 3.77 | .66   | 2.13-5.00     | .67             | .70              |
| 7.1 Opportunities for Skill—Sport and Physical Skills | 3.93 | .81   | 1.00-5.00     | .70             | .71              |
| 7.2 Opportunities for Skill—Life Skills       | 2.65 | 1.35  | 1.00-5.00     | .83             | .88              |
| 8.0 Integration of Family, School, and Community Efforts | 2.77 | .86   | 1.00-5.00     | .91             | .72              |
| Total Scale                                   | 3.93 | .43   | 2.88-4.91     | .75             | .84              |

Note: Scale ranges from a possible score of 1–5.
4.4. Inter-rater reliability

Inter-rater reliability, using the Kappa statistic, was performed to determine consistency among raters. As two researchers were in attendance for every observation, their scores for each item were compared for consistency. For every pair of observations conducted, a score was calculated and then a total score of inter-rater reliability for each subscale was determined. Table 2 outlines the inter-rater reliability statistics for the total measure ($\kappa = 0.75; p < 0.0005$, 95% confidence interval [.74, .76]) and each subscale ($k_{\text{range}} = .61-.88$), outlining consistent and substantial-to-near-perfect agreement between raters (Landis & Koch, 1977).

4.5. Convergent and predictive validity

As outlined above, youth from 24 of the 52 observed programs ($N = 322$, $M$ per team = 13.42, $SD = 7.41$) completed the YPQS and YES-S short to assess their perceptions of program quality and developmental experiences, respectively. These 322 youth (50% boys) ranged from nine to 18 years of age ($M_{\text{age}} = 13.66$, $SD = 2.91$) and their length of involvement in the given program ranged from 1 to 10 years ($M = 3.36$, $SD = 2.89$). Youth identified as Caucasian (59%), Black (14%), Asian (10%), multiracial (8%), Arabic (3%), and Aboriginal (1%) with 5% of youth not disclosing their ethnicity. Paper questionnaires were distributed by researchers to all youth at the end of a program session. Coaches were not present during questionnaire completion to minimize social desirability. Researchers answered youth’s questions related to comprehension.

Convergent validity assesses the agreement between scores of two measures that are believed to assess similar constructs (Schutz & Park, 2004). To assess convergent validity, researcher-scored PQAYS data were compared to the YPQS data that were completed by the youth participants involved in the same sport programs. This procedure was completed to assess whether the PQAYS measured similar constructs to the YPQS. The identified importance of holistic program quality (e.g., Eccles & Gootman, 2002) resulted in two scores of total program quality being calculated (i.e., two mean scores of averaged subscales—one for the observed measure of program quality and one for the youth-perceived measure of program quality) for each of the 24 programs. A Pearson’s correlation coefficient was used to assess if these two variables were correlated. Analysis showed that the PQAYS and YPQS were significantly and moderately correlated ($r = 0.52$, $p = 0.001$).

Predictive validity measures the extent to which a construct measured using scales predicts scores on another measure (Cronbach & Meehl, 1955). Knowing that program quality is one of the best predictors of positive developmental outcomes in youth programming (e.g., Durlak et al., 2010; Yohalem & Wilson-Ahlstrom, 2010), it is important to assess this relationship. However, to date, there has been no measure that can systematically measure program quality in youth sport. Thus, the regression included the total score of program quality regressing on the total score of the YES-S short to determine if program quality predicted developmental experiences within the 24 programs. Both total scores were averaged using the mean of the average of YES-S short subscales. A regression analysis was conducted which indicated that observed program quality significantly predicted youth’s perceptions of psychosocial experiences, whereby program quality accounted for 21% of the variance ($F(1, 22) = 5.73$, $p = .026$, $R^2 = .21$).

5. Discussion

The purpose of this paper is to report on two studies conducted to develop a valid and reliable observational measure to assess program quality processes in youth sport. Study one was conducted to establish content and face validity for the observational measure and used academic and applied expert panels. Study one resulted in the creation of the 51-item PQAYS measured across 10 subscales based on Eccles and Gootman’s (2002) eight program setting features. The second study further tested the reliability and validity of the PQAYS through internal consistency reliability, inter-rater reliability, and convergent and predictive validity. The results provide initial evidence to support the reliability and validity of this measure and demonstrated the potential for using the PQAYS as an observational measure of program quality within youth sport. It should be acknowledged that the results provide evidence for the validity and reliability of the PQAYS, but
only when the aforementioned optimal procedures presented in study one are followed (i.e., completing the Program Demographic Form, having two observers present for each observation, conducting a minimum of three observations, conducting pre- and post-interviews).

The development of a valid and reliable measure of program quality is justified because program quality has been outlined as one of the best predictors of developmental outcomes within youth programming (e.g., Roth & Brooks-Gunn, 2015; Yohalem & Wilson-Ahlstrom, 2010). If we consider that sport is the most popular extracurricular activity for North American youth, assessing the quality of sport programs must be a priority.

The eight setting features work together as building blocks (Eccles & Gootman, 2002; HSERF, 2005), whereby providing a positive climate and creating an appropriate structure act as the foundation for higher-order elements of program quality to occur (i.e. opportunities for skill-building). Singer, Newman, and Moroney (2018) argued that program quality should indeed be viewed as hierarchical, where “participants should experience a safe and supportive environment, so that ultimately they can engage in positive relationships and skill building” (p. 196). Therefore, to achieve quality programs, programmers should strive to incorporate a balance of all eight setting features into their programs (Eccles & Gootman, 2002). In sport, this means that coaches must be intentional in their approaches, with deliberate and strategic decisions made to create opportunities that maximize their athletes’ psychosocial development (Walker, Marczak, Blyth, & Borden, 2005). For example, the importance of coaches adopting explicit approaches to foster the development and transfer of life skills has recently been outlined in a continuum of intentionality (Bean, Kramers, Forneris, & Camiré, 2018) and bolsters the importance of adopting intentional approaches in delivering quality sport programs.

The development of the PQAYS has important practical implications. Researchers have called for more work to examine how youth sport programs influence youth’s developmental outcomes and program quality has been identified as an important variable to assess (e.g., Holt & Sehn, 2008; Petitpas, Cornelius, & Van Raalte, 2008). The PQAYS can be used when working with coaches to understand the components needed to deliver quality youth sport programs. As program quality is dependent on the fidelity of program implementation (Baldwin & Wilder, 2014), delivering a quality program is beneficial not only for youth psychosocial development, but also for the retention of youth within a given program. Further, using the PQAYS to understand the current state of youth sport may also be helpful in identifying gaps within coach education and developing material to help close those gaps.

As outlined in the PQAYS instructions, this tool can be used in different ways depending on the research goals. Specifically, if the measure is used as part of an intervention or single case study, conducting more than three minimum observations is suggested. Further, it is critical that all steps be carried out (e.g., pre-post interviews) to gain a comprehensive perspective of the program through multiple methods. In contrast, if the purpose of using the PQAYS lies in making comparisons across many sport programs, it may not be feasible to completely follow the outlined optimal procedure, from a time and resources point of view. Nonetheless, for ideal PQAYS usage, it is recommended to conduct pre-post interviews, observe a minimum of three program sessions, have multiple observers, and complete the Program Demographic Form.

A strength of the PQAYS is that it is explicitly structured to account for the eight setting features developed by Eccles and Gootman (2002). Previous measures (e.g., YPQA, YPQS, Youth and Program Strengths Survey) relied upon inference as it relates to which items comprised each of the eight setting features. Having the PQAYS structured explicitly around the eight features allows academics and practitioners to better assess program quality within the sport context, understanding where program strengths lie, and where improvements are needed.

Findings from study two outlined that Opportunities for Skill-Building–Life Skills was rated substantially lower than all other subscales of program quality. Much theoretical (e.g., Fraser-Thomas
et al., 2005; Petitpas et al., 2005) and empirical research (Bean & Forneris, 2017) has emphasized the importance of intentionally teaching life skills within the sport context. Specifically, in one study, sport programs that were intentionally structured to teach life skills scored higher on program quality compared to sport programs that did not intentionally teach life skills (Bean & Forneris, 2016a). Such findings reinforce previous assertions that coach education is needed related to how coaches can intentionally teach life skills to foster the positive development of youth within sport (Vella, Oades, & Crowe, 2011). The goal of the PQAYS is to help both researchers and practitioners (e.g., coaches, youth sport directors) better assess the quality of their programs and thus be in informed positions to make choices (e.g., coaches accessing further training) to ensure that youth are afforded quality sport experiences.

5.1. Limitations and future research
This study represents a first step towards contextualizing important elements of program quality within youth sport. Having PQAYS users follow the detailed guidelines outlined in the instructions allows for a comprehensive understanding of program quality, through the use of multiple methods (interviews, observational field notes, quantitative documentation) and procedures that enhance rigor (e.g., multiple observers over multiple sessions). Despite the rigors built within the PQAYS procedures, it is inevitable that when using this tool, observers will rely on their own perceptions and lived experiences (e.g., Haerens et al., 2013), which should be acknowledged as a limitation. However, the potential for engaging in a bracketing interview prior to data collection may help minimize this bias. The goal of a bracketing interview is to enhance the researcher’s reflexivity and awareness of potentially unacknowledged preconceptions that may influence the research process (Rolls & Relf, 2006). Future work with the PQAYS can include video recording program sessions, as the use of video can aid in the effective and objective use of observational measures (see Smith et al., 2016 for review). Although the Program Demographic Form and interviews with coaches can help researchers gather key information pertaining to programmatic structure, future research is warranted to refine the gathering of such information, which is crucial in contextualizing observations.

Further, few individual sport programs were included within the current study despite attempts to recruit in this context. To continue to explore this area, data have recently been collected using the PQAYS within the sport of golf. Moreover, we recognize that “sport” is not a homogenous context, but rather includes a myriad of different structures (e.g., rules, competition levels). As such, research is ongoing to continue to test the validity and reliability of the measure across a variety of youth contexts (e.g., recreational and competitive; male and female; indoor and outdoor). However, it is important to note that the eight setting features have been assessed qualitatively and identified as prevalent in competitive, recreational, and summer sport camp contexts (e.g., Povilaitis & Tamminen, 2017; Strachan et al., 2011). This past research provides a foundation to build on and reinforces the use of multiple methods to explore program quality.

Another limitation relates to the lower reliability for two subscales. However, Nunnally (1967) outlined that in “the early stages of research on predictor tests or hypothesized measures of a construct, . . . reliabilities of .60 or .50 will suffice” (p. 226). As the PQAYS constitutes a new measure, further testing is underway to examine reliability of these two subscales within a different sample, as well as investigating additional characteristics of relevance that do not overlap with other subscales within the measure. Moreover, when using an observational measure, inter-rater reliability is considered of greatest importance when assessing reliability (e.g., McHugh, 2012). Several researchers have outlined issues surrounding the use of Cronbach's alpha to measure internal consistency within observational measures (e.g., Sijtsma, 2009). Within the current study, all subscales achieved substantial or almost perfect agreement between raters (Landis & Koch, 1977). In addition, as noted above, the low reliability on a subscale may be due to the wide range of constructs that fall within the subscale.
As program quality is a relatively new area of study within sport research, few measures exist to test the validity of the PQAYS. For example, the YPQS, one of only a few self-report measures of program quality, has received little psychometric testing. Given that research on youth sport is growing, there is a need for more research to validate the PQAYS and develop additional quantitative measures (MacDonald & McIssac, 2016). Future research with the PQAYS will consist of assessing predictive and structural validity with a sample of a minimum of a 10:1 participant-to-item ratio to conduct a confirmatory factor analysis (Nunnally & Bernstein, 1994). Examining how certain elements of program quality influence outcomes (e.g., Bean & Forneris, 2016c) has only recently been examined in the literature; thus, more research is needed to investigate if certain elements of program quality have greater influence on youth development than others.

6. Conclusion
To date, access to tools for assessing program quality within youth sport has been limited (Holt, Deal, & Smyth, 2016; Holt & Jones, 2008), with the majority of measures being self-report. Thus, the PQAYS addresses a gap by offering a research-based observational measure that can be used to assess the quality of youth sport programming. Although research has provided recommendations for quality improvement within the youth programming context (e.g., Baldwin, Stromwall, & Wilder, 2015), using the PQAYS can aid in capacity building for coaches and programs in ways that optimize the positive development of youth.
Botvin, G. J. (2004). Advancing prevention science and practice: Challenges, critical issues, and future directions. Prevention Science, 5, 69–72. doi:10.1023/B:PREV.0000013984.83251.8b

Brewer, C. J., & Jones, R. L. (2002). A five-stage process for establishing contextually valid systematic observation instruments: The case of rugby union. The Sport Psychologist, 16, 139–161. doi:10.1123/tpsp.16.2.138

Camiré, M., Forneris, T., Trudel, P., & Bernard, D. (2011). Strategies for helping coaches facilitate positive youth development through sport. Journal of Sport Psychology in Action, 2, 92–99. doi:10.1080/21520704.2011.584246

Camiré, M., Trudel, P., & Forneris, T. (2012). Examining how model youth sport coaches learn to facilitate positive youth development. Physical Education and Sport Pedagogy, 19(1), 1–17. doi:10.1080/17408989.2012.726975

Catalano, R. F., Berglund, M. L., Ryan, J. A., Lonczak, H. S., & Hawkins, J. D. (2004). Positive youth development in the united states: Research findings on evaluations of positive youth development programs. The Annals of the American Academy of Political and Social Science, 591, 98–124.

Clayton, M. J. (1997). Delphi: A technique to harness expert opinion for critical decision-making tasks in education. Education Psychology, 17, 373–386. doi:10.1080/0144341970170401

Coakley, J., & Donnelly, P. (2009). Sports in society: Issues and controversies (2nd ed.). Toronto: McGraw-Hill Ryerson.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78, 98–104. doi:10.1037/0021-9010.78.1.98

Côté, J., & Abernethy, B. (2012). A developmental approach to sport expertise. In S. Murphy (Ed.), The Oxford handbook of sport and performance psychology (pp. 435–447). New York: Oxford University Press.

Côté, J., & Fraser-Thomas, J. (2011). Youth involvement and positive development in sport. In P. Crocker (Ed.), Sport psychology: A Canadian perspective (2nd ed., pp. 226–255). Toronto: Pearson.

Côté, J., & Hancock, D. J. (2014). Evidence-based policies for youth sport programmes. International Journal of Sport Policy Politics, 8(1), 1–15. doi:10.1080/1904069x.2014.919338

Côté, J., & Mallett, C. J. (2012). Review of junior sport framework briefing paper: Positive youth development through sport. Canberra: Australian Sports Commission.

Côté, J., Strachan, L., & Fraser-Thomas, J. (2008). Participation, personal development and performance through youth sport. In N. L. Holt (Ed.), Positive youth development through sport (pp. 34–45). New York: Routledge.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity for psychological tests. Psychological Bulletin, 52, 281–302. doi:10.1037/h0040957

Damon, W. (2004). What is positive youth development? The ANNALS of the American Academy of Political and Social Science, 593, 13–24. doi:10.1177/0002716203260092

Danish, S. J., Forneris, T., Hodge, K., & Heke, I. (2004). Enhancing youth development through sport. World Leisure Journal, 46, 38–49. doi:10.1080/04419035.2004.967435S.

Danish, S. J., Petitpas, A. J., & Hole, B. D. (1993). Life development intervention for athletes: Life skills through sports. The Counseling Psychologist, 21, 352-385.

Dryfoos, J. G. (1990). Adolescents at risk: Prevalence and prevention. New York: Oxford University Press.

Duda, J. L. (1989). The relationship between task and ego orientation and the perceived purpose of sport among male and female high school athletes. Journal of Sport and Exercise Psychology, 11, 318–335.

Duda, J. L. (2013). The conceptual and empirical foundations of empowering coaching: Setting the stage for the PAPA project. International Journal of Sport and Exercise Psychology, 11, 311–318. doi:10.1080/1612197X.2013.839414

Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British Journal of Psychology, 105, 399–412. doi:10.1111/bjop.12046

Durak, J. A., Mahoney, J., Bohner, A., & Parente, M. (2010). Developing and improving after-school programs to enhance youth’s personal growth and adjustment: A special issue of AJCP. American Journal of Community Psychology, 45, 285–293. doi:10.1007/s10464-010-9298-9

Durak, J. A., & Weissberg, R. P. (2007). The impact of after-school programs that promote personal and social skills. Chicago, IL: Collaborative for Academic, Social, and Emotional Learning.

Eccles, J. S., & Gootman, J. A. (2002). Community programs to promote youth development. Washington, DC: National Academy Press.

Elme, R. M., Young, J. A., Harvey, J. T., Charity, M. J., & Payne, W. R. (2013). A systematic review of the psychological and social benefits of participation in sport for children and adolescents: Informing development of a conceptual model of health through sport. International Journal of Behavioral Nutrition and Physical Activity, 10, 1–14. doi:10.1186/1479-5868-10-98

Erickson, K., & Côté, J. (2015). The intervention tone of coaches’ behaviour: Development of the Assessment of Coaching Tone (ACT) observational coding system. International Journal of Sports Science & Coaching, 10, 699–716. doi:10.1260/1747-9541.10.6.699

Erickson, K., Côté, J., Hollenstein, T., & Deakin, J. (2011). Examining coach-athlete interactions using state space grids: An observational analysis in competitive youth sport. Psychology of Sport and Exercise, 12, 645–654. doi:10.1016/j psycsport.2011.06.006

Flett, M. R., Gould, D. R., & Louer, L. (2012). A study of an underserved youth sports program using the youth program quality assessment. Journal of Applied Sport Psychology, 24, 275–289. doi:10.1080/10413200.2011.641061

Fraser-Thomas, J., Côté, J., & Deakin, J. (2005). Youth sport programs: An avenue to foster positive youth development. Physical Education and Sport Pedagogy, 10, 19–40. doi:10.1080/1740898042000334890

Gould, D., & Carson, S. (2008). Life skills development through sport: Current status and future directions. International Review of Sport and Exercise Psychology, 1, 58–78. doi:10.1080/17490539.2008.970187

Guèvremont, A., Findlay, L., & Kohen, D. (2008). Organized extracurricular activities of Canadian children and youth. Health Reports, 19, 65–69. doi:10.1111/josh.12154

Hooerens, L., Aelterman, N., Van den Bergh, L., De Meyer, J., Soenens, B., & Vansteenkiste, M. (2013). Observing physical education teachers’ need-supportive interactions in classroom settings. Journal of Sport & Exercise Psychology, 35, 3–17. doi:10.1123/jsep.35.1.3
Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods Psychology, 8*, 23–34. doi:10.20982/tqmp.08.1.p023

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34*, 177–189.

HighScope Educational Research Foundation. (2005). *Youth program quality assessment. Ypsilanti, MI: HighScope Press.*

Hodge, K., Danish, S., Forneris, T., & Miles, A. (2016). Life skills and basic psychological needs. In N. L. Holt (Ed.), *Positive youth development through sport (2nd ed., pp. 45–56).* New York: Routledge.

Holt, N. L. (Ed.). (2016). *Positive youth development through sport (2nd ed.).* New York: Routledge.

Holt, N. L., Deal, C. J., & Smyth, C. (2016). Future directions for positive youth development through sport. In N. L. Holt (Ed.), *Positive youth development through sport (2nd ed., pp. 229–240).* New York: Routledge.

Holt, N. L., & Jones, M. I. (2008). Future directions for positive youth development and sport research. In N. L. Holt (Ed.), *Positive youth development through sport (pp. 122–132).* New York: Routledge.

Holt, N. L., & Sehn, Z. L. (2008). Processes associated with positive youth development and participation in competitive youth sport. In N. L. Holt (Ed.), *Positive youth development through sport (pp. 24–33).* New York: Routledge.

Jones, M. I. (2015). *Research methods for sports studies (3rd ed.).* New York: Routledge.

Jones, M. I., & Lavellee, D. (2009). Exploring perceived life skills development and participation in sport. *Qualitative Research in Sport and Exercise, 1*, 36–50. doi:10.1080/19398440802567931

Kendellen, K., Camiré, M., Bean, C., Forneris, T., & Thompson, J. (2017). Integrating life skills into golf canada’s youth programs: Insights into a successful research to practice partnership. *Journal of Sport Psychology in Action, 8*, 34–46. doi:10.1080/21520704.2016.1205699

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174. doi:10.2307/2529310

Larson, R. W., & Walker, K. C. (2010). Dilemmas of practice: Challenges to program quality encountered by youth program leaders. *American Journal of Community Psychology, 45*(3–4), 338–349. doi:10.1007/s10464-010-9307-z

Larson, R. W., Walker, K. C., Rusk, N., & Diaz, L. B. (2015). Understanding youth development from the practitioner’s point of view: A call for research on effective practice. *Applied Developmental Science, 19*, 74–86. doi:10.1080/10888691.2016.1199727

MacDonald, D. J., Côté, J., Eys, M., & Deakin, J. (2012). Psychometric properties of the youth experience survey with young athletes. *Psychology of Sport and Exercise, 13*, 332–340. doi:10.1016/j.psychsport.2011.09.001

MacDonald, D. J., & McIsaac, T. (2016). Quantitative assessment of positive youth development in sport. In N. L. Holt (Ed.), *Positive youth development through sport (2nd ed., pp. 83–96).* New York: Routledge.

Maxwell, J. A., Chmiel, M., & Rogers, S. E. (2015). Designing integration in multimethod and mixed methods research. In S. N. Hesse-Biber & R. B. Johnson (Eds.), *Oxford handbook of multimethod and mixed methods research inquiry (pp. 688–706).* New York: Oxford University Press.

McCarthy, P. J., Jones, M. V., & Clark-Carter, D. (2008). Understanding enjoyment in youth sport: A developmental perspective. *Psychology of Sport and Exercise, 9*, 162–176.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemica Medica, 22*, 276–282. doi:10.11613/BM.2012.031

McLaughlin, M. W. (2000). Community counts: How youth organizations matter for youth development. Washington, DC: Public Education Network.

Merkel, D. L. (2013). Youth sport: Positive and negative impact on young athletes. *OpenAccess Journal of Sports Medicine, 4*, 151–160. doi:10.2147/OAJSM.S35356

Merry, S. (2000). *Beyond home and school: The role of primary supports in youth development.* Chicago, IL: Chaplin Hall Center for Children.

Nakaha, J. R., Grimes, L. M., Nadler, C. B., & Roberts, M. W. (2016). A treatment selection model for sibling conflict based on observational measurements. *Journal of Child and Family Studies, 25*, 124–135. doi:10.1007/s10826-015-0210-y

Nunnally, J. C. (1967). *Psychometric theory.* New York: McGraw-Hill.

Nunnally, J. C. (1978). *Psychometric theory (2nd ed.).* New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. (1994). Elements of statistical description and estimation. In J. C. Nunnally & I. H. Bernstein (Eds.), *Psychometric theory (3rd ed.).* New York: McGraw-Hill.

Papacharissis, 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.

Patton, O. M. (2002). *Qualitative research and evaluation methods (3rd ed.).* Thousand Oaks, CA: Sage.

Pechman, E. M., Russell, C. A., & Birmingham, J. (2008). *Out-of-school time (OST) instrument. Retrieved from http://www.most.ie/webreports/Fatima%20reports/ OST/OST%20Observation%20Instrument.pdf*

Peterson, R. A. (1994). A meta-analysis of Cronbach’s coefficient alpha. *Journal of Consumer Research, 21*, 381–391. doi:10.1086/jcr.1994.21.issue-2

Petitpas, A. J., Cornelius, A. E., Van Raalte, J. L., & Jones, T. (2005). A framework for planning youth sport programs that foster psychosocial development. *The Sport Psychologist, 19*, 63–80. doi:10.1123/ tsp.19.1.63

Petitpas, A. J., Cornelius, A. E., & Van Raalte, J. L. (2008). Youth sport through sport: It’s all about relationships. In N. L. Holt (Ed.), *Positive youth development through sport (pp. 61–70).* New York: Routledge.

Pierce, S., Gould, D., & Camiré, M. (2017). Definition and model of life skills transfer. *International Review of Sport and Exercise Psychology, 10*, 186–211. doi:10.1080/1750984X.2016.1199727

Povilaitis, V., & Tamminen, K. A. (2017). Delivering positive youth development at a residential summer sport camp. *Journal of Adolescent Research. doi:10.1177/ 0743558417702478*

Reeves, C. A., & Bednar, D. A. (1994). Defining quality: Alternatives and implications. *Academy of Management Review, 19*, 419–445. doi:10.5465/amr.1994.9412271805

Rolls, L., & Relf, M. (2006). Bracketing interviews: Addressing methodological challenges in qualitative interviewing in bereavement and palliative care. *Mortality, 11*, 286–305. doi:10.1080/ 13576270600774893

Roth, J. L., & Brooks-Gunn, J. (2015). Evaluating youth development programs: Progress and promise.
Appendix A

Use of the Program Quality Assessment in Youth Sport (PQAYS)

Purpose

- The PQAYS is an evidence-based observational instrument that has been designed to assess the quality of youth sport programs. The PQAYS is based on the eight program setting features that have been proposed to foster positive developmental outcomes in youth (Eccles & Gootman, 2002) and can be used as an external assessment for researchers and program...
evaluators. Use of this measure can aid in outlining areas of strength and areas of improvement within a sport program.

Instructions

Prior to Conducting Observations with the PQAYS

• Become familiar with the measure by reading through each of the items and footnotes. This will make it easier to take notes during your observations that will provide a more comprehensive assessment.

• If not already, become familiar with the sport context in which you will observe (e.g., rules, basic culture). To do this, complete the Program Demographic Form with the coach or have the coach complete the form. This form can be found at the end of the instructions. Information within this form will provide context to the program including the program’s regular practice and competition schedule, additional sessions they have besides practices and competitions that may be of value to observe, and when youth typically arrive prior to program sessions.

• In order to gain a better understanding of the team prior to starting the observations we recommend two steps be taken. First, review the vision, mission, values, and objectives of the team and/or organization, which can often be found online. Second, conduct an interview with the coach. This interview will allow you to have the most comprehensive understanding of the environment you will observe. A sample interview guide can be found at the end of this measure.

• Finally, coordinate with the coach regarding the best time to attend the program sessions.

Attending and Observing the Program

• It is strongly recommended to have two observers complete the PQAYS each time a program session is observed for inter-rater reliability.

• Make sure to have a hard copy of the PQAYS to use as a reference when conducting an observation. However, the full scoring and completion of the PQAYS should be done once the program session is over.

• Based on the arrival time outlined in the Program Demographic Form, plan to arrive at the same time as youth to begin your observation (typically between 15 and 45 min prior to regular practice/competition time). This will give you time to not only prepare yourself, but also to communicate with the coach(es), if necessary. The coach(es) often interact with youth prior to the commencement of a program session and such interactions can be valuable for observation.

• When observing a session situate yourself as unobtrusively as possible.

• Observations should be conducted up to 15 min after the scheduled program as well. This is to fully capture the program experience and coach-youth interactions.

• Observe a minimum of three practice sessions at different time points in the program’s duration (the more the better). Observing sessions throughout the season/program allows for the measurement of inter-rater reliability and allows for a good understanding of what regularly occurs within the program.

• If the purpose of using this measure is to make comparisons across various sport programs, then three observations is sufficient. However, if this measure is part of an intervention or case study, we recommend more than three observations.

• The majority of observations should be done at practices, as this is when most youth-coach interaction takes place; however, it is strongly encouraged to observe other sessions that have been outlined in the Program Demographic Form as regular program components, if possible.

• During the observations make sure to refer to all of the corresponding footnotes in the PQAYS for supplementary information.
• Take detailed field notes during the program session.
  o These field notes can be taken by hand or using an electronic device and should be objective and factual, specific, and chronological (use of time markers may be helpful). For example, field notes should include, but not limited to, sequences of events, descriptions of interactions, quotations of coach/youth interactions, and list of materials that provide evidence for individual items with the PQAYS.
  • If feasible, and approved by coach(es), use video recording to aid with corresponding field notes.
  • At the end of a given session, ask coach(es) any follow-up questions you may have.

After Attending the Program

• Complete the PQAYS and supporting “Comments” sections immediately after the program session is over, thereby basing program quality scores on observational evidence and preventing problems related to recall. The “Comments” sections are there to provide justification and evidence for the scores provided within the subscale, and provide space for text from the detailed field notes taken during the program observation.

Scoring

• Score each item from a 1 (never) to 5 (very often). A score of 5 means that the behavior or program characteristic being observed for a given item is highly evident and consistent. A score of 4 means that the behavior or program characteristic is evident but perhaps not as explicit or consistent. A score of 3 means that the behavior or program characteristic is evident but is fairly inconsistent. A score of 2 means that the behavior or program characteristic is not very evident or explicit (e.g., one occasion or example when you would expect more consistency) and finally a score of 1 indicates no evidence of a behavior or program characteristic. For a few specifically outlined items, there is the option of “N/A” (not applicable). This will ensure that a particular item does not lower the score for that subscale if it was not present (e.g., if there was no conflict between youth there was no need for the coach(es) to manage the conflict). However, the majority of the items within the measure should be evident; therefore, if it is not observed during the session, when it typically should be, observers should score a 1 and not N/A (e.g., if the coach(es) did not discuss the importance of developing life skills with youth, this should be scored as a 1 and not N/A).
  • Also, keep in mind that there may be multiple coaches working with the same team. The score needs to reflect all of the coaches and not just one. For example, if more than one coach is actively interacting with youth while two others are not then the program should not be scored a 5 but perhaps a 3 or 4 depending on the level of interaction of all three coaches combined.
  • Scoring of the PQAYS can be calculated by computing averages for each subscale. A total score of program quality can be calculated from computing an average of all items within the measure.

After all Program Observations are Completed

• A second interview should be conducted to follow-up with the coach(es) based on the observations. The purpose of this interview is to further understand elements of program quality observed which may clarify some aspects of the program. A sample interview guide that includes suggested questions to use in the post-observation interview can be found at the end of the measure.

Note: The above instructions detail the process that we strongly recommend when using the PQAYS. However, we understand that not all components may be feasible (e.g., some coaches may
not agree to be interviewed). In these cases, it is encouraged that individuals using this measure to
gain as comprehensive of an assessment of the program as possible.

**Program Demographic Form**

As mentioned above, this form should be completed prior to starting any observations. Key
questions to be asked while discussing this form with the coach concern the structure of the
program including how long the season is, how many practices occur per week, how long the
practices are, when do youth and coach(es) typically arrive for practices, how often competi-
tions or games occur, are there regular activities outside of the practice time that are impor-
tant to observe (e.g., mental training sessions, dry-land training, or social events). This would
ensure the whole program is being observed effectively and no component is being missed.
Moreover, answers to some of these questions may help score items within the last subscale,
which is *Integration of Family, School, and Community Efforts*, and give light to any planned
community events.

Organization name: ________________________________

Program name: ________________________________

Age Range of Youth Athletes: _____
Gender of Youth Athletes: _____

Date and time of scheduled games: ________________________________

Date and time of scheduled practices: ________________________________

Additional program components (e.g., dryland training, mental skills sessions, team building):

Typical time of youth arrival before schedule programming: ________________________________

Typical time of departure after scheduled programming: ________________________________

Any relevant certifications held by coaches (e.g., NCCP, HIVEFIVE®, First Aid): __________

Are parents welcome at practices? What is level of parental involvement? __________

Additional details shared by the coach(es) that will help in conducting the observations:

**Table of Contents**

1.1 Physical Safety
1.2 Psychological Safety
2 Appropriate Structure
3 Supportive Relationships
4 Opportunities to Belong
5 Positive Social Norms
6 Support for Efficacy and Mattering
7.1 Opportunities for Skill-building—Sport and Physical Skills
7.2 Opportunities for Skill-building—Life Skills
8 Integration of Family, School, and Community Efforts
### Session Information

Team name: ____________________

Name of coach(es) observed: ____________________ Observer(s): ____________________

Date & Time of Program Session: ________ Coach to youth ratio: ____________

#### 1.1 PHYSICAL SAFETY

|                              | Never | Rarely | Sometimes | Often | Very Often | N/A |
|------------------------------|-------|--------|-----------|-------|------------|-----|
| 1. Program space is free of obvious hazards (e.g., wet floor, poor lighting, open gates) and allows for reasonable safety | 1     | 2      | 3         | 4     | 5          |     |
| 2. Program is relatively free of distractions and if an unexpected distraction does occur the coach responds appropriately (e.g., curtail loud noises, respectfully removes disruptive parents) | 1     | 2      | 3         | 4     | 5          | 0   |
| 3. Program space is appropriate for the activities that the program is designed to include (e.g., not crowded, equipment can be moved if necessary, utilize available area) | 1     | 2      | 3         | 4     | 5          |     |
| 4. Appropriate supervision is provided to youth (e.g., youth are not left unattended) | 1     | 2      | 3         | 4     | 5          |     |
| 5. First-aid supplies are present or available on site | 1     | 2      | 3         | 4     | 5          |     |
| 6. If an injury occurs, program staff respond appropriately by attending to youth with first-aid, if necessary (e.g., use first-aid supplies, notify trainer) | 1     | 2      | 3         | 4     | 5          | 0   |
| 7. Appropriate sporting equipment worn/used (e.g., helmets worn, proper shoes with laces tied, equipment not broken) | 1     | 2      | 3         | 4     | 5          |     |
| 8. Coach(es) foster positive health practices (e.g., allow time for water breaks, stop activities when injury or physical danger occurs) | 1     | 2      | 3         | 4     | 5          |     |

**Comments:**

1. If no distractions occur, mark as Not Applicable.
2. If no injury occurs, mark as Not Applicable.
## 1.2 Psychological Safety

|   | Never | Rarely | Sometimes | Often | Very Often | N/A |
|---|-------|--------|-----------|-------|------------|-----|
| 1. Coach(es) foster predominately positive emotional climate, with the absence of negative emotional climate (e.g., refrain from sarcasm, encourage use of manners, promote inclusion of youth, no foul language) | 1 | 2 | 3 | 4 | 5 | |
| 2. Coach(es) ensure youth are respectful with one another and intervene appropriately if necessary (e.g., respectful language, no conflicts, absence of bullying) | 1 | 2 | 3 | 4 | 5 | |
| 3. Coach(es) mediate exclusive/conflict behaviour from youth appropriately (e.g., use as a learning experience for youth, acknowledge youth's feelings and explanations, involve youth in deciding appropriate consequences) | 1 | 2 | 3 | 4 | 5 | |

Comments:

3. If no conflict arises during program session, mark as Not applicable.
## 2. Appropriate Structure

|                                    | Never | Rarely | Sometimes | Often | Very Often | N/A |
|------------------------------------|-------|--------|-----------|-------|------------|-----|
| 1. Coach(es) start and end the program on time | 1     | 2      | 3         | 4     | 5          |     |
| 2. Coaches support one another (e.g., are on the same page, provide congruent messages to youth) | 1     | 2      | 3         | 4     | 5          |     |
| 3. Coach(es) are actively involved and engaged in the program session | 1     | 2      | 3         | 4     | 5          |     |
| 4. Clear expectations are established and/or reinforced related to behaviour (e.g., rules, activities) | 1     | 2      | 3         | 4     | 5          |     |
| 5. Purpose of activities is understood (e.g., coach gives clear explanations, little confusion among youth) | 1     | 2      | 3         | 4     | 5          |     |
| 6. Coach(es) ensure that activities are engaging (e.g., youth do not appear bored, appropriate for all youth despite differences in skill-level or age) | 1     | 2      | 3         | 4     | 5          |     |
| 7. The program session is organized appropriately (e.g., effective use of space, effective use of time, effective use of assistant coach and/or captains) | 1     | 2      | 3         | 4     | 5          |     |

**Comments:**

4. If there is only one coach, mark as Not Applicable.
5. If observation is not done early on in the program, observer may miss the establishment of expectations; therefore, focus on the reinforcement of the expectations. This could be something that is followed up within coach interview.

---

Bean et al., Cogent Social Sciences (2018), 4: 1467304
https://doi.org/10.1080/23311886.2018.1467304
### 3. SUPPORTIVE RELATIONSHIPS

|                                                                                           | Never | Rarely | Sometimes | Often | Very Often | N/A |
|-------------------------------------------------------------------------------------------|-------|--------|-----------|-------|------------|-----|
| 1. Coach(es) provide a warm environment (e.g., greet youth, smiling, approachable, maintaining eye contact, positive attitude) | 1     | 2      | 3         | 4     | 5          |     |
| 2. Coach(es) interact with youth (e.g., check in, show interest in youth beyond the sport) | 1     | 2      | 3         | 4     | 5          |     |
| 3. Youth appear to trust coach(es) (e.g., feel comfortable approaching or asking for help) | 1     | 2      | 3         | 4     | 5          |     |
| 4. Youth appear to have positive relationships with one another (e.g., friendly, get along with one another) | 1     | 2      | 3         | 4     | 5          |     |
| 5. Activities encourage the development of supportive relationships among youth (e.g., structured/non-structured opportunities to work with teammates, sense of openness within team) | 1     | 2      | 3         | 4     | 5          |     |

Comments:

6. If the sport being assessed is an individual sport, mark as Not applicable.
### 4. OPPORTUNITIES TO BELONG

|                      | Never | Rarely | Sometimes | Often | Very Often | N/A |
|----------------------|-------|--------|-----------|-------|------------|-----|
| 1. Youth appear to   | 1     | 2      | 3         | 4     | 5          |     |
| positively associate |       |        |           |       |            |     |
| with the program     |       |        |           |       |            |     |
| and their teammates  |       |        |           |       |            |     |
| (e.g., friendships,  |       |        |           |       |            |     |
| social ties)         |       |        |           |       |            |     |
| 2. Youth are inclusive and | 1    | 2      | 3         | 4     | 5          |     |
| cooperate with one   |       |        |           |       |            |     |
| another (e.g.,      |       |        |           |       |            |     |
| communicate during  |       |        |           |       |            |     |
| drills and plays,    |       |        |           |       |            |     |
| work together when  |       |        |           |       |            |     |
| necessary)?          |       |        |           |       |            |     |
| 3. Additional activities are | 1    | 2      | 3         | 4     | 5          |     |
| provided beyond      |       |        |           |       |            |     |
| regular programming  |       |        |           |       |            |     |
| that provide         |       |        |           |       |            |     |
| opportunities for    |       |        |           |       |            |     |
| youth to develop    |       |        |           |       |            |     |
| relationships with    |       |        |           |       |            |     |
| teammates and coach  |       |        |           |       |            |     |
| (es) (e.g., team     |       |        |           |       |            |     |
| building activities, |       |        |           |       |            |     |
| dinners, fundraisers)|       |        |           |       |            |     |

Comments:

7. If the sport being observed is individual, mark as Not Applicable.
8. Completion of the Program Demographic Form and pre-interview will help in answering this item. It is also important to follow-up with coach(es) throughout the program as activities and opportunities may change.
9. These activities do not have to take place outside of regularly scheduled program time, yet is an activity designed to provide opportunities to develop relationships (e.g., holiday party).
5. POSITIVE SOCIAL NORMS

|                                                                 | Never | Rarely | Sometimes | Often | Very Often | N/A |
|-----------------------------------------------------------------|-------|--------|-----------|-------|------------|-----|
| 1. Coach(es) model positive social norms (e.g., use respectful language, shows respect for other team and officials) | 1     | 2      | 3         | 4     | 5          |     |
| 2. Fair play is practiced (e.g., youth follow the rules of the sport in practice/competition)                        | 1     | 2      | 3         | 4     | 5          |     |
| 3. There is evidence of sportsmanship (e.g., youth cheer and congratulate teammates for good effort/success, shake hands with opponents at end of competition) | 1     | 2      | 3         | 4     | 5          |     |

Comments:

Bean et al., Cogent Social Sciences (2018), 4: 1467304
https://doi.org/10.1080/23311886.2018.1467304
### 6. SUPPORT FOR EFFICACY AND MATTERING

|                  | Never | Rarely | Sometimes | Often | Very Often | N/A |
|------------------|------|-------|----------|-------|------------|-----|
| 1. Coach(es) focus on youth progressing in a developmentally appropriate manner (e.g., progressive activities, modifications based on skill level) | 1    | 2     | 3        | 4     | 5          |     |
| 2. Coach(es) provide youth with choices and options throughout the program (e.g., jersey colour, drill or skill to practice, activity order) | 1    | 2     | 3        | 4     | 5          |     |
| 3. Coach(es) acknowledge achievements based on youth’s effort (e.g., foster a task-mastery climate) | 1    | 2     | 3        | 4     | 5          |     |
| 4. Youth have opportunities to mentor other teammates (e.g., help others with equipment or with a skill/ activity) | 1    | 2     | 3        | 4     | 5          |     |
| 5. Coach(es) give positive feedback to help enhance efficacy (e.g., praise and encouragement) | 1    | 2     | 3        | 4     | 5          |     |
| 6. Coach(es) give constructive feedback to enhance efficacy (e.g., technical instruction when youth struggle, guidance to why error occurred) | 1    | 2     | 3        | 4     | 5          |     |
| 7. Coach(es) recognize individual contributions (e.g., use youth names, providing individualized feedback) | 1    | 2     | 3        | 4     | 5          |     |
| 8. Coach(es) listen to youth and respond appropriately (e.g., does not ignore questions or concerns) | 1    | 2     | 3        | 4     | 5          |     |

**Comments:**

10. Higher levels of choice results in a higher score; use comments section and field notes to document the types and quality of choices (e.g., youth choosing a drill for warm-up would be scored higher than choosing position or jersey colour).

11. A task-mastery climate is one where coach(es) offers positive reinforcement when youth demonstrate hard work, improvement, and by eliciting an equal contribution from each youth.

12. If the sport being observed is individual, mark as Not Applicable.
7. OPPORTUNITIES FOR SKILL-BUILDING—SPORT AND PHYSICAL SKILLS

| 1. Coach(es) appear to understand the fundamental rules and skills of the sport | Never | Rarely | Sometimes | Often | Very Often | N/A |
|---|---|---|---|---|---|---|
|  | 1 | 2 | 3 | 4 | 5 |   |

| 2. Coach(es) discuss the importance of learning/developing physical skills (e.g., why and how a skill is useful) | Never | Rarely | Sometimes | Often | Very Often | N/A |
|---|---|---|---|---|---|---|
|  | 1 | 2 | 3 | 4 | 5 |   |

| 3. Coach(es) model the physical skills (e.g., demonstrates skills, use proper technique, break down challenging skills) | Never | Rarely | Sometimes | Often | Very Often | N/A |
|---|---|---|---|---|---|---|
|  | 1 | 2 | 3 | 4 | 5 |   |

| 4. Coach(es) provide opportunities for youth to practice physical skills (e.g., repetition when learning new skills, relates to session focus, self-guided) | Never | Rarely | Sometimes | Often | Very Often | N/A |
|---|---|---|---|---|---|---|
|  | 1 | 2 | 3 | 4 | 5 |   |

| 5. Coach(es) provide opportunities for youth to reflect on their progress of sport/physical skills throughout the session (e.g., verbal or video feedback, ask for youth input on progress/future development) | Never | Rarely | Sometimes | Often | Very Often | N/A |
|---|---|---|---|---|---|---|
|  | 1 | 2 | 3 | 4 | 5 |   |

Comments:
### 7. 2 OPPORTUNITIES FOR SKILL-BUILDING—LIFE SKILLS

|   | Never | Rarely | Sometimes | Often | Very Often | N/A |
|---|-------|--------|-----------|-------|------------|-----|
| 1. Coach(es) discuss importance of developing life skills with youth (e.g., how communication will enhance flow on the court, how teamwork is important for working towards goals, encourages responsibility of personal and team equipment) | 1 | 2 | 3 | 4 | 5 |   |
| 2. Coach(es) model the life skills explicitly discussed (e.g., if life skill is communication, displays effective communication; if life skill is emotional regulation, demonstrates breathing activity) | 1 | 2 | 3 | 4 | 5 |   |
| 3. Coach(es) provide opportunities for youth to improve life skills through practice (e.g., life skill-specific activity integrated, feedback provided) | 1 | 2 | 3 | 4 | 5 |   |
| 4. Coach(es) debrief how life skills can be applied and transferred outside of specific sport context (e.g., effectively communicating at home, problem-solving at school, coping strategies in other sports) | 1 | 2 | 3 | 4 | 5 |   |

**Comments:**

---

13. Examples of life skills include: emotional skills (e.g., focus, emotional regulation, conflict resolution, moral character); intellectual skills (e.g., decision-making, critical reasoning, goal-setting); and social skills (e.g., teamwork, communication, cooperation, connectedness to adults and teammates).
### 8. INTEGRATION OF FAMILY, SCHOOL, AND COMMUNITY EFFORTS

|                                                                 | Never | Rarely | Sometimes | Often | Very Often | N/A |
|------------------------------------------------------------------|-------|--------|-----------|-------|------------|-----|
| 1. Families are welcome to the program (e.g., doors are open for families to watch, coaches greet parents upon arrival, families are aware of program’s expectations and rules) | 1     | 2      | 3         | 4     | 5          |     |
| 2. Communication between coaches and parents and/or families is positive (e.g., respect between parent(s) and coach(es) is mutual) | 1     | 2      | 3         | 4     | 5          | 0   |
| 3. Families have opportunities to be involved in the program (e.g., volunteering, attending events, helping with practice, providing snacks) | 1     | 2      | 3         | 4     | 5          |     |
| 4. Coaches communicate with schools or other community organizations to support youth and provide other opportunities to develop life skills (e.g., decision-making, problem-solving) | 1     | 2      | 3         | 4     | 5          |     |
| 5. Program provides youth opportunities to work with their community and practice learned skills (e.g., fundraising within community, volunteering, mentoring opportunities) | 1     | 2      | 3         | 4     | 5          |     |

**Comments:**

---

14. Completion of the Program Demographics Form will help to answer certain items within this subscale. It is also important to follow-up with the coach throughout the program as activities and opportunities may change.

15. Item should be scored based on if parents are welcome to the program, not if they attend.
Sample Interview Guide Questions for Pre-Observation Interview

(A) Introduction, review purpose of interview
(B) Demographics—Age, gender, years of involvement (if not already completed)
(C) Interview Questions

(1) Tell me a little about how you became a coach and why you choose to coach?
(2) How did you become involved in [name of program]?
(3) How would you describe your overall approach to coaching youth (e.g., coaching philosophy)?
(4) What are your main goals and objectives for this program/team?
   (a) What strategies or activities do you incorporate into your coaching practice to achieve these goals and objectives?
   (b) Do you believe those strategies are having a positive influence?
   (c) What about challenges associated with coaching this program/team?
(5) Will/Have you communicated your expectations to youth at the beginning of the season? What were they? Explain.
   (a) How were these expectations developed?
   (b) How will you try to reinforce them? In what ways/examples?
(6) What roles, if any, do parents play in the program? Are they welcome to attend the practices/games?
(7) As part of your program, are there activities that extend beyond regular practices and competitions, such as team bonding, fundraising, or volunteering? If so, what are they? How often do they occur? Explain.
(8) Overall, what has your experience been like coaching so far this season?
(9) Is there anything else you would like to share with me that will help in gaining a sense of your program/team or your work with them?

Sample Interview Guide Questions for Post-Observation Interview

There are number of questions below. You may not need to ask all of these questions but can choose the ones you feel would be most helpful to follow-up on based on the scoring of the PQAYS. In addition, you may want to add some of your own questions.

(A) Introduction, review purpose of interview
(B) Demographics—Age, gender, years of involvement (if not already completed)
(C) Interview Questions

(1) Since we last talked, tell me about your program experiences (e.g., changes/successes/challenges).
(2) What do you feel you have learned about your coaching this season?
(3) Have there been any new activities or strategies you have used? Why?
   (a) Have these been successful? Explain.
(4) Has anything changed related to your overall approach to coaching youth? If so, what and why?
(5) Do you have strategies to ensure a psychologically safe environment for youth (e.g., helping youth feel welcome and included)? Explain.
(6) How would you describe your relationship with youth this season?
(7) Do you consider yourself to be a model for youth in the program?
   (a) What do you do specifically to act as a model within the program?

(8) How would you describe the relationships youth have with each other?
   (a) Are there strategies you use to foster/create positive relationships among youth? Examples.

(9) As a result of participating in this program, do you believe youth are developing life skills (e.g., emotional regulation, focus, goal setting, respect, teamwork)? How? Examples.

(10) Would you say that you intentionally taught these life skills?

(11) Do you feel that youth in the program have a voice (e.g., are able to help make decisions, opportunity to share their thoughts)?
   (a) Choice in the activities/drills included in your program?
   (b) Strategies you have found effective? For what reasons?

(12) By participating in this program, do you believe youth are developing their sense of competence (a belief in themselves)?
   (a) How have you tried to foster this competence? Specific strategies?

(13) How do you provide feedback to youth? Examples.

(14) Do you think it is important for youth to have a say or choice in the activities/drills included in your program?
   (a) What strategies you have found effective? For what reasons?

(15) Do you feel it is important to encourage youth to help/learn from one another?
   (a) Did you see any occurrences of youth mentoring other youth during the program?
   (b) How do you try to do this in your program? Examples.

(16) What strategies do you use to keep youth engaged? Which strategies do you find are the most effective? Least effective?

(17) Based on what we have talked about today, is there anything else you would like to discuss?
