Evaluation and Program Planning 55 (2016) 67–76

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Evaluation development for a physical activity positive youth development program for girls

Sarah Ullrich-French*, Amy N. Cole, Anna K. Montgomery

Washington State University, United States

ARTICLE INFO

Article history:
Received 22 April 2015
Received in revised form 5 November 2015
Accepted 10 December 2015

Keywords:
Physical activity
Positive youth development
Collaborative evaluation development

ABSTRACT

Girls on the Run (GOTR) is an after school program for girls in third through fifth grade which utilizes a physical activity based positive youth development curriculum that culminates with completing a 5 K run. Unfortunately, there is little empirical data documenting GOTR participant changes that align with the curriculum and describe the evaluation process. Therefore, this study presents an evaluation of GOTR consisting of three main processes: curriculum content analysis and stakeholder focus groups ($N = 11$) to identify key outcomes of the program; community-based participatory research to collaborate with program personnel to further identify important outcomes; and the design and pilot testing of an instrument ($N = 104$) for assessing changes in the theoretically grounded outcomes over time. Findings demonstrated a positive collaborative process that led to important information to be used for an impact evaluation of Girls on the Run and for future evaluation development efforts for physical activity based positive youth development.

Girls on the Run (GOTR) is a voluntary after-school physical activity based positive youth development (PA PYD) program for girls in 3rd through 5th grade. The GOTR mission is to "inspire girls to be joyful, healthy and confident using a fun, experience-based curriculum which creatively integrates running" (Girls on the Run, 2015). The positive youth development (PYD) philosophy is grounded in the assumption that all youth have the potential for positive growth and well-being, but that potential requires intentionally designed opportunities to develop strengths and assets (Holt, 2008). Based in developmental systems theory (Ford & Lerner, 1992), PYD assumes a positive potentiality in all youth to exhibit positive outcomes (Benson, Scales, Hamilton, & Sesma, 2006). Therefore, PYD programs should be designed to emphasize the development of personal and social assets rather than focusing on reducing deficits or problematic behaviors (National Research Council and Institute of Medicine (NRCIM), 2002). PYD programs for pre-adolescent girls are ideally situated to develop the skills and assets to protect against negative developmental outcomes. The GOTR program encompasses this philosophy and incorporates regular physical training, an emphasis on personal health and well-being, and a physical challenge in completing a 5 K. This program therefore draws on the potential of the PYD approach and the unique opportunities in the physical activity context to integrate physical, emotional, and social assets for at-risk populations (Holt, 2008).

The physical activity context offers many benefits including, but not limited to, physical and mental health (Physical Activity Guidelines Advisory Committee, 2008). Of especial interest is the opportunity to build social skills in physical activity settings by providing a context that requires conflict resolution, cooperation, team building, goal setting, and leadership (Fraser-Thomas, Côté, & Deakin, 2005; Hellison, Martinek, Walsh, & Holt, 2008; Holt, 2008). However, it should be noted that this rich social context can also lead to negative outcomes such as lower moral reasoning, acceptance of aggressive behaviors, and even delinquency via an over-emphasis on social comparison (Gardner, Roth, & Brooks-Gunn, 2009; Weiss, Smith, & Stuntz, 2008; Coakley, 2014). Therefore, it is vital that programs are intentionally structured to facilitate positive assets and outcomes (Danish, Forneris, Hodge, & Heke, 2004; Gould & Carson, 2010; Gould, Flett, & Lauer, 2012; Larson, Hansen, & Moneta, 2006) and that the participant experience is carefully evaluated to document whether program outcomes match the intentional structure of the curriculum. The GOTR program is a good example of a well-structured PA PYD program in the physical activity context.

It is becoming increasingly important to document PYD effectiveness. There is strong encouragement for program evaluation data to support, improve, and ultimately sustain PYD

* Corresponding author at: Washington State University, P.O. Box 641410, Pullman, WA 99164-1410, United States.
E-mail address: sullrich@wsu.edu (S. Ullrich-French).

http://dx.doi.org/10.1016/j.evalprogplan.2015.12.002
0149-7189/

programs (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2002). Unfortunately, rigorous and carefully planned program evaluation data is often lacking (Catalano et al., 2002; Brooks-Gunn, 2003) and indeed Roth and Brooks-Gunn have lamented, "the enthusiasm for youth development programs far outstrips the empirical evidence of their effectiveness" (2003, pp. 171–172). There are recent examples of high quality PYD evaluation efforts, such as the general PYD *Positive Action* program (Beets et al., 2009) and the PA PYD *First Tee* program (Weiss, Stuntz, Bhalla, Bolter, & Price, 2013). However, there continues to be limited data documenting programs' effectiveness, especially for programs in the sport and physical activity context. And further, there is limited information describing the process of evaluation development in this context. Following this, trend there is limited evaluation data that directly evaluates key elements of GOTR.

It is important to distinguish rigorous, experimental research efforts requiring a high degree of control, from developmental program evaluation requiring a more flexible and user focused emphasis (Patton, 2011). The scope and rigor of program evaluation will and should vary from program to program based on individual program goals and resources (Eccles & Gootman, 2002). Indeed, a focus on rigorous outcome evaluations may be a barrier for program evaluation efforts. Instead, evaluations that are flexible and useful for programs can also serve an important function in our understanding of PYD programs (Patton, 2011). Community-based participatory research (CBPR) perspectives have challenged a traditional top-down approach (i.e., research driven) and advocated for more participation from communities/programs to bridge the gap between research and practice (Wallerstein & Duran, 2010). Critical issues raised by CBPR include flexibility, collaborative partnerships, recognition and value of community priorities, trust and mutual respect (Israel, Eng, Schulz, & Parker, 2012). An important aspect of building capacity within programs to conduct evaluations is to create a culture that values (Arnold & Cater, 2011) and reduces resistance to evaluation (Patton, 2011), something that may be in part accomplished through involvement of stakeholders in the evaluation development process (Patton, 2011). We therefore provide a description of the development process and initial pilot testing of a PA PYD program evaluation using an approach that emphasized CBPR values (Israel et al., 2012).

1. The program

GOTR programs are organized within a geographically-based council. Councils are managed by an executive director and program decisions are made by a Board of Directors comprised of current and past coaches, and community members, and guided by the GOTR International administration guidelines and staff recommendations. Girls volunteered to participate in the after-school program in teams comprised of approximately 15 girls and two coaches. The program was conducted at elementary schools, or at a close, local community center. The program met twice a week over a 10 week season, completing a total of 20 standardized and scripted lessons that were facilitated by a trained GOTR coach. Coaches were volunteers who applied and completed a background check. Coaches also attended a mandatory training session prior to each season in which they are instructed in the program philosophy, behavioral management, injury prevention in youth, program curriculum facilitation, and how to get support when coaching. Council board and staff, girls, their parents, and coaches were all identified as important stakeholders for this evaluation.

There are three versions of the curriculum. At the time of this project, the second curriculum was implemented and described below. Coaches used a manual with specific lesson plans, including standardized format, materials, and scripts. The lessons followed a common format, beginning with a "Getting on Board" activity to introduce and discuss the daily theme. A "Warm Up" activity incorporated physical movement that integrated the day's theme. The "Workout" also integrated the daily theme while incorporating running or physical conditioning. The lessons ended with a "Wrap Up" summarizing the activities and theme of the day and giving "energy awards" for those demonstrating a positive attitude. Following each section processing questions allowed the coaches and the girls to discuss and apply the theme and activities. The program's curriculum was divided into three sections (see Table 1 for a sample lesson from each section). The first addressed self-care by enhancing positive identity by way of understanding and valuing oneself. The second addressed connectedness with the goal of supporting healthy relationships, such as ways to initiate and maintain relationships and valuing of others. The third section emphasized empowerment through connection with community to both celebrate and share each individual's strengths by setting and striving towards their goals, as well as connecting with and shaping their context. Girls explore the benefits of connecting their own world with others by sharing their strengths through a team-designed and implemented community service project. At the end of the season, girls, together with an adult running buddy, completed a 5 k running event that is not timed. This 5 k was framed as a run/walk/skip or hop to give the girls a framework for setting and achieving their own individual goals and an opportunity to experience a sense of personal accomplishment. All participants wear a #1 running bib number and the goal is to complete the event. Girls were encouraged through all activities to make their own choices or thoughts and actions providing an inner strength of self. The content and structure of GOTR thus align with the foundational PYD philosophy. From the first program started in Charlotte, North Carolina by Molly Barker in 1996, GOTR has grown to now serve girls in more than 200 cities in both the United States and Canada (Girls on the Run, 2015).

Table 1
Curriculum content examples.

	Example lesson	Lesson description
Self-care (8 lessons)	Lesson 2: plugging into the Girls on the Run cord	Getting on board: introduce and practice skills to identify and stay positive, such as visualization. Warm up: physical movement around the space to exchange of positive words with each other. Workout: running laps with each lap completed each girl lists an example of strategies used to maintain positivity. Processing question: "*Do you feel plugged into the positive Girls on the Run cord right now? How do you know? What does it really feel like*?"
Connection (7 lessons)	Lesson 10: standing up to peer pressure	Getting on board: introduce the concept of peer pressure and generate examples of peer pressure situations. Warm up: practice "stop; breathe; listen; respond" skill in active practice scenarios. Workout: self-selected goal of number of laps completed to help listen to one's inner voice and to "choose the right pace for me". Processing question: "*Are there any situations in your life right now that you think you could use this strategy*?"
Empowerment (5 lessons)	Lesson 19: learning about community	Getting on board: introduce and discuss the concept of community and do a cooperative trust activity. Warm up: relay run where successively everyone completes the same activity. Workout: Set a group goal of total laps completed where everyone contributes to their own ability toward the group goal. Processing question: "*How did this activity demonstrate community*?"

Notes: examples drawn from version two of the curriculum, which was used at the time of this study.

2. Research on GOTR

Evaluation evidence about this program is limited in scope. The initial research on GOTR examined the outcomes of self-esteem, body size dis-satisfaction, eating attitudes and behaviors, physical activity behavior, and commitment towards physical activity at pre- and post-program using non-random survey designs. Findings revealed significant increases in self-esteem, positive eating attitudes and behaviors, and body size satisfaction (DeBate & Thompson, 2005), physical activity commitment (DeBate, Zhang, & Thompson, 2007) and physical activity behavior (DeBate, Pettee Gabriel, Zwald, Huberty, & Zhang, 2009). More recent research examining self-efficacy for physical activity and physical activity behavior (Bean, Miller, Mazzeo, & Fries, 2012) found increases in positive outcomes to be sustained up to 3 months post-program. Additionally, researchers have investigated body image outcomes (Rauscher, Kauer, & Wilson, 2013; Racine, DeBate, Gabriel, & High, 2011) and self-esteem related to appearance and body image (Sifers & Shea, 2013). Collectively, these studies provide initial evidence of positive outcomes associated with participation in GOTR, but utilized a limited assessment of the curriculum or description of the evaluation development process.

Although there is initial supporting evidence for positive outcomes associated with the program, there appears to be lack of the evidence that assesses the program's goals and content beyond self-esteem and body image, which only comprise a part of the program content. Additionally, current research fails to employ a theoretical mechanism to help explain changes in participants. Both of these are especially important for detecting changes that are aligned with program philosophy and content (Linnan & Steckler, 2002; Patton 2011). For example, the work of Iachini, Beets, Ball, & Lohman (2014) represents an important step by conducting a process evaluation to assess program implementation as it relates to program outcomes. What is still lacking is a clearly articulated theoretical framework for explaining mechanisms for how observed changes occur. Inclusion of evaluation tools such as a logic model (Fig. 1) would also improve the evaluation process regarding GOTR. Logic models are useful evaluation tools for focusing, organizing, and interpreting the evaluation data (Cooksy, Gill, & Kelly, 2001) and would clarify the connection between the program's resources and activities to its desired outcomes. Additional research is needed to address the absence of a theoretical mechanism to explain change over time in curriculum based outcomes that also includes stakeholders in the process.

Previous research has identified girls' risk regarding lack of identity, lack of connectedness, and feelings of powerlessness (DeBate et al., 2007). These needs are tied to the three curriculum themes of the self (to capture issues of identity), connectedness, and empowerment. Though these themes have been used to guide the GOTR curriculum, they have not been well-articulated within a theory to help explain the mechanisms of change. Self-Determination Theory (STD: Ryan & Deci, 2000) provides a well-supported theoretical framework that conceptualizes how the social context provides a foundation for positive development and well-being. A core social contextual factor described in SDT that supports wellbeing and optimal functioning is autonomy support. The SDT based construct of autonomy support reflects participants'

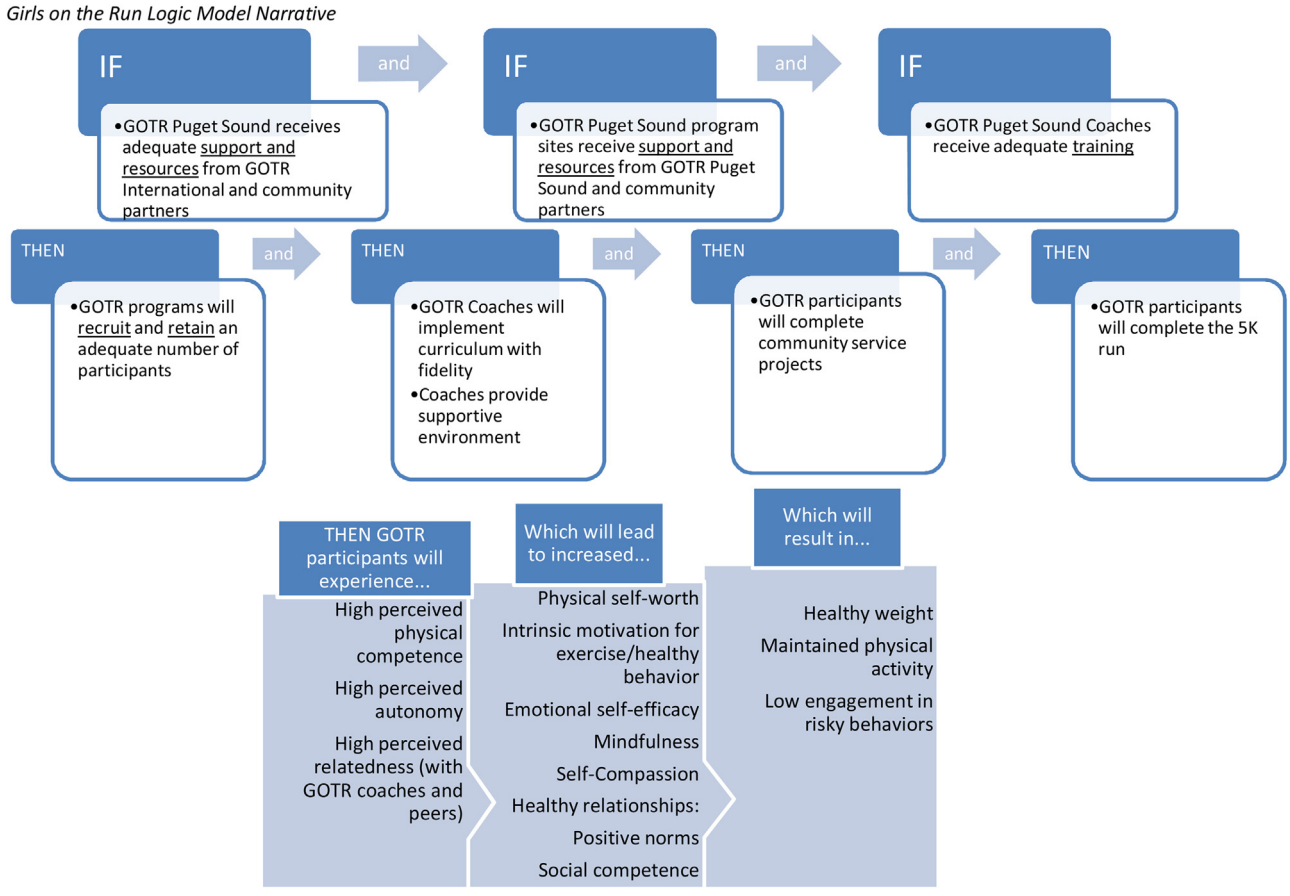

Fig. 1. Girls on the Run logic model narrative.

perceptions of how those in authority create a context in which one has positive feelings about their abilities, empowerment to make their own choices, and a feeling of social connection (Ryan & Deci, 2000) and aligns directly with how the GOTR program is structured and how coaches are trained to work with girls in an encouraging and supportive manner. Thus, the theory provides a useful framework for effectively evaluating GOTR. Using this theoretically grounded perception of the degree to which the context is autonomy supportive could help explain variation in positive outcomes.

The purpose of this study was to describe the development process and pilot testing of a quantitative outcome evaluation instrument for GOTR. We specifically aimed to use valuable stakeholder input to develop an instrument that assesses outcomes aligned with the GOTR curriculum and includes an assessment of a theoretical mechanism to help explain changes in these outcomes. The ultimate goal was to develop a practical, yet psychometrically sound quantitative instrument for this program to document effectiveness utilizing key elements of CBPR. In order to achieve this goal, we begin the process by collecting rich qualitative data to provide context for developing and then testing the instrument, which allowed us to ground instrument development in the experiences of key GOTR stakeholders (e.g., participants, parents, and GOTR personnel) (Bamberger, Rugh, & Mabry, 2012; Rowan & Wulff, 2007). This process was implemented to verify the theoretical framework from which we were working with stakeholders and to make the development of items more transparent, which is important for ensuring validity of the instrument (Rowan & Wulff, 2007). We used three-steps to develop the instrument and then conducted a quantitative pilot study using the instrument. The first step was to systematically review the curriculum and identify constructs that align with the program goals, philosophy, and content. Second, through focus groups with coaches, parents, and participants from one GOTR council, we gathered stakeholder feedback to guide a prioritization of an evaluation instrument to assess GOTR that was also aligned with SDT. Third, we created a logic model to guide the final instrument development. Finally, we conducted a pilot study with a select subsample of teams from the same GOTR council to test the initial evaluation instrument at pre- and post-program.

3. Three step instrument development process

3.1. Step one: program curriculum review methods

Three academic researchers conducted an in-depth review of the GOTR program and curriculum to aid in the identification of relevant outcomes that reflect changes that are expected to occur based on the goals and objectives of the program. We first independently reviewed the curriculum structure, mission, philosophy, and goals of GOTR to establish an understanding of the program goals. Next, we conducted a review of the curriculum content to identify specific content objectives of the program. Reviewers met to discuss the independent perspectives on the curriculum and through group consensus, identify measurable constructs that align with the curriculum.

3.2. Step one: program curriculum review results

As a group, the academic researchers generated a list of measurable constructs that were relevant to the program and that had both empirical and theoretical grounding. The proposed constructs were selected to align with curriculum and SDT and appear in Table 2 with a conceptual description and example of relevance to GOTR goals and objectives. Consensus agreement was obtained for the final list of constructs.

3.3. Step two: focus groups methods

Focus groups were conducted to gather stakeholder insight into the GOTR experience from coach, participant, and parent perspectives. Participants for focus groups were solicited by the GOTR council's staff by mailing an invitation to participate in focus groups to their GOTR mailing list. Focus groups were conducted solely by researchers in community locations and no GOTR staff were present during the discussions. Due to practical constraints, three focus groups were conducted. The first focus group was with 5 of the 96 coaches in the regional council. Three coaches were in their 20s, while a fourth coach was 44. Coach experience ranged from one to four seasons of past coaching. The second focus group was with GOTR participants ($N = 4$) from the region, which had a total of 400 girls participating at that time. Three girls were 11 and one was 10 years old. All but one girl had participated one previous season; the fourth girl had three previous seasons of experience with GOTR. The third focus group was with parents of previous GOTR participants ($N = 4$) and included one father and three mothers. Focus groups lasted between 30 and 60 min.

3.3.1. Focus group guide

The coach focus group guide consisted of three primary sections. We asked coaches (1) "*What are the main messages that coaches want the girls to learn in GOTR?*" and (2) "*Do the girls 'get' these messages* ?" and "*How do you know?*". After introducing each of the proposed outcome constructs that were identified by the research team following the curriculum review we asked "*How relevant is each outcome to GOTR?* " and "*What do you teach or model that relates to each outcome?*". We also asked "*What do you think are the most important parts of the GOTR curriculum?*" and "*How can we best evaluate GOTR?*" to gather insight into priority areas of the curriculum and to determine the appropriate length of time available for the evaluation, the training and support needed to administer the evaluation, and anticipated barriers. The parent focus group guide consisted of four questions to tap into the parent perspective of the program: (1) "*What are the main lessons that girls learn in GOTR?*", (2) "*What changes have you seen in your daughters as a result of GOTR?*", (3) "*How do you know?*", and (4) "*Is there anything you would change about GOTR?*" . Finally, the participant (girls) focus group guide consisted of four questions to gather participant perspectives about the program: (1) "*What did you like the most about being in GOTR?*", (2) "*Is there anything you would change about GOTR?*", (3) "*What are some of the things you learned in GOTR (about self, others, running or taking care of yourself)*?", and (4) "*Did anything about you change because you were in GOTR (attitudes, thoughts, feelings, or behaviors)*?".

3.3.2. Focus group data analysis

Each focus group was audio recorded and the recordings were reviewed in-depth independently by two researchers. Initial themes were identified independently and a consensus meeting was conducted to discuss themes. Audio recordings were reviewed again using a constant comparison method (Patton, 2015). Clear themes emerged and consensus was reached among the coding team.

3.4. Step two: focus groups results

Overall, coaches reported many positive experiences and attributes associated with GOTR participation. This was especially true for girls who participate multiple times, and for the concepts which are repeated themes across the season (e.g., positive self-talk). Themes of self-awareness and positive self-perceptions, awareness of others around them, and a sense of community were all identified by coaches as outcomes they saw from participants of

GOTR. Coaches reflected on girls' ability to address the challenges and issues raised, noting how well they responded to and picked up on complex themes, and then demonstrated the concepts such as conflict management and positive self-talk. One coach shared,

> "I think to find positives about themselves is really important, like whatever that is, a lot of girls at that age have trouble saying even one thing that's positive about them, and so just to start building that confidence, that is something that I like to see in the program."

When presented with a list of relevant constructs, coaches could connect each construct presented with GOTR experience; however, though self-awareness was identified as a key theme, there was concern over measuring mindfulness among the girls. Additionally, though they identified increase confidence and positivity among the girls as an important theme, they expressed concern about the feasibility of measuring intrinsic motivation. Finally, coaches had significant concern over their ability to administer an evaluation that lasted more than 15 min. They requested specific instructions and thought evaluation administration could be included in the coach training. The overall themes that emerged from coaches regarding GOTR goals and objectives were increased confidence and positivity, importance of understanding peers and community, self-awareness, and girls' ability to address the issues presented in the curriculum.

Parents were enthusiastic about their daughters' confidence and motivation, which was identified as one key theme. Parents discussed their daughters' positive attitudes, focus, and maturity, their stronger sense of responsibility and effort, and their ability to take on the challenge of the 5 K run. Another theme that emerged was parents' perception of various positive social outcomes experienced by their daughters. They noted outcomes such as teamwork, bonding with running buddies, and bringing lessons home to the family. One parent said,

> "I'm hoping that she's going to be more willing to keep trying at things that are difficult for her . . . it was teaching her if she works hard enough she can be successful at something that she is really struggling with."

Additionally, parents discussed the healthy activity and nutrition changes their daughters were making, noting that their daughters were making more healthy choices overall.

Results from the focus group with GOTR participants revealed that girls identified many social benefits to participating in GOTR. The ability to deal with their emotions and conflict with peers was one main theme. For example, one girl noted that she uses the strategy she learned in GOTR of "Stop, breathe, listen, and respond" to help deal with drama from friends. Girls identified important new skills and attributes, such as meeting new friends and learning how other people should react. The girls also identified that they learned to speak out, the value of friendship, and that when peers talk about them it is just one opinion. One girl stated,

> "Often times people would say something about me and I would like take it super seriously and I like have this grudge for like

Table 2
Curriculum and construct mapping.

Construct	Conceptual description	Program/curriculum content
Autonomy support	The extent to which girls feel like their GOTR coaches encourage positive feelings about their abilities; empower them to make their own choices; and help them to feel socially connected to the group (Ryan & Deci, 2000)	• Coach training to create positive social environment • Lesson structure includes discussion time and sharing • Team building activities • Focus on community
Emotional self-efficacy	Feelings of being competent at managing emotional experiences. Children with high emotional self-efficacy feel confident that they can cope effectively with negative emotions and thoughts using positive self-talk and self-soothing strategies (Caprara et al., 2008)	• Lesson 6: being emotional is healthy! • Lesson 4: positive self-talk and why I choose it!
Mindfulness	A psychological quality that involves bringing one's complete attention to the present experience on a moment-to-moment basis. Paying attention to one's emotions or other experience in a particular way, intentionally or on purpose, in the present moment, and non-judgmentally (Kabat-Zinn, 2003)	Lesson 8: centering: the importance of slowing down
Intrinsic motivation	Reasons or motives for doing something "for its own sake." When a person's motivation does not come from outside pressures (Ryan & Deci, 2000)	• Intertwined throughout curriculum • Program goal of long-term positive behaviors
Positive social norms	Beliefs about how acceptable or unacceptable bullying behaviors are (Siu, Shek & Law, 2012)	• Lesson 12: standing up for myself; • Lesson 13: gossiping hurts everyone • Lesson 14: bullying
Relatedness	One's feelings of belonging, closeness, and social connections with coaches and peers in GOTR. It is a general sense of acceptance and connection to the social group (Ryan & Deci, 2000)	• Coach training creating positive social environment • Core theme of connectedness • Team building activities • Focus on community
Autonomy	When someone feels a sense of having choice and control in what they do they are volitional, or autonomous (Ryan & Deci, 2000)	Lesson 12: standing up for myself
Competence	When one feels that they can be effective in achieving their goals they feel competent. It is feeling that your actions can produce the desired outcomes (Ryan & Deci, 2000)	• 5 K goal • "That every girl, can embrace who she is, can define who she wants to be, can rise to any challenge, can change the world." (Girls on the Run, 2015)
Physical self-worth	This is typically reflected in one's feelings of happiness, satisfaction, pride, respect, and confidence in their overall physical self. This construct is linked closely with self-esteem as well as ability and appearance perceptions (Harter, 1985)	• Focus on positive self-perceptions across curriculum • Relevant to the physical activity context of running and to the context of pre-adolescent girls
Self-compassion	Three components: (1) self-kindness versus self-judgment, (2) a sense of common humanity versus isolation, and (3) mindfulness versus rumination (Neff, 2009)	• Core themes of self-care and community • Lesson 4: positive self-talk and why I choose it!
Social competence	A socially competent child is one who regulates their emotions well, can communicate effectively with others, and engages in prosocial behavior (e.g., helping, sharing, including others) (Semrud-Clikeman, 2007)	• Lessons 12: standing up for myself • Lesson 13: gossiping hurts everyone

two weeks about them, but now it's sort of like, that's their opinion, and that's not what I think about myself, so, I just want to like let it go."

The theme of improved confidence in their own abilities emerged when girls were asked how they have changed after participating in GOTR. They identified physical improvements, making more friends, helping others, speaking up and being brave (especially in helping others), teamwork, and learning strategies to manage conflict. One girl recognized the benefits of persistence and making physical improvements, saying, "I just want to keep going, to get better and better at running. I used to hate running, but now, I'm just getting better and better, so I'm starting to like it better." Finally, the girls mentioned the power of encouragement from their running buddies. They felt supported and encouraged to take on the challenge of running. A summary of focus group themes appears in Table 3.

3.5. Step three: evaluation instrument development methods

Based on the curriculum content review and focus groups we created a logic model (Fig. 1) to help guide recommendations for outcomes to be included in a pilot evaluation instrument. The formation of a logic model for GOTR aided in the identification of appropriate outcomes to assess and served as a communication tool for stakeholder evaluation buy-in and capacity building (see Wholey, 1994, for more information on logic models). In addition to the previous steps, we also considered the developmental literature, self-determination theory, and evaluations of other PYD programs to arrive at a list of potential program outcomes. Emotional self-efficacy, positive norms, mindfulness, physical self-worth, physical activity, intrinsic motivation for physical activity, social competence, and self-esteem were considered as relevant outcomes of the program while perceived competence, perceived relatedness in GOTR, perceived autonomy in GOTR, and autonomy support in GOTR were considered as possible mechanisms, or theoretical predictors of variability in the program outcomes.

An initial proposal including the logic model, constructs listed above, and outlining a coach-administered pre- and post- program evaluation was sent to key GOTR-Puget Sound (GOTRPS) stakeholders for feedback. These stakeholders included the executive director, program manager, and the program committee chair, who then forwarded it on to the GOTRPS program committee, made up of past and present coaches and members of the council's board of directors. Stakeholders reviewed the proposal before meeting with researchers in person to discuss the proposal and what could be gained through completing an evaluation. The proposal defined the possible constructs, provided example items demonstrating how they could be measured, and an example of how to administer the evaluation by integrating it into the curriculum. The draft was intended to present the ideas to the stakeholders and through discussion learn their perspective on the viability of the approach and to help build the value of evaluation and ultimately build internal capacity.

Using a collaborative approach that engaged GOTR program stakeholders for feedback (Butterfoss, Francisco, & Capwell, 2001; Brandon & Fukunaga, 2014), discussion and consideration of empirical, theoretical, and practical constraints led to some initial eliminations to the list of outcomes. The construct of self-compassion was considered too challenging to measure with girls as young as 8 years old and was therefore eliminated. Although ideally all three psychological needs based on SDT would be examined, we determined that perceived competence was the most important psychological need to assess in this context. We also included autonomy support as a primary mechanism by which to explain girls' perceptions of a positive environment which could be used to help predict changes across program participation.

The GOTR staff and program committee were given the opportunity to provide anonymous feedback to help prioritize of the remaining outcomes given the limited time allotted to administering the evaluation instrument during program time. An online survey asked these stakeholders to rate the importance and priority of each of the remaining potential outcome variables. Open discussion between researchers and program leaders using all the collected evidence guided consensus on what would be included on the evaluation instrument.

3.6. Step three: evaluation instrument development results

The two outcomes with the highest priority rankings were the girls' sense of emotional self-efficacy and their endorsement of prosocial norms. There was consensus that it would be important to measure both of these constructs. There was also consensus that GOTR positively affects girls' perceived physical self-worth and that was also an important outcome to be included in an evaluation. Social competence was also rated by the committee as relevant and appeared to have a middle priority ranking. However, although a relevant construct to measure, it was not considered as important as some of the other constructs by the group. Therefore, it was not included in the pilot instrument. Finally, intrinsic motivation for physical activity and mindfulness where the least favorably ranked outcomes, generally appearing as somewhat relevant and receiving the lowest rankings on the priority list. While this could be related to the unfamiliarity with the terms and the constructs they represent, these constructs were not included in the pilot instrument due to practical constraints

Table 3
Focus group themes and potential outcomes.

Focus group themes				
Coaches	Increased confidence and positivity	Importance of understanding peers and community	Self-awareness	Girls' ability to address the challenges/issues raised
Parents	Confidence and motivation	Positive social outcomes		Healthy changes in activity and nutrition
Girls	Gaining confidence in own abilities	• Dealing with peers and emotions • Power of the running buddy for encouragement		
Potential outcomes	• Perceived competence • Physical self-worth • Self-esteem • Motivation	• Relatedness • Autonomy support • Positive norms • Emotional efficacy	Mindfulness	Physical activity/ nutrition behaviors

and the importance of fully including stakeholders as part of the evaluation process (Butterfoss et al., 2001; Brandon and Fukunaga, 2014). However, we believe that these would be interesting to measure in future research as they may be mediators of the long-term effects of GOTR.

The pilot instrument therefore was comprised of emotional self-efficacy, prosocial norms, and physical self-worth to capture the goals and content of the GOTR program. These constructs address the immediate or short-term outcomes reflected in the logic model (Fig. 1). Autonomy support and perceived competence served as theoretical and empirical mechanisms for explaining changes in the program outcomes. Given programming staff concern regarding the attention span of the girls and the complexity of some of the survey items we integrated the survey into lesson 2 of the curriculum with a detailed script and explanation for the survey procedures. We included physical activities to be incorporated within the survey (short breaks), as well as structured discussion questions to process the items on the survey ("*what do you think these questions are about?*"), which aligned with the structure and method of GOTR lessons. The goal was to provide a survey experience that was fully integrated with the curriculum to help coaches administer and participants complete the evaluation.

4. Quantitative pilot study

4.1. Methods

We conducted the pilot study with a sub-set of teams participating in one season of GOTR. The goal was to test the administration procedures, examine psychometric properties of the evaluation instrument, and collect initial data assessing program outcomes. Based on the program outcome assessment we hypothesized that (1) emotional self-efficacy, prosocial norms, perceived competence, and physical self-worth would increase from pre- to post-program, and (2) based on SDT (Ryan & Deci, 2000), higher perceptions of autonomy support would positively predict increases in emotional self-efficacy, prosocial norms, perceived competence, and physical self-worth.

4.2. Procedures

In the fall season of 2012, seven GOTR teams were selected by program staff to participate in a pilot evaluation. The teams were purposefully selected to provide diversity in team location and coach's experience with the curriculum. This evaluation received an Institutional Review Board Certification of Exemption from the first author's institution before collecting data. At registration, all participants had a parent or guardian provide their permission for girls to participate in the evaluation survey. A training video was made to introduce the coaches to the survey instrument and to provide demonstration and instruction of the survey administration procedures. Additionally, coaches were provided a standardized script for the survey administration. These steps were taken to help coaches feel more comfortable and capable in conducting the evaluation. Coaches administered the pre-program survey during the second lesson. We observed the pre-program survey administration with three of the teams. Although the intended time allotted for the survey was 15 min, the teams we observed ranged from approximately 20–40 min for survey administration. The survey administration was complex and some of the item structures were clearly difficult to understand for both coaches and participants. Post-program surveys were administered by coaches in lessons 23 or 24 following the 5 K running event and took approximately 20–30 min. Coaches were asked to provide feedback on the evaluation process to inform refinement of the administration procedures.

4.3. Participants

Participants in the pilot evaluation included 104 girls in 3–6th grade and represented 7 different elementary schools located within King County, Washington. Ages ranged from 7 to 11 ($M = 8.76$, SD = 0.90). About half ($n = 48$) had never participated in GOTR, 14 participated once before, and 10 participated two or more times previously. Demographic information was obtained from program registration records and only 47 participants reported race and ethnicity (Caucasian = 24, African American = 4, American Indian = 1, Asian = 5, Hispanic = 5, and Multi-Racial = 8).

4.4. Measures

Emotional self-efficacy was assessed using an adapted form of the Self-Efficacy Questionnaire for Children (SEQ-C, Muris, 2001). This set of questions contained 7 statements regarding how well participants felt they could handle their emotions. Items were modified from "I am good at. . ' to "I can. .' and items were all positively written (one item was changed from 'I don't. . to "I spend a lot of time. .'). Utilizing a 4-point ordinal scale, 1 = NO!, 2 = No, 3 = Yes, 4 = YES!, girls were asked to circle the word that told how much they agreed with each statement. Modifications were based on developmental literature and stakeholder feedback. An example is 'I can cheer myself up when something bad has happened.' Evidence of reliability and validity has been documented for younger and older adolescents (Muris, 2001; Suldo and Shaffer, 2007). Internal consistency reliability for the current study was acceptable at pre-test ($\alpha = 0.70$) and post-test ($\alpha = 0.73$).

Prosocial norms were assessed using an adapted measure from Huesmann and Guerra (1997), Werner and Nixon (2005), and Werner and Hill (2010) containing 5 statements to measure what participants thought about how to treat other people. Modifications were minor to improve wording (e.g., replacing "it is my responsibility. . " to "I should. ."). One item was removed as it was perceived to be redundant. Utilizing a 4-point ordinal scale, 1 = NO!, 2 = No, 3 = Yes, 4 = YES!, girls were asked to circle the word that told how much they agreed with each statement. An example is "It is OK to keep someone you don't like from sitting with your group of friends (e.g., at the lunch table). " Despite previous use of this measure in past research, internal consistency reliability was low at pre-test ($\alpha = 0.43$) and post-test ($\alpha = 0.61$).

Perceived competence for physical activity was assessed using a measure developed by Amorose (2003). Three statements were used to measure participant's perceptions about their own competence in physical activities. Utilizing a 4-point ordinal scale, 1 = NO!, 2 = No, 3 = Yes, 4 = YES!, girls were asked to circle the word that told how much they agreed with each statement. An example is "I am good at physical activities." Amorose (2003) provided evidence supporting reliability and validity of this measure with older adolescents. Internal consistency reliability was acceptable at pre-test ($\alpha = 0.76$) and post-test ($\alpha = 0.86$).

Physical self-worth was assessed using an adapted measure from Harter's Self-Perception Profile for Children (Harter, 1985). Three of the original six items were selected based on developmental appropriateness and examination of data on this measure with a similar age sample. This set of questions contained 3 sets of statements about kids' perceptions of themselves physically using a structured alternative format. Statements were worded to describe two different types of kids. Participants were asked to select what type of kid they were more like and then check if that was "sort of true" or "really true" for themselves. Really true indicated a higher response. An example is "Some kids are proud of

themselves physically BUT other kids do not have much to be proud of physically." Scores range from 1 to 4. Participants were instructed to only check on the side of the kid that was more like them. Although this measure has been used extensively and has demonstrated adequate psychometric properties (Harter, 1985), it is a challenging format to administer and the internal consistency reliability was low at pre-test ($\alpha = 0.59$) and posttest ($\alpha = 0.62$).

Autonomy support was assessed at post-program only using a shortened version of the measure from Standage, Duda and Ntoumanis' (2005) adaption of the Learning Climate Questionnaire (Williams & Deci, 1996). Seven of the items from this measure were selected based on developmental appropriateness, to capture the theoretical basis for autonomy support, and using data from a similar age sample to identify a concise yet reliable measure. This set of questions contained 7 statements to measure participant's perceptions of how supportive the GOTR coaches and running buddies were. Utilizing a 4-point ordinal scale, 1 = NO!, 2 = No, 3 = Yes, 4 = YES!, girls were asked to circle the word that told how much they agreed with each statement. An example is "I feel that at GOTR I was provided with choices and options." The full measure has been used successfully, demonstrating adequate reliability and validity in similar sample and context (e.g., McDonough, Ullrich-French, Anderson-Butcher, Amorose, & Riley, 2013). Internal consistency reliability was marginally acceptable ($\alpha = 0.68$).

4.5. Data analysis

Data screening for missing data, normality, and outliers was conducted. Next, descriptive statistics and internal consistency reliability were conducted. Bivariate correlations were also calculated. Paired *t*-tests were calculated to test for change in the outcome variables of emotional self-efficacy, prosocial norms, and physical self-worth. Hierarchical multiple regression was used to examine whether autonomy support predicted change in the outcome variables. Standardized residual scores were created for the outcome variables to establish scores representing change from pre- to post-program. Because GOTR is a physical activity-based program, perceived physical competence was entered in the first step along with age, past GOTR experience, and coach experience as covariates. The second step included post-program autonomy support perceptions. Change in explained variance was examined in addition to model fit for the second step.

4.6. Quantitative pilot study results

The missing data (13–35% cases) was not missing completely at random so we used listwise deletion. No cases were identified as outliers (i.e., $Z > 3$). Emotional self-efficacy and perceived competence for physical activities were reliably measured at both pre- and post-test (see Table 4). Prosocial norms and physical self-worth were not reliable and were not included in further analyses. Autonomy support was marginally reliable, but was retained in subsequent analyses. Descriptive statistics appear in Table 4. Emotional self-efficacy was moderate at pre- and moderately high at post-program. Perceived physical competence was relatively high at both pre- and post-program. Autonomy support was relatively high post-program.

To test for change from pre- to post-program in emotional self-efficacy and perceived physical competence, paired *t*-tests were conducted. There was a significant increase in emotional self-efficacy ($t = -3.52$, $p = 0.01$), but no change in perceived physical competence ($t = -0.35$, $p = 0.73$). The covariates of perceived physical competence, age, past GOTR experience, and coach experience were not significant in the hierarchical regression models and were dropped from the models. Therefore, simple regression was conducted with autonomy support predicting residual change in emotional self-efficacy and perceived physical competence. Residual change in emotional self-efficacy was significantly and positively predicted (F (1, 64) = 21.58, $p < 0.01$, $R^2 = 0.25$) by autonomy support ($\beta = 0.50$, $p < 0.01$). Also, residual change in perceived physical competence was significantly and positively predicted (F (1, 66) = 4.72, $p < 0.03$, $R^2 = 0.07$) by autonomy support ($\beta = 0.26$, $p < 0.03$). Higher perceptions of autonomy support at post-program predicted increases in emotional self-efficacy and perceived physical competence.

Table 4
Descriptive Statistics.

	Variable	1	2	3	4	5
1	Emotional self-efficacy PRE					
2	Perceived physical competence PRE	0.45**				
3	Emotional self-efficacy POST	43**^	0.26*			
4	Perceived physical competence POST	0.27*	0.59**^	0.40*		
5	Autonomy support POST	0.32**	0.21	0.57**	0.31**	
	N	88	90	79	81	81
	Internal consistency reliability (α)	0.70	0.76	.73	0.86	0.68
	Mean	2.86	3.46	3.10	3.48	3.68
	SD	0.55	0.57	0.53	0.56	0.30

Notes: ^intraclass correlation.
*$p < .05$.
**$p < .01$.

5. Discussion

This study described the development of an outcome evaluation of the PYD program GOTR and initial pilot testing of the evaluation instrument. We describe collaboration with staff at one regional council administering the program, stakeholder focus groups, and a pilot study used to develop and implement a pilot program evaluation. Through an iterative consensus process we created a program logic model and initial outcome evaluation survey instrument that balanced practical and logistical constraints and stakeholder perspectives with measures with evidence of validity and reliability. The instrument included participants' perceptions of emotional self-efficacy, prosocial norms, physical competence, physical self-worth, and autonomy support.

Much of the existing evaluation research on GOTR examines the relationship between participation and body image variables, such as body size satisfaction, fat attitudes, and eating attitudes (DeBate & Thompson, 2005; DeBate et al., 2009; Racine et al., 2011; Pettee Gabriel, DeBate, High, & Racine, 2011; Martin, Waldron, McCabe, & Choi, 2009). The limited process evaluation research conducted on GOTR emphasizes its implementation through the use of implementation checklists (Iachini et al., 2014). While this is an important step in understanding the variability in delivery of GOTR and how that relates to its outcomes, our study provides an important resource in the development of a pilot evaluation to effectively and reliably measure outcomes grounded in the curriculum and stakeholder feedback. Using an interactive evaluation approach that encouraged collaboration with a number of GOTR stakeholders (King & Stevahn, 2012; Brandon & Fukunaga, 2014; Butterfoss et al., 2001), we created a pilot instrument that reflected the desired interests and outcomes of those involved. Focus groups revealed that stakeholders were less concerned about some of the constructs assessed by previous research (e.g., body image), and more interested in the social interactions of girls with their teammates and coaches, as well as what the girls think about their physical abilities (as opposed to physical appearance). This decision reflects the curriculum of GOTR, based on empowering the participants by reducing emphasis on appearance and emphasizing what girls can do.

The pilot test of this instrument resulted in reliable measurement of emotional self-efficacy and perceived physical competence, and marginally reliable assessment of autonomy support. Change was found for the outcome of perceived emotional self-efficacy, which reflects existing research on girls' self-efficacy in GOTR that found improvements up to three months after participation (Bean et al., 2012). Autonomy support from coaches and staff significantly predicted increases in emotional self-efficacy and perceived physical competence as expected based on theory and PYD literature (e.g., McDonough et al., 2013).

The modifications made to the measures, and especially the truncated versions of measures likely contributed to the low reliability in some of the measures. Additionally, the participants were quite young and there is limited psychometric validation support for many measures using this age group. The choices to minimize survey length to accommodate the coaches' concern about time allotted to evaluation resulted in the inability to analyze all the data due to unreliable measures. A second issue that arose was that the coaches reported difficulty with the survey administration procedures. Coaches generally reported that they felt the survey procedures were too complex and time consuming. These issues demonstrate the tradeoff between theory and practice (Patton, 2011). However, it should be noted that given the number of limitations of the sample, procedures, and measurement, the evaluation was able to provide some valuable information about changes that occur across a GOTR season.

Additionally, the process of creating the pilot evaluation provided valuable information to guide future administration procedures and the evaluation instrument itself. Because strong relationships were developed with stakeholders, we were able to have open discussions about the procedures, the resulting data, and the conclusions that could be made. Feedback from coaches, GOTR staff, and pilot evaluation results were integrated in developing recommendations for revising and refining the instrument and procedures for evaluation. This approach demonstrated a successful integration of both researcher and stakeholder perspectives (Butterfoss et al., 2001; Bryson, Patton, & Bowman, 2011).

5.1. Lessons learned

Our process resulted in specific recommendations for future evaluation. Based on observations of survey administration, coach feedback, the high percentage of missing data, and low reliability of several of the measures, we determined that the structured alternative format of the physical self-worth items was not a feasible format to ask these questions. We recommend using a consistent format for all response formats across measures for consistency and ease of administration. We also recommend including all 6 items from the original physical self-worth scale (while this study had reliability issues with the scale, it has been shown to work in other evaluations of both GOTR and Girls on Track, a comparable program for middle school-aged girls (Sifers & Shea, 2013)). Because coaches noted a lot of difficulty in the participants' understanding of the prosocial norms items and because of the very low reliability for this measure we do not recommend including this construct in future evaluations. Rather we recommend removing these items in favor of adding items to the other constructs measured (e.g., physical self-worth, autonomy support) to maintain a concise survey length while increasing measurement precision of other outcomes.

Although results suggest that autonomy support positively predicted both increases in emotional self-efficacy and perceived physical competence, the autonomy support measure was marginally reliable. It was determined that all of the measure's original 15 items should be used to confirm results of this study and provide stronger evidence for the results. This measure has strong evidence for reliability and validity, even with similar samples and context (e.g., McDonough et al., 2013) and therefore is believed to provide a good indicator of the participants' perceptions of the supportive social climate of GOTR.

Based on coach feedback we recommend using a more scripted and simplified approach to administering the survey rather than incorporating physical activities and processing discussion questions within the survey administration procedures. Finally, we recommend conducting a second pilot using the revision notes above on the instrument and procedures and also to include a larger sample.

The process we describe aligns with evaluation approaches that encourage stakeholder input (Bryson et al., 2011). Through the extensive involvement of stakeholders throughout the process we also utilized many core principles of CBPR (Israel et al., 2013). The value of establishing relationships creates a more real-world basis and sustainable approach to conducting evaluation work (Bamberger et al., 2012). The importance of incorporating practical and logistical considerations along with data collection rigor into the development of an evaluation is absolutely vital. We found the flexibility and mutual cooperation resulting from a community-academic-practice partnership resulted in an effective culture supporting the evaluation process (Israel et al., 2012).

This study provided a description of the outcome evaluation development process for a PYD program. Through a collaborative process with key program stakeholders we were able to create a working relationship that balanced program and evaluation needs and resulted in an effective partnership that fosters future collaborations. The process described used empirical and theoretical support, aligned with key features of evaluation and CBPR, and provided recommendations for an evaluation instrument that can be used as a basis for conducting future outcome and impact evaluations for this and similar PYD programs.

Acknowledgements

We would like to thank the Girls on the Run of Puget Sound Executive Director Kerin Brasch and their board of directors, Becky Antilla and the council's staff and program committee, coaches, participants and their parents for their positive enthusiasm, cooperation, and involvement throughout this project. Part of this project was the Masters thesis of Anna Montgomery partially funded by the Edward Graff Academic Excellence Award. A WSU College of Education Faculty Funding Award also provided partial funding. Thank you to Douglass Jackson for his feedback on this manuscript.

References

Amorose, A. J. (2003). Reflected appraisals and perceived importance of significant others' appraisals as predictors of college athletes' self-perceptions of competence. *Research Quarterly for Exercise and Sport*, *74*, 60–70.

Arnold, M. E., & Cater, M. (2011). From then to now: emerging directions for youth program evaluation. *Journal of Youth Development*, *6*, 83–94.

Bamberger, M., Rugh, J., & Mabry, L. (2012). *Real world evaluation: working under budget, time, data and political constraints*, 2nd ed. Thousand Oaks, CA: Sage.

Bean, M. K., Miller, S., Mazzeo, S. E., & Fries, E. A. (2012). Social cognitive factors associated with physical activity in elementary school girls. *American Journal of Health Behavior*, *36*(2), 265–274.

Beets, M.W., Flay, B.R., Vuchinich, S., Snyder, F.J., Acock, A., Li, K.K., . . . & Durlak, J. (2009). Use of a social and character development program to prevent substance use, violent behaviors, and sexual activity among elementary-school students in Hawaii. American Journal of Public Health, 99(8), 1438–1445

Benson, Scales, P. C., Hamilton, S. F., & Sesma, A. (2006). *Positive youth development: theory, research, and applications*. Hoboken, NJ: John Wiley & Sons, Inc..

Brandon, P. R., & Fukunaga, L. L. (2014). The state of the empirical research literature on stakeholder involvement in program evaluation. *American Journal of Evaluation*, *35*, 26–44.

Brooks-Gunn, J. (2003). Do you believe in magic? What we can expect from early childhood intervention programs. *Social Policy Report*, *17*, 3–15.
Bryson, J. M., Patton, M. Q., & Bowman, R. A. (2011). Working with evaluation stakeholders: a rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, *34*(1), 1–12.
Butterfoss, F. D., Francisco, V., & Capwell, E. M. (2001). Stakeholder participation in evaluation. *Health Promotion Practice*, *2*(2), 114–119.
Catalano, R. F., Berglund, M. L., Ryan, J. A. M., Lonczak, H. S., & Hawkins, J. D. (2002). Positive youth development in the United States: research findings on evaluations of positive youth development programs. *Prevention and Treatment*, *5*, article 15.
Coakley, J. (2014). *Sports in society: issues and controversies*, 11th ed. New York, NY: McGraw-Hill.
Cooksy, L. J., Gill, P., & Kelly, P. A. (2001). The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and Program Planning*, *24*(2), 119–128.
Danish, S., Forneris, T., Hodge, K., & Heke, I. (2004). Enhancing youth development through sport. *World Leisure Journal*, *46*(3), 38–49.
DeBate, R. D., & Thompson, S. H. (2005). Girls on the Run: improvements in self-esteem, body size satisfaction and eating attitudes/behaviors. *Eating and Weight Disorders*, *10*(1), 25–32.
DeBate, R. D., Pettee Gabriel, K., Zwald, M., Huberty, J., & Zhang, Y. (2009). Changes in psychosocial factors and physical activity frequency among third- to eighth-grade girls who participated in a developmentally focused youth sport program: a preliminary study. *The Journal of School Health*, *79*, 474–484.
DeBate, R., Zhang, Y., & Thompson, S. H. (2007). Changes in commitment to physical activity among 8-to-11 year old girls participating in a curriculum-based running program. *American Journal of Health Education*, *38*(5), 276–283.
Community programs to promote youth development. In J. S. Eccles, & J. A. Gootman (Eds.). Washington, DC: National Academy Press.
Ford, D. H., & Lerner, R. M. (1992). *Developmental systems theory: an integrative approach*. Thousand Oaks, CA: Sage.
Fraser-Thomas, J. L., Côté, J., & Deakin, J. (2005). Youth sport programs: an avenue to foster positive youth development. *Physical Education & Sport Pedagogy*, *10*(1), 19–40.
Gardner, M., Roth, J. L., & Brooks-Gunn, J. (2009). Can after-school programs help level the playing field for disadvantaged youth? *Equity Matters Research Review*, *4*, 1–39.
Girls on the Run. (2015). Retrieved from: http://www.girlsontherun.org/on April 5, 2015
Gould, D., & Carson, S. (2010). The relationship between perceived coaching behaviors and developmental benefits of high school sports participation. *Hellenic Journal of Psychology*, *7*, 298–314.
Gould, D., Flett, R., & Lauer, L. (2012). The relationship between psychosocial developmental and the sports climate experienced by underserved youth. *Psychology of Sport and Exercise*, *13*(1), 80–87.
Harter, S. (1985). Self-perception profile for children. *Unpublished manual*. Denver, CO: University of Denver.
Hellison, D., Martinek, T., Walsh, D., & Holt, N. (2008). Sport and responsible leadership among youth. In N. L. Holt (Ed.), *Positive youth development through sport* (pp. 49–60).New York, NY: Routledge.
Positive youth development through sport. In N. L. Holt (Ed.), (2008). New York, NY: Routledge.
Huesmann, R. L., & Guerra, N. G. (1997). Children's normative beliefs about aggression and aggressive behavior. *Journal of Personality and Social Psychology*, *72*, 408–419.
Iachini, A. L., Beets, M. W., Ball, A., & Lohman, M. (2014). Process evaluation of Girls on the Run: exploring implementation in a physical activity-based positive youth development program. *Evaluation and Program Planning*, *46*, 1–9.
Methods in community-based participatory research for health, In B. A. Israel, E. Eng, A. J. Schulz, & E. A. Parker (Eds.). 2nd ed. San Francisco, CA: Jossey-Bass Publishers.
King, J. A., & Stevahn, L. A. (2012). *Interactive evaluation practice: mastering the interpersonal dynamics of program evaluation*. Los Angeles, CA: Sage.
Larson, R. W., Hansen, D. M., & Moneta, G. (2006). Differing profiles of developmental experiences across types of organized youth activities. *Developmental Psychology*, *42*(5), 849–863.
Linnan, L., & Steckler, A. (2002). *Process evaluation for public health interventions and research*. San Francisco, CA: Jossey-Bass.
Martin, J. J., Waldron, J. J., McCabe, A., & Choi, Y. S. (2009). The impact of Girls on the Run on self-concept and fat attitudes. *Journal of Clinical Sports Psychology*, *3*, 127–138.
McDonough, M. H., Ullrich-French, S., Anderson-Butcher, D., Amorose, A. J., & Riley, A. (2013). Social responsibility among low-income youth in physical activity-based positive youth development programs: scale development and associations with social relationships. *Journal of Applied Sport Psychology*, *25*, 431–447. http://dx.doi.org/10.1080/10413200.2012.751563.
Muris, P. (2001). A brief questionnaire for measuring self-efficacy in youths. *Journal of Psychopathology and Behavioral Assessment*, *23*, 145–149.
National Research Council and Institute of Medicine (NRCIM) (2002). Community programs to promote youth development: committee on community-level programs for youth. In J. Eccles, & J. A. Gootman (Eds.), *Division of behavioral and social sciences and education, board on children, youth, and families*Washington, DC: National Academy Press.
Patton, M. Q. (2011). *Developmental evaluation: applying complexity concepts to enhance innovation and use*. New York, NY: The Guilford Press.
Patton, M. Q. (2015). *Qualitative research & evaluation methods: integrating theory and practice*, 4th ed. Thousand Oaks, CA: Sage.
Pettee Gabriel, K. K., DeBate, R. D., High, R. R., & Racine, E. F. (2011). Girls on the Run: a quasi-experimental evaluation of a developmentally focused youth sport program. *Journal of Physical Activity and Health*, *8*(2), 285–294.
Physical Activity Guidelines Advisory Committee Report (2008). *Physical activity guidelines advisory committee report*. Washington, DC, U.S: Department of Health and Human Services.
Racine, E. F., DeBate, R. D., Gabriel, K. P., & High, R. R. (2011). The relationship between media use and psychological and physical assets among third- to fifth-grade girls. *The Journal of School Health*, *81*, 749–755.
Rauscher, L., Kauer, K., & Wilson, B. D. M. (2013). The healthy body paradox: organizational and interactional influences on preadolescent girls' body image in Los Angeles. *Gender & Society*, *27*(2), 208–230.
Roth, J. L., & Brooks-Gunn, J. (2003). Youth development programs: risk, prevention, and policy. *Journal of Adolescent Health*, *32*, 170–182.
Rowan, N., & Wulff, D. (2007). Using qualitative methods to inform scale development. *The Qualitative Report*, *12*(3), 450–466.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78.
Sifers, S. K., & Shea, D. N. (2013). Evaluations of Girls on the Run/Girls on track to enhance self-esteem and well-being. *Journal of Clinical Sport Psychology*, *7*, 77–85.
Standage, M., Duda, J. L., & Ntoumanis, N. (2005). A test of self-determination theory in school physical education. *British Journal of Educational Psychology*, *75*, 411–433.
Suldo, S. M., & Shaffer, E. J. (2007). Evaluation of the self-efficacy questionnaire for children in two samples of American adolescents. *Journal of Psychoeducational Assessment*, *25*, 341–355.
Wallerstein, N., & Duran, B. (2010). Community-based participatory research contributions to intervention research: the intersection of science and practice to improve health equity. *American Journal of Public Health*, *100*(S1), S40–S46.
Weiss, M. R., Smith, A. L., & Stuntz, C. P. (2008). Moral development in sport and physical activity: theory, research, and intervention, In T. S. Horn (Ed.), *Advances in sport psychology* (pp. 187–210).3rd ed. Champaign, IL: Human Kinetics.
Weiss, M. R., Stuntz, C. P., Bhalla, J. A., Bolter, N. D., & Price, M. S. (2013). 'More than a game': impact of the first tee life skills programme on positive youth development: project introduction and year 1 findings. *Qualitative Research in Sport, Exercise and Health*, *5*, 214–244.
Werner, N. E., & Nixon, C. L. (2005). Normative beliefs and relational aggression: an investigation of the cognitive bases of adolescent aggressive behavior. *Journal of Youth and Adolescence*, *34*(3), 229–243. http://dx.doi.org/10.1007/s10964-005-4306-3.
Werner, N. E., & Hill, L. G. (2010). Individual and peer group normative beliefs about relational aggression. *Child Development*, *81*(3), 826–836. http://dx.doi.org/10.1111/j.1467-8624.2010.01436.x.
Wholey, J. S. (1994). Assessing the feasibility and likely usefulness of evaluation. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 15–39).San Francisco, CA: Jossey-Bass.
Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: a test of self-determination theory. *Journal of Personality and Social Psychology*, *70*, 767–779.

Sarah Ullrich-French is an Associate Professor of Kinesiology and Graduate Faculty in Prevention Science at Washington State University. Sarah received her M.S. (2003) and Ph.D. (2006) in Kinesiology, with an emphasis on Sport and Exercise Psychology from Purdue University. She has been at Washington State University since 2008. Sarah is an avid runner, she competed in collegiate cross-country, then moved to marathons, and ulta-marathons. She studies the positive potential of physical activity experiences on the well-being and development of youth.

Amy Cole is a doctoral candidate in the Individual Interdisciplinary Doctoral Program at Washington State University. Amy received her M.A. (2011) in Sport Management from WSU and has taught a variety of programming and theory courses in that field. Amy is passionate about physical activity and enjoys teaching a variety of group exercise classes. She is most interested in the design, implementation, and evaluation of physical activity programs that empower vulnerable populations.

Anna Montgomery is a certified child life specialist, graduated with her Masters in Human Development from Washington State University in 2013. This evaluation was conducted as part of Anna Montgomery's thesis research work. Working on this evaluation was a great fit for Anna, who loves running and supporting positive youth development. At the time of article submission, Anna was working in South Africa with Thembalethu Care Organization as an Orphan and Vulnerable Children Project Support Specialist, focusing on empowering children and caregivers to live lives that are filled with faith, hope and love.