Examining the Relationships between Dosage and Outcomes in Sport-Based Positive Youth Development

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Leeann Marie Lower

Graduate Program in Education: Physical Activities and Educational Services

The Ohio State University

2015

Dissertation Committee:

Dr. Brian Turner, Advisor

Dr. Donna Pastore

Dr. Dawn Anderson-Butcher
Abstract

Sport is a critical context for youth development (Riley & Anderson-Butcher, 2012; Weiss, 2008). Through engagement in sport-based positive youth development (PYD), vulnerable youth can experience relief from risk prevalent in their environment and develop assets to overcome adverse circumstances (Anderson-Butcher, Iachini, Riley, Wade-Mdivanian, Davis & Amorose, 2013; Petitpas, Cornelius, Van Raalte, & Jones, 2005). Organizational effectiveness of sport-based PYD programs is imperative due to the impact programs can have on youths’ successful transition to adulthood.

A sport-based PYD program is considered effective when exposure to an intentionally designed intervention, with theoretical justification, produces the desired program outcomes (Dusenbury, Brannigan, Falco, & Hansen, 2003; Law & Shek, 2011; Nation, Crusto, Wandersman, Kumpfer, Seybolt, Morrissey-Kane, & Davino, 2003). Research examining organizational effectiveness, within the context of sport-based PYD, can inform program planning, improve program practices, and enhance PYD. The purpose of the current study was to examine program dosage as a programmatic mechanism of PYD, to determine how program implementation and engagement influences youth development in the context of sport-based PYD.

To investigate program dosage as a programmatic mechanism of PYD, within the context of sport-based PYD, data were collected from 52 staff and 417 youth participants
engaged in a sport-based PYD summer program. Program staff completed self-assessments of their implementation of the program, including adherence to program structure, use of curriculum, and creation of a caring, mastery-oriented climate. Additionally, youth reported perceptions of program engagement (i.e., behavioral engagement, belonging, leader support), as well as social and athletic competencies (i.e., PYD indicators). Youth attendance was collected as secondary data, as a measure of quantity of program dosage.

Confirmatory factor analyses were conducted to test the measurement models, for which implementation of the program curriculum and youth attendance were found to be non-significant measures of quantity of program dosage. To test the relationships between program dosage (i.e., program implementation and engagement) and PYD, structural equation modeling (SEM) was employed. Results found program engagement to be a significant positive mediator of social and athletic competence development. Comparatively, program implementation was found to be a significant positive antecedent of program engagement in the athletic competence model and a non-significant positive antecedent of program engagement in the social competence model. Overall, the study demonstrated PYD in four of the five indicators, including general social competence, effort, teamwork, and athletic competence.

Collectively, the program implementation and engagement dimensions of program dosage were found to be significant programmatic mechanisms of PYD. The findings have implications for sport-based PYD programs and practitioners, specifically in the areas of program design and evaluation, as well as staff recruitment, training, and
support. Future research should seek to build from the current study to further explore how programs can effectively develop youth through sport.
Acknowledgments

I first want to acknowledge my Lord and Savior, Jesus Christ, who blesses me with renewed purpose and joy every day. He walks beside me and directs my path; His faithfulness and love are unceasing. Through Christ all things are possible; because of Christ I have hope for my future.

I would also like to acknowledge my family for their continued support and encouragement – Dave and Pat Lower, Dawn and Anthony Naranjo (plus Dri, Lexi, and Zeke), and Dani and Glen Armstrong. Through success and struggle, you believe in me. Thank you for your steadfast love and care; you mean the world to me.

To Abbie Lawson - we did it! I could not have gotten this far without you. I am grateful for your friendship and the countless hours you have spent helping me to achieve my goals. You are one in a million.

I would like to thank the Learning in Fitness and Education through Sports (LiFE Sports) Initiative for helping me to facilitate this study. My colleagues Lauren Paluta, Tark Newman, Luke O’Quinn, and Mickey Fraina were instrumental in providing support and assistance throughout the dissertation process. I am grateful for your
friendship and constant willingness to provide insight, suggestions, and a helping hand. LiFE Sports is a great team, which I am proud to be a part of.

Lastly, I would like to acknowledge my advisors – Dr. Brian Turner, Dr. Dawn Anderson-Butcher, Dr. Donna Pastore, Becky Wade-Mdivanian, Dr. Jerry Davis, and Dr. Don Stenta. Through your mentorship, I have been able to learn and experience far more than I ever imagined. Thank you for investing in me, creating new opportunities for me, and guiding me through my doctoral program. I am grateful for your commitment to my education, professional development, and personal wellness.
Vita

2004.......................................................... William Mason High School

2007.......................................................... B.A. Math-Economics & Business Administration, Anderson University

2009-2011 .............................................. Graduate Administrative Assistant, Campus Recreation Department, Baylor University

2011.......................................................... M.S. Sport Management, Baylor University

2011-2014 .................................................. Graduate Teaching Associate, College of Human Ecology, The Ohio State University

2014 to present ......................................... Graduate Research Associate, College of Social Work, The Ohio State University

Publications


Fields of Study

Major Field: Physical Activities and Educational Services
Studies in Sport Management

Specialization: Development through Sport, Recreation, and Play
General Interdisciplinary Specialization in Quantitative Research Methodology
Table of Contents

Abstract ........................................................................................................................................ ii

Acknowledgements .................................................................................................................. v

Vita .............................................................................................................................................. vii

Table of Contents ................................................................................................................... ix

List of Tables ............................................................................................................................. xii

List of Figures ............................................................................................................................ xiv

Chapter 1: Introduction ............................................................................................................... 1

  Sport-Based Positive Youth Development ............................................................................ 2

  Organizational Effectiveness ................................................................................................. 4

  Program Implementation ......................................................................................................... 6

  Program Engagement ............................................................................................................. 8

  Problem .................................................................................................................................. 9

  Purpose .................................................................................................................................. 17

  Research Questions .............................................................................................................. 17

  Definitions of Key Terms ....................................................................................................... 19

  Overview of Remaining Chapters ......................................................................................... 20

Chapter 2: Review of Literature ................................................................................................. 22

  Status of Youth ....................................................................................................................... 22
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk and Resilience Framework</td>
<td>26</td>
</tr>
<tr>
<td>Positive Youth Development</td>
<td>32</td>
</tr>
<tr>
<td>Sport-Based Positive Youth Development</td>
<td>34</td>
</tr>
<tr>
<td>Organizational Effectiveness</td>
<td>42</td>
</tr>
<tr>
<td>Organizational Effectiveness in Positive Youth Development Programs</td>
<td>50</td>
</tr>
<tr>
<td>Mechanisms of Positive Youth Development</td>
<td>52</td>
</tr>
<tr>
<td>Chapter 3: Methods</td>
<td>87</td>
</tr>
<tr>
<td>Overview</td>
<td>87</td>
</tr>
<tr>
<td>Research Context</td>
<td>88</td>
</tr>
<tr>
<td>Research Design</td>
<td>96</td>
</tr>
<tr>
<td>Samples</td>
<td>98</td>
</tr>
<tr>
<td>Validity and Reliability</td>
<td>101</td>
</tr>
<tr>
<td>Instruments</td>
<td>106</td>
</tr>
<tr>
<td>Data Collection</td>
<td>125</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>131</td>
</tr>
<tr>
<td>Chapter 4: Results</td>
<td>143</td>
</tr>
<tr>
<td>Overview</td>
<td>143</td>
</tr>
<tr>
<td>Demographics</td>
<td>143</td>
</tr>
<tr>
<td>Data Treatment</td>
<td>145</td>
</tr>
<tr>
<td>Validity and Reliability</td>
<td>148</td>
</tr>
<tr>
<td>Data Reduction</td>
<td>173</td>
</tr>
<tr>
<td>Normality</td>
<td>175</td>
</tr>
<tr>
<td>Chapter/Section</td>
<td>Page</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Measurement Models</td>
<td>179</td>
</tr>
<tr>
<td>Structural Equation Models</td>
<td>200</td>
</tr>
<tr>
<td>Chapter 5: Discussion</td>
<td>252</td>
</tr>
<tr>
<td>Organizational Effectiveness of a Sport-Based PYD Program</td>
<td>252</td>
</tr>
<tr>
<td>Impact of the Study</td>
<td>288</td>
</tr>
<tr>
<td>Implications for Practice</td>
<td>289</td>
</tr>
<tr>
<td>Limitations and Future Directions</td>
<td>295</td>
</tr>
<tr>
<td>Conclusion</td>
<td>302</td>
</tr>
<tr>
<td>References</td>
<td>306</td>
</tr>
<tr>
<td>Appendix A: Staff Instrument</td>
<td>336</td>
</tr>
<tr>
<td>Appendix B: Youth Pre-Test Instrument</td>
<td>339</td>
</tr>
<tr>
<td>Appendix C: Youth Post-Test Instrument</td>
<td>343</td>
</tr>
<tr>
<td>Appendix D: Session Log Training Handout</td>
<td>348</td>
</tr>
<tr>
<td>Appendix E: Parental Consent Form</td>
<td>351</td>
</tr>
<tr>
<td>Appendix F: Youth Assent Form</td>
<td>354</td>
</tr>
</tbody>
</table>
List of Tables

Table 1. Descriptions of program staff positions.............................................................92
Table 2. Factor loadings and Cronbach alphas for session log observed variables........151
Table 3. Model fit statistics for the tests of invariance in social competence scale scores
across two time points........................................................................................................158
Table 4. Model fit statistics for the tests of invariance in effort scale scores across two
time points........................................................................................................................................161
Table 5. Model fit statistics for the tests of invariance in teamwork scale scores across
two time points.............................................................................................................................164
Table 6. Model fit statistics for the tests of invariance in self-control scale scores across
two time points.............................................................................................................................167
Table 7. Model fit statistics for the tests of invariance in athletic competence scale scores
across two time points....................................................................................................................169
Table 8. Descriptive statistics of observed variables.........................................................176
Table 9. Descriptive statistics of transformed observed variables......................................178
Table 10. Correlation matrix of observed variables..............................................................202
Table 11. Covariance matrix of observed variables..............................................................204
Table 12. Structural equation modeling goodness of fit indices for social competence
Model A

Table 13. Structural equation modeling goodness of fit indices for social competence
Model B

Table 14. Structural equation modeling goodness of fit indices for athletic competence
Model A

Table 15. Structural equation modeling goodness of fit indices for athletic competence
Model B
List of Figures

Figure 1. Social competence measurement model at pre-test with standardized solutions ................................................................. 181
Figure 2. Social competence measurement model at post-test with standardized solutions ........................................................................................................ 184
Figure 3. Athletic competence measurement model at pre-test with standardized solutions ........................................................................................................ 186
Figure 4. Athletic competence measurement model at pre-test with standardized solutions ........................................................................................................ 187
Figure 5. One-factor program dosage measurement model with standardized solutions 192
Figure 6. Program implementation measurement model with standardized solutions .... 196
Figure 7. Program engagement measurement model with standardized solutions .... 197
Figure 8. Two-factor program dosage measurement model with standardized solutions ........................................................................................................ 199
Figure 9. Path diagrams of proposed structural models .............................................. 201
Figure 10. Path diagrams of social competence Model A structural models ............. 207
Figure 11. Path diagrams of social competence Model B structural models ............. 218
Figure 12. Social competence Model B1 with standardized solutions ................................. 227
Figure 13. Path diagrams of athletic competence Model A structural models ...............230

Figure 14. Path diagrams of athletic competence Model B structural models ...............241

Figure 15. Athletic competence Model B1 with standardized solutions ....................250
Chapter 1: Introduction

Within urban, impoverished areas of the United States live youth exposed to severe risk, who are faced with challenges and stressors unique to their circumstances. Children may be living in single-parent households with limited supervision, lacking quality housing or healthy food, unable to access prosocial activities in a safe environment, struggling within failing school systems and influenced by negative relationships (Federal Interagency Forum on Child and Family Statistics, 2013; Fraser, 2004). Urban youth are likely to be exposed to greater risk due to their environment (Hellison, 2000; Li, Nussbaum, & Richards, 2007; Pedersen & Seidman, 2005), which can inhibit positive development. The environmental risk can have a significant impact on their current and future health, satisfaction, and productivity.

From an ecological perspective, youth development is influenced by the interaction between the individual youth and their external environment (i.e., family, community, school, peers; Anderson-Butcher, 2011; Anthony, Alter, & Jensen, 2009). Risk factors can be found in the various systems of influence within a youth’s life, including their individual traits (e.g., behavioral disorders), as well as the influence of their peers (e.g., anti-social peers), parents/guardians (e.g., lack of supervision), school (e.g., school failure), and community (e.g., violence). Youth exposed to risk factors are
vulnerable to experiencing harm and displaying problem behaviors which can result in negative developmental outcomes (Anderson-Butcher, 2011; Anthony et al., 2009; Fraser, Kirby, & Smokowski, 2004).

Individual and environmental conditions associated with risk do not always lead to developmental digression, harm, or problem behavior. Youth who are able to overcome the adversity of their circumstances are considered resilient (Anthony et al., 2009). Youth development theorists explain resiliency as a product of both risk and protective influences (Anthony et al., 2009; Fraser et al., 2004). When resilient youth take advantage of their internal and external assets to overcome risk, these assets are considered protective factors as they help youth prevail over adversity (Fraser et al., 2004; The Search Institute, 1997). Protective factors can be found in multiple systems of influence: individual (e.g., self-control), interpersonal and social (e.g., caring relationships), and environmental (e.g., safety), which mitigate the risk youth are exposed to (Anderson-Butcher, 2011; Anthony et al., 2009; The Search Institute, 1997).

**Sport-Based Positive Youth Development**

The positive youth development (PYD) movement emerged as a response to the needs of youth vulnerable to risk. Rather than focus on managing the negative outcomes of risk, such as problem behaviors, PYD takes a strengths-based approach focusing on developing individual and environmental assets which can mitigate risk (Anderson-Butcher, 2011; Anthony et al., 2009; Lerner, Almerigi, Theokas, & Lerner, 2005; Vierimaa, Erickson, Cote, & Gilbert, 2012). PYD programs are intentionally designed to address and reduce risk factors and enhance protective factors through evidence-based
practices. These programs emerged as a means to intentionally prepare youth for adulthood through purposefully designed interventions (National Collaboration for Youth Members, 1998). PYD interventions consist of structured, sequential activities which build competencies and develop valuable life skills that are transferable to multiple contexts (e.g., school, home, community; Anderson-Butcher, 2005; National Collaboration for Youth Members, 1998).

Sport is a vital context for PYD, consisting of organized sport, recreation, physical activity, and play (Riley & Anderson-Butcher, 2012; Weiss, 2008). Millions of youth in the United States engage in organized and unstructured sport and physical activities, accounting for a substantial amount of discretionary time (Anderson-Butcher et al., 2011; Danish, Forneris, & Wallace, 2005; Ewing & Seefeldt, 2002; Petitpas et al., 2005; Theokas, Danish, Hodge, Heke, & Forneris, 2008; Weiss, 2008; Weiss et al., 2012). Many youth are attracted to sport due to the nature of the activities which include social interaction, competition, fun, challenge, as well as the popularity of the pastime (Papacharisis et al., 2007; Petitpas et al., 2005; Weiss et al., 2012). Youth sport programs can provide a safe, supervised environment, in which youth engage in healthy, constructive activities and develop competencies through participation.

Sport is a powerful vehicle for positive youth development, acting as a “hook” to attract youth engagement, while providing experiences which develop both sport and life skills (Hartmann, 2003). Sport-based PYD programs are intentionally designed with the dual focus of integrating sport and life skill instruction, demonstration, modeling, and activities, while still addressing risk factors and enhancing protective factors found in the
individual, social, and environmental systems (Anderson-Butcher et al., 2013; Petitpas et al., 2005). Exemplar sport-based PYD programs and models which have recently emerged include, Learning in Fitness and Education through Sports (LiFE Sports; Anderson et al., 2011), Sports United to Promote Education and Recreation (SUPER; Theokas et al., 2008), and Girls on the Run (Gabriel, DeBate, High, & Racine, 2011).

**Organizational Effectiveness**

The organizational effectiveness of sport-based PYD programs is critical, due to the impact that these programs have on vulnerable youth exposed to severe risk with limited alternative options to engage in supervised, prosocial activities with caring adults. In light of the constantly changing internal and external environment, sport-based PYD programs must determine effective strategies and practices in order to have maximum impact on their target population (Hossein, Ramezanineghad, Yosefi, Sajjadi, & Malekakhlagh, 2011; Schreiber, 1983). Within PYD literature, a program is considered effective if it is evidence based, which consists of: 1) demonstrating the relevancy of the program, 2) theoretical explanations for the program design, and 3) evaluation of program outcomes (Baldwin, Caldwell, & Witt, 2005).

This researcher contends that program effectiveness within PYD must move past a theoretical explanation of program design and incorporate an empirical investigation of the interaction between program design and outcomes. Such an assessment integrates the goals and process models of organizational effectiveness. The goals model is the most widely accepted framework of effectiveness, focusing on goal (i.e., outcome) attainment as a demonstration of effectiveness (Chelladurai & Haggerty, 1991; Hossein et al., 2011).
Therefore the goals model requires evaluation of a program’s outcomes. Comparatively, the process model focuses on the internal processes of a program, including program implementation, curriculum, and procedures (Chelladurai & Haggerty, 1991; Hossein et al., 2011; Shilbury & Moore, 2006). Through process and outcome evaluation, a sport-based PYD program can determine which program components (i.e., processes) act as mechanisms of PYD (i.e., outcomes).

Sport-based PYD programs facilitate intentional interventions designed for prevention and youth development. For an intervention to act as a mechanism of PYD, the program participants must be exposed to the intervention (i.e., program dosage). Program dosage reflects one’s degree of exposure to the program intervention, incorporating both quantity and quality of exposure (Nation et al., 2003). To examine program dosage as a mechanism of PYD, implementation of the intervention and program engagement must be examined.

A sport-based PYD program may implement an intervention designed to address specific risk and protective factors and then evaluate the outcomes of the program, but have limited understanding of what influenced those outcomes. The achievement of program outcomes is insufficient evidence of organizational effectiveness. More specifically, claiming effectiveness through outcome evaluation alone may result in a Type III error in which a program erroneously attributes achieved outcomes to the program interventions (Dobson & Cook, 1980). To accurately claim organizational effectiveness, a program must demonstrate that when properly implemented, the intervention will result in prevention and youth development (Dusenbury et al., 2003;
Law & Shek, 2011). Through evaluation of implementation fidelity and program engagement, a program can establish internal validity and program dosage in order to account for the programmatic mechanisms of PYD.

**Program Implementation**

Program implementation fidelity reflects program delivery, including dimensions of adherence, quality, dosage, and response (Dusenbury et al., 2003; Fagan et al., 2008; Tucker & Blythe, 2008). In relation to the current study under investigation, the program staff are responsible for adhering to the intervention structure, curriculum, and climate when delivering the program. Additionally, the staff must engage the participants to expose them to the intervention in order to facilitate the developmental growth.

The intervention being studied in this research adopts an experiential learning framework, focusing on experiences as the predominant learning agent, which is difficult to evaluate due to the complexity of learning (Gosen & Washbush, 2004; Kold, Boyatzis, & Mainemelis, 2001). Facilitation of these experiences utilizes a three-step process of briefing, facilitating, and debriefing to purposefully guide the participants through the learning process (Schoel, Prouty, & Radcliffe, 1998). Within PYD, a progressive sequence of activities constitute the experience during which the learning will take place. The intervention curriculum consists of the content and skills to be taught through sport and play-based activities. The current intervention integrates sport and social skill curriculum, utilizing the Sport Education model and Teaching Personal and Social Responsibility model (TPSR; Hellison, 2003; Siedentop, 1998).

The Sport Education model is a prominent sport curriculum framework found
effective within the context of institutional sport (Presse, Block, Horton, & Harvey, 2011; Siedentop, 1998; Wallhead & O’Sullivan, 2005). Siedentop (1998) proposed six primary features to be implemented within sport programs, including: seasons, team affiliation, formal competition, record keeping, culminating events, and festivity. To embed social skill instruction within the Sport Education model, the intervention studied integrates the TPSR model, which has limited empirical evidence (Hellison, 2003; Hellison & Walsh, 2002). The TPSR model consists of five developmental levels designed to enhance personal and social responsibility within the context of physical education. The developmental levels include: 1) respecting the rights and feelings of others, 2) participation and effort, 3) self-direction, 4) helping others and leadership, and 5) transferring learning outside the gym. Additionally this model emphasizes integration of sport and social skill instruction, transfer of learning, student empowerment, and teacher-student relationships throughout the curriculum.

Through a program’s structure, curriculum, and staff practices the climate is created. Several critical features of a PYD setting have been proposed and supported in the literature, including physical and psychological safety, supportive relationships, opportunities to belong, promotion of a mastery motivational climate, opportunities for skill building, and fun (Allodi, 2007; Anderson-Butcher et al., 2011; Eccles & Gootman, 2002). A caring, mastery-oriented climate has been suggested to contribute to PYD; however, there have been inconsistent findings regarding the impact of caring-based programming and staff practices (Newton, Watson, Gano-Overway, Fry, Kim, & Magyar, 2007; Riley, 2013). In relation to a mastery motivational climate, empirical research has
consistently found this type of environment associated with positive youth outcomes (Barkoukis, Tzorbatzoudis, & Grouios, 2009; Boyce, Gano-Overway, & Campbell, 2009; Ntoumanis & Biddle, 1999). The interpretation of these findings is limited as the primary methodological approach has examined perceived climate at one point in time, while motivational climate is a dynamic construct (Ames & Archer, 1988; Ntoumanis & Biddle, 1999).

**Program Engagement**

For youth participants to benefit from a program’s intervention (i.e., structure, curriculum, climate), they must be exposed to the intervention (Anderson-Butcher, Newson, & Ferrari, 2003). Program engagement involves an interaction between the intervention, participants, and program facilitator, for which the facilitator entices participation in the intervention and the participant responds through active involvement. Participant engagement has almost exclusively been assessed quantitatively (e.g., attendance), with inconsistent findings regarding the outcomes of engagement (Anderson-Butcher, Cash, Saltzburg, Midle, & Pace, 2004; Anderson-Butcher et al., 2003; Bohnert, Fredricks, & Randall, 2010).

To better understand participant engagement, an examination of the staff practices which engage participants is warranted. More specifically, by establishing a caring, mastery-oriented environment through developing positive, supportive relationships with youth participants, staff can enhance the emotional engagement of youth. Through enhancing the participants’ sense of belonging and attachment to the program and staff and providing emotional and autonomy support, youth become more invested and
motivated to engage. Perceived belonging and leader support have consistently been found associated with positive outcomes (Anderson-Butcher & Fink, 2005; DuBois & Silverthorn, 2005; McDonough, Ullrich-French, Anderson-Butcher, Amorose, & Riley, 2013; Ullrich-French & McDonough, 2013), however research has not yet examined these perceived emotions in relation to behavioral engagement.

In order for a sport-based PYD program to be effective, the program intervention must be designed and implemented based upon theoretically and empirically supported practices and program components. Additionally, the quality of exposure to the intervention is dependent on youth engagement, largely influenced by the program design and staff practices. The literature suggests that many PYD programs lack implementation fidelity (Gottfredson & Gottfredson, 2002; Hallfors & Godette, 2002; Paulsell, Kisker, Love, & Raikes, 2002), with few empirical studies examining both process and outcome evaluation to determine the programmatic mechanisms of PYD as well as the strengths and weaknesses of a program design and staff practices (Wright & Burton, 2008). Through an integrative approach of investigating both process and outcome evaluation within the context of sport-based PYD, administrators can gain a better understanding of the programmatic mechanisms of PYD within sport, which can then inform best practices and program decisions.

**Problem**

In the United States, there is an underserved youth population whose positive development is jeopardized by exposure to risk factors present within their multiple systems of influence (Li et al., 2007; Pedersen & Seidman, 2005). Sport-based PYD
programs have emerged to address the risk prevalent among vulnerable populations and promote PYD through sport. The current program under investigation uses sport and play-based activities to enhance the athletic and social skills of economically disadvantaged youth. The program intervention utilizes an experiential learning framework (Kolb, Boyatzis, & Mainemelis, 2001), with curriculum based upon the Sport Education and TPSR models (Hellison, 2003; Siedentop, 1998), and is intended to be implemented within a caring, mastery-oriented environment.

The use of experiential learning has been empirically investigated in a variety of learning environments with a consistent result of positive outcomes (Gosen & Washbush, 2004). However, Gosen and Washbush (2004) have found that most experiential learning based studies have lacked rigor and objective measures. How learning takes place is difficult to measure as learning is unique to each individual. Therefore, there is a need to further explore the use of experiential exercises as a medium for development. Moreover, research evaluating an experiential based program must utilize objective measures which match the learning goals associated with the experiential exercises and adopt a longitudinal research design to accurately measure whether development occurred.

A program curriculum is designed to target learning in certain developmental areas, for which the current program under investigation focuses on sport and social skill development, utilizing the Sport Education and TPSR models. The Sport Education model has been heavily studied within the context of physical education and institutional sport, and consistently found an effective mechanism for youth development (Presse et al., 2011; Wallhead & O’Sullivan, 2005). This model has yet to be empirically tested
within contexts outside of institutional sport (i.e., school-based physical education; (Wallhead & O’Sullivan, 2005). Sport-based PYD is an important context in which the Sport Education model can be applied and tested. There is a need to establish the external validity of the Sport Education model in order to determine its effectiveness as a programmatic mechanism of PYD within alternative contexts. Comparatively the TPSR model has been studied within the context of PYD; however, limited empirical studies can be found in the literature (Hellison & Walsh, 2002).

When evaluating the effectiveness of curriculum models, it is imperative to establish implementation fidelity of the model in order for the results to be reliable (Metzler, 2001). Of the limited empirical research examining the TPSR model, only two studies were found which evaluated implementation fidelity of the TPSR model in addition to learning outcomes (Pascual, Escarti, Llopis, Gutierrez, Marin, & Wright, 2011; Wright & Burton, 2008). Both studies used a qualitative approach, highlighting the need for a rigorous quantitative examination of TPSR implementation fidelity and outcomes to accurately determine the effectiveness of the TPSR model within sport-based PYD.

The intervention structure and curriculum, along with staff practices, contribute to the intervention climate, for which the current program under investigation is designed to create a caring, mastery-oriented environment. Although caring and positive relationships have been found associated with a plethora of positive outcomes (Allodi, 2010; Anderson-Butcher, 2010; Eccles and Gootman, 2002; Larson & Walker, 2005), there have been inconsistent findings regarding the impact of staff practices reflecting care and
support (Riley, 2013). Further research is needed to examine the impact of staff practices on PYD, rather than solely focusing on participant perceptions.

In relation to a mastery-oriented environment, empirical research has consistently found positive developmental outcomes associated with a mastery climate (Barkoukis et al., 2008; Boyce et al., 2009; Miller et al., 2004; Ntoumanis & Biddle, 1999). However, most research on motivational climates have measured this construct at one-point in time, failing to account for the dynamic nature of one’s environment, influenced by inconsistent situational cues (Ntoumanis & Biddle, 1999). This calls for a longitudinal research design which can capture changes in situational cues, for a more accurate measure of the motivational climate.

Even if a sport-based PYD program intervention integrates evidence-based models (e.g., experiential learning, Sport Education, TPSR, mastery motivational climate), the program might still not achieve their desired outcomes. Unfortunately, many programs are not able to fully achieve their objectives due to challenges and barriers to implementation fidelity and varying participant engagement (Bohnert et al., 2010; Dusenbury et al., 2003; Gottfredson & Gottfredson, 2002; Hallfors & Godette, 2002; Paulsell et al., 2002). The potential for sport-based PYD programs to promote PYD is often not actualized in large part due to transient or untrained staff members that implement the program as well as intervening circumstances. Approximately “90% of youth sport coaches in the United States do not have formal training in coach education or youth development” (Petitpas et al., 2005, p. 65). The achievement of intended program objectives is dependent upon program implementation fidelity. A sport-based
PYD program designed to teach sport and social skills and promote PYD in a caring, mastery-oriented climate must employ staff who are knowledgeable and trained on effective instructional techniques, PYD principles, and behavioral management strategies (Anderson-Butcher et al., 2013).

Theoretically, connecting a program’s intervention with desired program outcomes is insufficient evidence that the intervention is effective, as implementation fidelity across staff members and across days is often inconsistent due to intervening circumstances. Fully implementing the necessary structure of an intervention may be difficult on days with time constraints or weather delays; implementing new curriculum will be more difficult than implementing familiar curriculum, especially for new staff members; creating a caring, mastery goal-oriented climate may be challenging when youth are more rambunctious and aggressive or when the intervention incorporates competition.

When evaluating the effectiveness of a program, examination of implementation fidelity is pertinent in order to establish program dosage and determine areas of organizational effectiveness. A greater understanding of how program implementation and engagement influences youth development, as well as what specific components of the program (e.g., program structure, curriculum, climate, engagement) significantly contribute to growth in youth outcomes, can have major implications for organizational effectiveness. Such evidence can inform program design, intervention strategies, staff training, and support the justification for sport-based PYD programs.

Youth participation in sport-based PYD programs is typically voluntary, for
which degree of engagement is often determined by satisfaction with the activity and staff (Anderson-Butcher et al., 2003). For this reason, responsiveness to the interventions is considered a dimension of implementation fidelity (Dusenbury et al., 2003; Fagan et al., 2008; Tucker & Blyth, 2008). As attendance is not required at most programs, many youth may attend sporadically, limiting the youth’s exposure to the program’s intervention which will influence their outcomes. Research has demonstrated a decline in sport participation as youth get older, with approximately 35% of youth dropping out of sport per year (Anderson-Butcher et al., 2003; Ewing & Seefeldt, 2002). Sport-based PYD programs are faced with participation inconsistencies as well as increased attrition as youth get older, making recruitment and retention more difficult, limiting the program’s opportunity to make an impact on youth. Youth that attend sporadically will not be exposed to the full measure of the program’s intervention, which will limit their developmental outcomes.

A sport-based PYD program intervention may be implemented with high fidelity, but without sufficient dosage the intervention will have little impact or a temporal effect on program participants (Nation et al., 2003). There is a gap in the literature regarding the influence of the quality of exposure to an intervention on change in youth outcomes. Researchers have primarily examined quantity of intervention exposure through program documented or self-reported measures of attendance (Anderson-Butcher et al., 2003; Anderson-Butcher et al., 2004; Anderson-Butcher & Cash, 2010; Anderson-Butcher & Fink, 2005; Gottfredson, Cross, Wilson, Rorie, & Connell, 2010). Very few empirical studies have examined quality of participant engagement, consisting of either behavioral,
emotional, or cognitive engagement (Roth, Malone, Brooks-Gunn, 2010; Weiss, Little, & Bouffard, 2005). Roth and colleagues (2010) suggested a linkage between participant engagement and program quality, as reflected by high implementation fidelity, highlighting the need for research which examines both program engagement and implementation as programmatic mechanisms of PYD.

It is pertinent to have a greater understanding of the degree of engagement necessary to produce positive change. Youth that attend a program regularly, but do not actively pay attention to or demonstrate effort in the intervention, are receiving a limited exposure to the intervention. A more in-depth understanding of how program engagement mediates developmental change can inform program design and intervention strategies as well as guide staff training on intervention implementation.

Several PYD program design features are critical for youth engagement, including creating a sense of belonging and providing leader support (Anderson-Butcher & Conroy, 2002; Ullrich-French & McDonough, 2013). Research suggests that youth perceptions can be more critical than staff practices (Eccles & Gootman, 2002). Therefore, an examination of both staff practices and youth perceptions related to the intervention climate (i.e., caring, belonging, leader support) can provide a greater understanding of the mechanisms of PYD.

As a whole, researchers suggest the need to look at mechanisms that underlie the relationships among participation and youth outcomes (Anthony et al., 2009; Eccles, Barber, Stone, & Hunt, 2003; Gould & Carson, 2008; Hedstrom & Gould, 2004; Weiss & Smith, 2002), more specifically the impact of participation on positive outcomes and
problem behaviors (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2004). Research in
the PYD context documents positive youth outcomes, and yet other research does not
(Anthony et al., 2009; Eccles et al., 2003; Fraser-Thomas, Côté, & Deakin, 2005; Gould
& Carson, 2008). There is a need to further explore the impact and mechanisms of sport-
based PYD interventions.

Existing studies in PYD are often limited by methodological constraints and
research design. The use of advanced statistical techniques is often missing in studies of
settings-level features (Martinez & Raudenbush, 2008). Additionally, many studies are
limited to a correlational or descriptive design, which explores youth experiences but
does not contribute to program impact. When studying youth development, only a
longitudinal design will capture change over time. Only a few research studies utilize
pretest/posttest designs which allow for information related to the influence of staff
practices on youth development. Ultimately, longitudinal research which takes into
account the interconnected associations between program implementation, engagement,
and outcomes, as well as between staff practices and youth perceptions, will extend the
current body of knowledge on the mechanisms of PYD within sport.

Lastly, a challenge for many sport-based PYD programs that should not be
overlooked is the struggle to secure the necessary funding and resources to fully deliver
the program and support the overall infrastructure (Anderson-Butcher, 2011). “The bulk
of new public and private funding is going to school systems, leaving hundreds of
community-based programs insecurely and inadequately funded” (Halpern, 2000, p. 186).
This issue is exacerbated in urban communities where programs are “overburdened and
underfunded” (Anderson-Butcher, 2011, p. 2843). Acquiring additional resources and funding is difficult in the current economy, with hundreds if not thousands of programs competing for financial support. With limited resources, programs are not able to reach as many youth or offer a wide-variety of sport activities. Continuing the effort to provide evidence of the effectiveness of sport-based PYD programs contributes to the justification for external resources and support for programs, which has a direct impact on the underserved populations the programs are targeting.

**Purpose**

Research evaluating the effectiveness of sport-based PYD programs is needed to provide evidence which can inform program decision making, improve program practices, guide effective intervention strategies, and ultimately advance the goals of PYD. The purpose of the current study was to examine program dosage, which incorporates both implementation fidelity and engagement, as programmatic mechanisms of PYD, with a focus on the development of social and athletic competencies. A longitudinal research design of both staff and youth participants was employed in order to conduct a more rigorous assessment of the effectiveness of a sport-based PYD program and gain a better understanding of the relationships and contributions of the mechanisms of PYD. The current study sought to address the gaps in existing research on sport-based PYD programs and provide evidence which can enhance PYD strategies.

**Research Questions**

The current study was designed to examine the effectiveness of a sport-based PYD program through an integration of process and outcome evaluation. The primary
research question that guided the current investigation was:

*How does program dosage, essentially exposure to a program with high implementation fidelity, mediate PYD?*

Program dosage requires both program implementation and engagement and was operationalized as a comprehensive construct. Program implementation fidelity was assessed through examining the staff’s perceived implementation of the intended program structure, curriculum, and climate. Additionally, program engagement was evaluated by the youth’s perceived behavioral engagement during the program, as well as perceived belonging and leader support which are suggested to engage the youth participants.

The sport-based PYD program under investigation is designed to enhance the social competence, effort, teamwork, self-control, and athletic competence of youth participants through a 19 day summer program. As a means to more fully understand how program dosage mediates PYD within the context of the program under investigation, the primary research question was sub-divided by specific program outcome.

*R1: How does program dosage mediate social skill development?*

*R2: How does program dosage mediate athletic skill development?*

The answer to these questions will provide valuable theoretical and practical implications which can extend PYD research and practice. With a greater understanding of the programmatic mechanisms of PYD within sport, programs will be more equipped to strategically design, deliver, and evaluate their intervention. Ultimately, the value will emanate among the vulnerable populations served.
Definitions of Key Terms

The following are operational definitions of the constructs which will be utilized through the study:

1. Positive youth development (PYD) is defined as a strengths-based approach to promoting healthy development among children and adolescents (Lerner et al., 2005; Vierimaa et al., 2012).

2. Athletic competence is defined as one’s athletic skill, knowledge, or capacity.

3. Social competence is defined as one’s ability to interact with other individuals in a social context (Anderson-Butcher, Amorose, Lower, Gibson, Riley, & Ruch, in press; Martinek & Hellison, 1997).

4. Effort is defined as the exertion of physical and/or mental energy into goal-directed behavior.

5. Teamwork is defined as working collectively and cooperatively with others in order to achieve a common goal (Lower, Newman, & Anderson-Butcher, in review).

6. Self-control is defined as awareness and control over one’s behavior and actions (Gresham & Elliott, 1990).

7. Implementation fidelity is defined as adherence to a program design, reflected by implementing a program as intended (Dusenbury et al., 2003; Fagan et al., 2008; Mowbray et al., 2003; Tucker & Blythe, 2008).

8. Organizational effectiveness is defined as a program with high implementation fidelity, achieving anticipated program outcomes (Dusenbury et al., 2003; Law & Shek, 2011).

9. Behavioral engagement is defined as “active participation and includes factors such as
effort, concentration, attendance, following the rules, and avoiding trouble” (Bohnert et al., 2010, p. 593).

10. Belonging is defined as a connection or bond within a social structure (Anderson-Butcher & Conroy, 2002).

11. Leader support is defined as care, warmth, acceptance, and closeness extended by a leader to his/her follower (Bowers, Geldhof, Schmid, Napolitano, Minor, & Lerner, 2012; Eccles & Gootman, 2002).

**Overview of Remaining Chapters**

Chapter 1 provided an introduction to PYD within the context of sport, outlining program design, implementation, engagement, and outcomes to set the stage for an examination of programmatic mechanisms of PYD. Within this section two research questions were proposed concerning program dosage as a mediator of PYD. The second chapter examines PYD, organizational effectiveness, and program implementation and engagement with more depth, exploring the specific models and theories adopted by the current program under investigation. The existing body of knowledge is presented to provide conceptual and empirical justification for the current study. The third chapter reviews the methods used to answer the proposed research questions. Moreover, chapter three discusses the research context and design, subjects, measures used, and data collection and analysis procedures employed in the current study. The fourth chapter presents the results of the current study, focusing on the population demographics, individual measures, measurement models, and structural models. The final chapter provides a discussion of the results, utilizing theoretical and empirical support to explain
the findings. Chapter 5 also addresses the limitations of the current study and proposes suggestions for future research.
Chapter 2: Review of Literature

Overview

The purpose of the following chapter is to provide an in-depth review of the current body of knowledge on PYD and organizational effectiveness. The chapter will begin with a description of vulnerable youth in the United States and discuss the emergence of the PYD movement, with emphasis on the sport context. Models of organizational effectiveness will be presented to provide a framework for evaluating effectiveness of sport-based PYD programs. The chapter will conclude with an overview of the literature investigating programmatic mechanisms of PYD. A discussion of current theoretical arguments and previous empirical studies will highlight future directions for the study of organizational effectiveness within PYD. The review of literature will utilize the current program under investigation as the framework for the models and theories presented.

Status of Youth

There are approximately 73.7 million children in the United States, which accounts for 24% of the total population (Federal Interagency Forum on Child and Family Statistics, 2013). Many of these children face adversity and difficulty due to their present circumstances. Based on the U.S. Department of Health and Human Services

Poverty affects youth in many ways including the quality and availability of food, housing, health care, and transportation, limiting a child’s social and educational opportunities as well as quality of life (Fraser, 2004). In 2012, the United States Department of Agriculture (USDA) classified 22% of households with children as “food insecure” (Federal Interagency Forum on Child and Family Statistics, 2013). Food insecurity describes a socio-economic household condition in which individuals within the household have limited access to food, which may result in hunger due to prolonged lack of food (USDA, 2014). In regards to a youth’s physical environment and safety, 46% of households with children have one or more housing problems, which may include inadequate housing, crowded housing, or a high housing cost burden (Federal Interagency Forum on Child and Family Statistics, 2013). Inadequate food or housing can create additional stress within the home, exacerbating the conditions which can negatively affect youth development.

There are currently about 7 million children (9%) without health insurance, which can be problematic due to the current health status of youth in the U.S. (Federal Interagency Forum on Child and Family Statistics, 2013). In 2007-2008, the average diet quality of children was scored 50%, based on the Healthy Eating Index which reflects the percentage of recommended dietary intake; this statistic falls considerably below the recommended dietary guidelines (Federal Interagency Forum on Child and Family Statistics, 2013). Approximately one out of every three children and adolescents are
overweight or obese, with 18% of children (ages 6-17) reported as obese in 2009-2010 (Centers for Disease Control and Prevention [CDC], 2013; Federal Interagency Forum on Child and Family Statistics, 2013). A lack of access to the necessities for healthy living, safety, education, and social support can significantly influence youth development and future success.

Disadvantaged youth are more likely to live in environments where they are faced with disparity and challenges (Fraser, 2004). Approximately 24% of children live with only their mothers; single parents may not be able to provide adequate supervision or social support which can impact their child’s behavior (Federal Interagency Forum on Child and Family Statistics, 2013; Fraser, 2004). In 2011, the adolescent birth rate was 15 per 1,000 adolescents (ages 15-17), with 47% of high school students reporting having had sexual intercourse. In regards to drugs and alcohol, in 2012 25% of high school seniors reported using illicit drugs in the past 30 days and 24% reported consuming five or more alcoholic beverages in a row in the past two weeks (Federal Interagency Forum on Child and Family Statistics, 2013). These risky behaviors have implications for current and future challenges as well as the potential for elongated drug and alcohol use.

Problem behaviors can have a direct or indirect effect in other areas of a child’s life. In 2012, 8% of youth (ages 16-19) were neither enrolled in school nor working, for which older youth (ages 18-19) were found to be “almost five times as likely to be detached from school and work activities” than youth ages 16-17 (Federal Interagency Forum on Child and Family Statistics, 2013, p. 54). Lack of social support, safety, and access to educational and extracurricular opportunities as well as inadequate housing,
physical activity, and dietary intake are considered risk factors associated with poverty which can negatively impact youth development. Youth exposed to these risk factors often exhibit greater problem behavior as a result of their environmental circumstances. These external forces disrupt and impede PYD (Ginwright & Cammarota, 2002).

**Socio-cultural Factors**

Urban minority youth are more likely to have grown up in conditions associated with a disadvantaged, impoverished community (Hellison, 2000; Li et al., 2007; Pedersen & Seidman, 2005). Policy makers concerned with youth development have historically focused on impoverished, urban youth, as they are more at risk for deteriorated development and negative outcomes than advantaged suburban youth (Pedersen & Seidman, 2005). Conditions faced within this context are highlighted by lack of access to resources, lack of opportunity for prosocial activities, and general disparity compared to socioeconomically advantaged communities (Halpern, 2000; Li et al., 2007).

Urban youth are more likely to be exposed to multiple negative conditions and risk that can compromise their development (Li et al., 2007; Pedersen & Seidman, 2005). Inner-city neighborhoods are becoming increasingly unsafe (e.g., risky street culture) with inadequate opportunities for recreational activities for children and adolescents (Halpern, 2000). Negative economic conditions can have an adverse effect on a child’s family and social environment, creating more challenge and risk that youth are exposed to. Halpern (2000) suggests that social agents (e.g., family, peer group, school, neighborhood) can be a source of difficulty and stress for poor urban youth as they may contribute neglect, hostility, or rejection rather than support. However, these social agents
can also act as social support systems for youth exposed to severe risk and challenge in their environment, which can nullify the adverse effects of risk factors. In general, the primary determinant of environmental risk is poverty.

**Risk and Resilience Framework**

The risk and resilience framework emerged from an ecological perspective that youth development is significantly influenced by the interaction of the internal characteristics of a child (e.g., biological, psychological, social) and their external environment created by their family, school, community, and peer group (Anderson-Butcher, 2011; Anthony et al., 2009). Bronfenbrenner (1977), one of the leading ecological theorists, defined the ecology of human development as “the scientific study of the progressive, mutual accommodation, throughout the life span, between a growing human organism and the changing immediate environments in which it lives” (p. 514). The interaction between the individual and the environment is considered complex, dynamic, and interdependent (Bronfenbrenner, 1977). Adoption of this risk and resilience framework has led to a critical examination of the internal and external factors that influence youth development in order to understand how to reduce problem behavior and enhance PYD (Anthony et al., 2009).

**Risk Factors**

Disadvantaged youth are considered to be exposed to a variety of risk factors which can negatively influence their development. Risk factors have been defined as influences (e.g., events, conditions, or experiences) that increase one’s probability of problem behavior, harm, or developmental digression (Anderson-Butcher, 2011; Anthony
et al., 2009; Fraser et al., 2004). Ginwright and Cammarota (2002) label risk factors common in urban environments (e.g., racism, unemployment, violence) as “destructive forces”, as they threaten the well-being of communities and developing youth (p. 86). In the social sciences, risk factors are considered probabilities of future outcomes (Fraser, 2004), for which risk can appear as a combination of conditions (e.g., internal and external factors) or a singular condition (Anthony et al., 2009). Youth exposed to several risk factors have a greater propensity for certain negative outcomes (e.g., problem behavior; Fraser et al., 2004).

Risk factors can emerge in various contexts from individual (e.g., aggression, mental health needs, behavioral disorders), to interpersonal and social (e.g., lack of social support, school failure, association with antisocial peers), and environmental sources of influence (e.g., poverty, violence, lack of access to prosocial activities; Anderson-Butcher, 2011; Anthony et al., 2009; Fraser et al., 2004; Gresham, 1997; Hawkins, Catalano, & Miller, 1992; Li et al., 2007). Culture can play a significant role in risk factors among youth (Ginwright & Cammarota, 2002). Some risk factors may be unique to certain groups within the population or operate differently depending on the culture (Fraser, et al., 2004). The individual and environmental conditions considered to be a risk factor in one culture may not be as severe of a risk in other cultures based upon the cultural structures in place. Social influences working with youth must interpret risk factors in the context of the youth’s cultural, religious, and social background to fully understand what influences have greater potential to cause negative outcomes.

Risk factors have been found to lead to a variety of negative future outcomes.
Problem behavior consistently found associated with risk factors include delinquency, drug and alcohol abuse, violence, school failure, and adolescent pregnancy (Anderson-Butcher, 2011; Anthony et al., 2009). Risk factors have been found to lead to violent crimes, unemployment, lost productivity, and possibly loss of life (Hawkins et al., 1992). Poverty, in particular, has been associated with stress, behavioral issues, and academic failure (Li et al., 2007). For youth to develop and thrive and avoid negative future outcomes, these risk factors many youth are exposed to must be mitigated.

When youth are exposed to multiple risk factors or one risk factor over an extended period of time, there is a potential cumulative effect which is suggested to increase one’s vulnerability to risk (Fraser, et al., 2004). As the number of risk factors and prolonged exposure increase, youth become more susceptible to the risk (Fraser, et al., 2004). Sources of social influence may not be able to reduce all risk factors, but even the reduction or elimination of one risk can make a profound impact on a youth’s future outcomes. By understanding risk mechanisms, youth programs can develop targeted interventions for greater impact.

**Protective Factors**

Ecological theorists assert that youth exposed to risk factors have a greater likelihood of negative future outcomes, however this likelihood is not a certainty as some youth demonstrate resiliency in the face of adversity (Fraser, 2004). Resilient youth take advantage of individual and environmental assets and strengths to overcome the risk factors they are exposed to (Fraser, et al., 2004; The Search Institute, 1997). These assets and strengths act as protective factors, shielding youth from the influence of risk factors,
essentially mitigating the risk (Anderson-Butcher, 2011; Anthony et al., 2009). Internal
resources reflect individual traits which can enhance a youth’s ability to adapt and
overcome negative environmental conditions (Anthony et al., 2009). In comparison,
external resources may reflect peer, family, school, and/or community assets (Anthony et
al., 2009).

Protective factors include multiple systems: individual (e.g., social competence,
positive self-perception, adaptability), interpersonal and social (e.g., caring relationships,
constructive use of time, attachment to parents), and environment (e.g., explicit
boundaries and expectations, physical and psychologically safe environment, access to
services, support, and resources) that can mediate risk youth are exposed to (Anderson-
Butcher, 2011; Anthony et al., 2009; Fraser et al., 2004; Gresham, 1997; Lawson &
Anderson-Butcher, 2000; Li et al., 2007; Malecki & Elliot, 2002; The Search Institute,
1997; Weiss, Kipp, & Bolter, 2012). Youth programs can work towards strengthening
these internal and external assets in order to foster youth resilience. Similar to risk,
multiple protective factors or exposure to protective factors over an extended period of
time can have a cumulative, additive effect. Youth with a greater number of protective
factors are more likely to have positive behaviors (e.g., succeed in school, value diversity,
delay gratification; The Search Institute, 1997).

Protective processes are mechanisms which modify risk and include: reduction of
risk impact, reduction of negative chain reactions, development of self-perceptions, and
opening of opportunities (Fraser et al., 2004). Reduction of risk impact is the process of
reducing one’s exposure to a risk factor. If exposure is reduced, the likelihood of negative
outcomes is also reduced. Reduction of negative chain reactions refers to the connection between risk and outcome. This process can insert a protective factor to disrupt the connection between risk and outcome (Anthony et al., 2009). Self-perceptions, such as self-esteem and self-efficacy can strengthen a youth’s adaptive behavior. Enhancing a youth’s ability to adapt will provide them the skills (e.g., coping skills) to overcome adversity and continue to function.

The fourth protective process includes any opportunity which fosters social and educational growth, creating avenues for youth to experience success. Through achieving success, self-perceptions and motivation are enhanced, increasing the likelihood of continued or future success. Youth programs are able to intentionally design their program to facilitate protective processes (e.g., taking youth out of an unsafe environment, addressing family and community risk factors, developing youths’ positive self-perception, and teaching youth life skills) in order to enhance PYD.

Resilience

Many PYD theorists consider resilience to be a product of the interaction between risk and protective factors (Anthony et al., 2009; Fraser et al., 2004). Anthony and colleagues (2009) characterize resilience as a child’s ability to adapt when exposed to risk, as to overcome the adversity. Taking into consideration the multi-system conceptualization of an individual within their environment, protective factors can be enhanced and/or provided to mitigate risk, resulting in resilient youth who are less vulnerable to the risk they are exposed to. Resilience can help youth, exposed to serious risk, handle and rise above difficult situations, decreasing the potentially negative
influence of risk factors.

Enhancing the resilience of youth should be a priority among programs working with youth. In order to effectively foster resilience, programs must have a clear understanding of the conditions that influence resilience (Martinek & Hellison, 1997). Youth vulnerability to risk is unique to the individual and changes as youth develop and mature (Anderson-Butcher, Newsome, & Ferrari., 2003). Over time, youth will react differently to risk factors, based on changes in their individual and environmental conditions (Fraser et al., 2004). Risk factors will also have greater influence on youth at different times during their maturation. For example, interpersonal risk factors may have a greater affect as youth approach adolescence. Fraser and colleagues (2004) argue that resilience is a dynamic quality that fluctuates based on individual, social, and environmental conditions. Understanding the dynamic nature of resilience will better prepare programs to identify the influential risk factors and relevant protective factors in order for those programs to develop effective, targeted interventions.

Youth programs can help cultivate the resilience of youth through promoting and strengthening protective factors while simultaneously reducing and addressing risk factors. Gresham (1997) supports the assertion that merely addressing deficits of protective factors is inadequate if the interfering or competing risk factors remain. Both protective and risk factors must be targeted to increase the probability of positive development.

Informal and formal interventions can be designed to reduce risk and enhance resilience. An informal intervention is based on “incidental learning which takes
advantage of naturally occurring behavioral incidents or events” in which positive behaviors and norms are taught and reinforced and poor behaviors are addressed (Gresham, 1997, p. 233). In comparison, formal interventions can take place through purposeful curriculum designed to strengthen specific protective factors and reduce distinct risk factors. Formal interventions typically focus on common risk and protective factors considered universal (rather than incident specific risk and protective factors), which can potentially decontextualize these conditions (Gresham, 1997). Therefore, a combination of both formal and informal interventions, addressing universal factors while simultaneously targeting factors that are most relevant to the target population (Hawkins et al., 1992), is optimal (Halpern, 2000).

**Positive Youth Development**

PYD is a strengths-based approach to promoting healthy development among children and adolescents (Lerner et al., 2005; Vierimaa et al., 2012). Rather than focus on the correction of poor behavior, a strengths-based perspective focuses on the development of an individual’s abilities and assets, for which skill development is perceived as a means of preventing problem behavior (Amodeo & Collins, 2007; Weiss et al., 2012). The PYD approach is rooted in a perspective that systematic change is possible during the course of development (Lerner et al., 2005), suggesting the power of youth programs to positively impact youth development and prepare youth for adulthood (Anderson-Butcher, 2005). Lerner and colleagues (2005) suggest that the potential for systematic change is made possible by the dynamic interaction between the internal and external influences exerted upon youth (previously described as the ecological
Based upon the assertion that youth development has a plasticity quality, how youth spend their time outside of school is important (Durlak & Weissberg, 2007; Halpern, 2000; Huebner & Mancini, 2003). It is estimated that more than 7 million youth are unsupervised at some point in time during their out-of-school hours (Durlak & Weissberg, 2007). Unstructured, unsupervised time can increase a youth’s exposure to risk, increasing the likelihood of problem behavior (Durlak & Weissberg, 2007; Halpern, 2000). Furthermore, Durlak and Weissberg (2007) assert that “young people benefit when they spend time engaged in structured pursuits that offer opportunities for positive interactions with adults and peers, encourage them to contribute and take initiative, and contain challenging and engaging tasks” (p. 5). Youth programs can provide structured out-of-school activities and experiences, exposing youth to interventions which foster PYD (Lerner et al., 2005; Pedersen & Seidman, 2005).

PYD has been described as “a process which prepares young people to meet the challenges of adolescence and adulthood through a structured, progressive series of activities and experiences which help them obtain social, emotional, ethical, physical, and cognitive competencies” (National Collaboration for Youth Members, 1998). PYD is also characterized by transfer of learning, for which the skills learned in one domain can be transferred to other domains in a youth’s life, enhancing adaptability and healthy living (Weiss, 2008). Exemplary youth development has been conceptualized by six C’s – competence, confidence, connection, character, caring, and contribution (Lerner et al., 2005; Vierimaa et al., 2012), all of which a youth program can enhance.
PYD programs emerged as a means to facilitate purposefully designed interventions, which consist of activities and experiences, to assist PYD and prepare youth for adulthood. To facilitate positive development and reduce problem behavior, PYD programs seek to enhance protective factors (e.g., develop competencies, strengthen caring relationships, create opportunities for prosocial activity) and reduce risk factors youth are exposed to (e.g., unsafe environment, community attachment, lack of social support; Anderson-Butcher, 2005). Participation in PYD programs has been associated with a plethora of positive outcomes, including increased school connectedness, effort, and academic performance (Anderson-Butcher, 2010; Anderson-Butcher, 2011; Anderson-Butcher et al., 2003; Durlak & Weissberg, 2007; Pedersen & Seidman, 2005), enhanced protection (Anderson-Butcher et al., 2003), improved feelings and attitudes (Durlak & Weissberg, 2007), behavioral adjustment (Durlak & Weissberg, 2007), increased self-esteem (Pedersen & Seidman), enhanced moral reasoning and prosocial behavior (Weiss, 2008), and overall well-being (Anderson-Butcher, 2011).

**Sport-Based Positive Youth Development**

**Youth Sport**

Youth sport is an important context for PYD (Riley & Anderson-Butcher, 2012; Weiss, 2008) and has evolved to encompass a broad framework consisting of organized sport, recreation, physical activities, and play (Anderson-Butcher, Riley, Iachini, Wade-Mdivanian, & Davis, 2011; Weiss, 2008; Weiss et al., 2012). The various forms of youth sport can be facilitated through a variety of youth programs (e.g., school, after-school programs, YMCA, community leagues, travel teams) and found in informal settings as
well (e.g., playgrounds, parks, open-gym, in the home; Anderson-Butcher et al., 2011; Sabo & Feliz, 2008).

In the United States, sport touches millions of youth, with about 54% of that population participating in organized sport and 46% participating in unstructured sport (Anderson-Butcher et al., 2011; Danish et al., 2005; Ewing & Seefeldt, 2002; Petitpas et al., 2005; Theokas et al., 2008; Weiss, 2008; Weiss et al., 2012). It is estimated that approximately two out of every three youth participate on a sports team, of which approximately one-third participate on three or more teams (Sabo & Feliz, 2008). These trends are steadily increasing, suggesting the broad reach of sport and potential utility to contribute to PYD (Anderson-Butcher et al., 2011; Brustad et al., 2001; Ewing & Seefeldt, 2002; Theokas et al., 2008).

Through sport, youth have the opportunity to engage in physical activity within the context of a social environment. Participation in youth sport has implications for physical, social, and psychological development which can lead to lasting positive effects (Anderson-Butcher, 2011; Anderson-Butcher et al., 2011; Larson et al., 2006; Petitpas et al., 2005). Positive outcomes associated with participation in youth sport include, increased health and quality of life (Baumert, Hendersen, & Thompson, 1998; Sabo & Feliz, 2008; Wickel & Eisenmann, 2007), athletic competence (Anderson-Butcher et al., 2013; Papacharisis, Goudas, Danish, & Theodorakis, 2007) social development (Anderson-Butcher, et al., 2011; Donaldson & Ronan, 2006), a sense of connectedness (Broh, 2002); academic achievement (Fredricks & Eccles, 2006; Sabo & Feliz, 2008), improved goal setting and problem solving (Papacharisis et al., 2007), as well as
decreased problem behaviors (Anderson-Butcher, et al., 2011).

Although sport is linked to a wide array of positive outcomes under optimal conditions, sport participation can also lead to negative outcomes (Papacharisis et al., 2007; Theokas et al., 2008; Weiss et al., 2012) such as increased anxiety and negative affect (Brustad, Babkes, & Smith, 2001), decreased motivation (Anderson-Butcher, et al., 2011), and increase in risky behaviors (Eccles et al., 2003). The determinants of sport outcomes are largely based on the sport program (e.g., philosophy, program design, staff) and individual youth experiences and less on sport participation as a whole (Papacharisis et al., 2007; Petitpas et al., 2005; Theokas et al., 2008). Therefore a sport program has the potential to lead to either positive or negative outcomes. It is critical that youth sport programs intentionally design and implement their program, including staff practices (i.e., coaching behaviors), curriculum, and environment created, to lead to positive outcomes.

**Positive Youth Development through Sport**

Youth sport programs have an advantageous platform as many kids are attracted to sport because participation is enjoyable and popular (Papacharisis et al., 2007; Petitpas et al., 2005; Weiss et al., 2012). That enjoyment may stem from sport interests, the inherent qualities of the activity, or the social aspect of playing with friends (Theokas et al., 2008). The wide-spread reach of sports creates a significant socialization opportunity, for which sport can be used to make an impact on the positive development of youth.

Sport is a powerful medium for positive youth development, as it can act as a “hook” to capture the engagement of youth while providing experiences which can
develop youth holistically (Hartmann, 2003). Through sport participation, youth can develop social, physical, and psychological skills that lead to a variety of positive outcomes (Anderson-Butcher, 2011; Anderson-Butcher et al., 2011 Weiss, 2008; Weiss et al., 2012). Youth sport programs have the opportunity to make a larger impact by extending their objectives past traditional sport skill development (Theokas et al., 2008). Sport is typically characterized by structured, interactive activities, with certain rules and norms (Theokas et al., 2008). The unique qualities inherent in sport (e.g., cooperative, active) can provide natural avenues through which to integrate life skill (e.g., teamwork, effort, goal setting) and moral (e.g., character, sportsmanship) development as well as address risk and protective factors, meeting the social, emotional, and physical needs of youth (Gould & Carson, 2008; Martinek & Hellison, 1997; Weiss et al., 2012).

Life skills are the internal assets that enable youth to achieve success in the different environments in which they live (e.g., home, school, peer group, neighborhood; Gould & Carson, 2008; Papacharisis et al., 2007; Theokas et al., 2008). These assets may be cognitive, behavioral, intrapersonal, or interpersonal (Danish et al., 2005; Gould & Carson, 2008; Theokas et al., 2008). Many life skills are transferable in that they can be applied in a variety of contexts (Gould & Carson, 2008; Papacharisis et al., 2007). Through sport, youth are provided opportunities to work on a team, learn how to play fairly, treat opponents with respect, display effort, leadership, and athletic competence, and positively interact with adults and their peers, all of which involve valuable life skills that can contribute to PYD.

Life skills act as protective factors which can both mitigate risk youth are exposed
to and contribute to resiliency. There is a prevailing notion held by many coaches and parents that life skills are naturally developed through sport (Gould & Carson, 2008). For example, within team-based sports, participants automatically learn teamwork by engaging in the sport experience. However, most life skills researchers and practitioners argue that life skills are not developed merely through sport participation (Danish et al., 2005; Danish & Nellen, 1997; Gould & Carson, 2008). Similar to sport skills, life skills must be intentionally taught through instruction, demonstration, modeling, and practice (Danish & Hale, 1981). Sport serves as the context through which life skill instruction can be facilitated.

When considering which life skills should be taught, youth development researchers have identified a core set of life skills. For example, Lerner and colleagues (2005), identified the six C’s of positive youth development previously outlined, including: competence, confidence, connection, character, caring, and contribution. Additionally, Bensen, Leffert, Scales, and Blyth (1998) proposed four major categories of internal assets: commitment to learning, positive values, social competencies, and positive identity. Although researchers suggest a core set of life skills, it is critical to keep in mind the influence of the environment. A target population’s needed life skills will look different based upon the age, culture, and economic status of the youth (Danish et al., 2005).

In general, an integration of competencies, including social, emotional, cognitive, and physical, is optimal to contribute to a youth’s ability to successfully adapt in adverse conditions (Catalano et al., 2004; Fraser et al., 2004). Competence may come naturally to
youth or require great effort, but in either case – competency can be developed. In addition to athletic competence, which is an inherent anticipated outcome of sport participation and reflects developed sport knowledge and skills, the current study will focus on the social competence, effort, teamwork, and self-control life skills that can be developed through sport-based PYD programs.

Social competence refers to one’s ability to interact with other individuals in a social context (Anderson-Butcher et al., in press; Martinek & Hellison, 1997), which can help youth build positive relationships with family members, peers, and community leaders, resulting in a social support system to assist them when faced with life’s challenges. The exertion of physical and/or mental energy into goal-directed behavior reflects effort, which can be seen through youth engaging and working hard. The effort life skill can have significant impact in alternative contexts (e.g., school, home, future job), contributing to one’s ability to be successful.

Similar to effort, teamwork is a critical life skill which can be transferred to sport and other contexts. Teamwork involves working collectively and cooperatively with others in order to achieve a common goal (Lower, Newman, Anderson-Butcher, in review). Lastly, self-control is essentially being aware and in control of one’s behavior and actions, reflecting a form of personal social responsibility. Many risk factors can be averted by youth demonstrating self-control. Sport-based PYD practitioners must intentionally teach (i.e., instruction, demonstration, modeling, practice; Danish & Hale, 1981) and reinforce life skills through sport in order for the development to be maximized.
Participation in sport can not only strengthen individual assets but also interpersonal, social, and environmental assets including family, peer, school, and community related strengths (Anderson-Butcher, 2011; Riley & Anderson-Butcher, 2012). Through sport, family members are provided the opportunity to support and interact-with each other. Parents can demonstrate care by attending youth sport activities, paying attention to their children, and getting involved, which also allows for parents to practice effective family management techniques (Anderson-Butcher, 2011). Communication, conflict resolution, and a feeling of connectedness can all be enhanced through family engagement in sport.

Communication and positive relationships can also be formed within the peer group through sport participation. Many youth engage in sport activities for the sake of seeking out friendships and the need to belong; the interaction that naturally takes place in sport can enhance a youth’s interpersonal skills and foster a positive social support system (Anderson-Butcher, 2011). One avenue in which youth sport participation takes place is in the school system. Schools offer opportunity for sport participation through physical education classes, extracurricular activities, and school sports teams. Through participation in school sports, youth develop a sense of connectedness to the school and often experience academic success due to the expectations associated with sport participation (e.g., minimal grade point average, attendance rate; Anderson-Butcher, 2011).

School sports and other youth sporting events not only influence the participants and their families but also the broader community (Anderson-Butcher, 2011).
Community members can engage in sport through volunteering, participating, or spectating. Athletic facilities are often made available to the community when unused, in addition to drawing sporting events which can boost the economic condition and overall climate of a community (Anderson-Butcher, 2011). By fostering family, peer, school, and community strengths, sport can contribute to PYD through multiple systems.

Several sport-based PYD models and programs have emerged in recent years with the goal of using sport as a vehicle for PYD. Examples of such programs include, SUPER (Theokas et al., 2008), LIFE Sports (Anderson et al., 2011), The First Tee (Weiss, Stuntz, Bhalla, Bolter, & Price, 2013), Girls on the Run (Gabriel et al., 2011), Harlem RBI (Berlin, Dworkin, Eames, Menconi, & Perkins, 2007), and Tenacity (Berlin et al., 2007).

To highlight two examples of sport-based PYD, SUPER is a sport program that uses the environment of sport as a “‘training ground’ for life” (Danish et al., 2005, p. 50). SUPER teaches middle and high school-aged students to use a variety of sport and life skills to improve athletic performance, recognize situations (both inside and outside of sport) in which these skills can be used, and then apply the skills within the sport context as well as transfer the skills to alternative contexts. The program consists of 18 modules (i.e., workshops or sport clinics) which incorporate three sets of activities: learning sport skills, learning life skills, and playing the sport. Each module lasts approximately 30 minutes and is led by college student-athletes. The SUPER program has been implemented for a variety of sports, including: basketball, soccer, golf, rugby, and volleyball.

Another example is Girls on the Run, a developmentally-focused youth sport
program (Gabriel et al., 2011). Girls on the Run uses sport as a vehicle to teach sport and life skills simultaneously in order to develop the physical, psychological, and social assets of 3rd through 5th-grade girls. The program consists of a 12-week curriculum, broken into three sections based on the physical, psychological, and social assets targeted. Each lesson lasts approximately 1.5 hours and includes an introduction to the lesson, stretching activities and processing the topic, a workout, cool down stretching and debriefing, and a closing with encouragement provided. The 12-week program is facilitated by a trained and certified coach and assistant coach and is designed to lead up to a culminating 5km running event.

In summary, sport-based PYD programs have a dual focus of teaching both sport and life skills concurrently, with the purpose of using sport as a vehicle for PYD (Anderson-Butcher et al., 2013; Petitpas et al., 2005). Youth development researchers contend that life skills cannot be caught, but must be taught (Danish et al., 2005; Gould & Carson, 2008; Theokas et al., 2008). To ensure the development of life skills through sport, sport-based PYD programs must intentionally imbed life skills content within instruction, demonstration, modeling, and activities (Danish & Hale, 1981). As PYD programs have the power to make a significant, positive impact on youth development, it is critical that programs intentionally design and implement targeted interventions and evaluate the effectiveness of the intervention to ensure maximum impact.

**Organizational Effectiveness**

Organizational effectiveness is an integral concept embedded within many program models and organizational theories (Cameron, 1986; Cameron & Whetten, 1983;
Goodman & Pennings, 1977; Rojas, 2000). With a constantly changing internal and external environment, service-based programs must embrace strategies and practices which lead to effectiveness in order to meet the needs of their target population (Hossein et al., 2011; Schreiber, 1983). It is critical for programs to evaluate effectiveness to ensure continual improvement and maintain accountability. In order to evaluate effectiveness, programs must critically consider what constitutes effectiveness for their specific program and how to measure that construct. By evaluating effectiveness, programs will be more equipped to justify its decisions and financial needs, as well as strategize for continual improvement and future growth.

Organizational effectiveness is a complex, paradoxical construct that has been described as ambiguous and conceptually in disarray (Cameron & Whetten, 1983; Chelladurai & Haggerty, 1991; Shilbury & Moore, 2006; Steers, 1976; Zammuto, 1982). Katz and Kahn (1971) propose that the term has come to represent a “totality of organizational goodness”, incorporating any and all elements that may convey organizational success or worth, resulting in an exhaustive construct. With various stakeholders vying for different, sometimes competing criteria of effectiveness, multiple perspectives and models of organizational effectiveness have emerged and operational forms (e.g., satisfaction, efficiency, growth) utilized (Cameron, 1986; Cameron & Whetten, 1983; Campbell, 1976; Campbell, Brownas, Peterson, Dunnette, & 1974; Hossein et al., 2011; Kahn, 1977; Katz & Kahn, 1971; Lewin & Minton, 1986; Shilbury & Moore, 2006; Steers, 1976; Zammuto, 1982).

Overall, there is a lack of consensus regarding a universal organizational
effectiveness construct. It is impossible to develop a universal conception of organizational effectiveness because the construct is continually changing (Cameron & Whetten, 1983) and has a distinct meaning for different types of organizations and programs (e.g., PYD program). Therefore, unique programs must determine a conceptualization of organizational effectiveness that aligns with the program’s mission and key stakeholders.

Models of Organizational Effectiveness

The organizational effectiveness construct is complex as a program consists of multiple constituencies, processes, and components, in which judgments of effectiveness can take place in each of these different domains. Several models of organizational effectiveness have been proposed in the literature. These models include goals, system resource, process, multiple-constituency, and competing values. Each model consists of distinct characteristics, strengths, and criticisms, with no optimal model identified in the literature, rather preferred contexts for each model.

**Goals model.** The goals model (similar to rational goal model) is a traditional measure of effectiveness, in which effectiveness is determined by the extent of goal attainment (Chelladurai & Haggerty, 1991; Hossein et al., 2011; Price, 1972; Seashore, 1983; Yuchtman & Seashore, 1971; Zammuto, 1982). This model focuses attention on the purposes of a program (Seashore, 1983) and the future condition a program aims to realize (Etzioni, 1971). Many constituents perceive goal attainment as the primary means of legitimizing the existence and activities of a program (Etzioni, 1971). The goals model has been criticized for its inability to distinguish goals (Hossein et al., 2011; Price, 1972);
prioritize goals (Hossein et al., 2011; Seashore, 1983); or address competing goals (Hossein et al., 2011; Seashore, 1983). Despite the weaknesses of the model, the goals model is the most widely accepted and utilized among all organizational effectiveness models (Chelladurai & Haggerty, 1991; Hossein et al., 2011).

**System resource model.** As not all goals can clearly be defined and measured, alternative models of organizational effectiveness emerged (Rojas, 2000). The system resource model (similar to open systems model) uses the criterion of bargaining position in the acquisition of scarce and valuable resources to evaluate organizational effectiveness, highlighting the interdependence of a program and its environment (Evan, 1976; Hossein et al., 2011; Shilbury & Moore, 2006; Yuchtman & Seashore, 1971; Yuchtman & Seashore, 1967; Zammuto, 1982). The model assumes that resources are a prerequisite for the attainment of program goals and therefore are a measure of effectiveness. The value of resources is based on their utility as a means to enable program functions and activities (Chelladurai & Haggerty, 1991; Yuchtman & Seashore, 1971). Etzioni (1971) includes the optimal distribution of resources to the various organizational needs and functions as an additional dimension of the system resource model.

A criticism of the system resource model is the supposed weak association between the acquisition of resources and organizational effectiveness. Resources may be required to attain program goals, but their acquisition does not guarantee the optimal efficient use of those resources, which is necessary in order to infer that greater resources equates to greater organizational effectiveness. Individuals intimately familiar with the
internal processes, functions, and activities of a program are best suited to determine appropriate resource distribution.

**Process model.** The process model (similar to systems model) links the goals model to the system resource model, with an emphasis placed on the internal processes of a program, which transform a program’s resources into the desired outputs (Chelladurai & Haggerty, 1991; Hossein et al., 2011; Steers, 1976). A program’s internal processes might include: implementation of policies, procedures, curriculum, and activities or the transactions between a program and its staff (Shilbury & Moore, 2006). Internal processes must be flexible and adaptable to environmental changes, such as low program attendance. The process model assumes that optimal internal processes can be determined. A criticism of the process model is its neglect of other important aspects of effectiveness (e.g., resources, outputs; Hossein et al., 2011).

**Multiple-constituency model.** The need to address the political nature of programs, when considering organizational effectiveness, lead to the creation of the multiple-constituency model (Hossein et al., 2011; Rojas, 2000). This model focuses on a program’s constituents’ definitions of effectiveness (Zammuto, 1982). The term constituency is intended to refer to any group or individual that may influence or benefit from the activities of the program. The multiple-constituency model is dependent upon a constituent’s ability to identify and articulate their preferences in regards to what constitutes organizational effectiveness (Cameron, 1986; Cameron & Whetten, 1983). Examining organizational effectiveness, using the multiple-constituency model, takes
into consideration the interdependence of an organization with their environment (Zammuto, 1982).

Controversy emerges regarding which constituents’ views of organizational effectiveness should receive preference (Hossein et al., 2011; Zammuto, 1984). Since it is difficult to claim any one perspective as correct and/or valuable over other perspectives, Chellardurai (2001) asserts that organizations should consider the effectiveness construct plural (i.e., effectivenesses) rather than singular. This argument suggests that an organization may be effective and ineffective simultaneously, depending on the distinct constituent perspective, which can often be conflicting.

**Competing values framework.** In an effort to create a robust model to assess the complexity of organizational effectiveness, the competing values framework (CVF) was developed (Hossein et al., 2011; Rojas, 2000). The CVF identifies three sets of competing values which drive organizational decision-making, these sets include developing people versus the organization (organizational focus), organizational stability versus flexibility (organizational structure), and organizational means (e.g., internal processes) versus ends (e.g., outputs; Hossein et al., 2011; Quinn & Rohrbaugh, 1981; Quinn & Rohrbaugh, 1983; Shilbury & Moore, 2006). The CVF suggests that an organization is effective when it satisfies several criteria of effectiveness, balancing the competing values (Shilbury & Moore, 2006). Similar to the multiple-constituency model, a criticism of the CVF is the need for a program to determine which constituents’ perspectives will receive preference, informing the values and criteria to focus on (Shilbury & Moore, 2006). Despite the criticism, the CVF can be considered the most
advanced organizational effectiveness model established in the literature, incorporating components from each of the previously discussed models.

**Evaluation of Organizational Effectiveness**

Program evaluation is a systematic process which determines the effectiveness of a program design and delivery as well as the achievement of anticipated outputs (i.e., goals; Caffarella & Daffron, 2013; Schalock, 1995). A purpose of program evaluation is to verify that objectives are being met, assess program implementation, identify necessary changes for improvement, and justify program existence (Chelladurai, 2001; Myers, 1999; Posavaac & Carey, 2003; Schalock, 1995). Programs can use the evaluative feedback to procure continued support from external donors and potentially substantiate the need for expansion.

The systematic process of evaluation begins with determining which model of organizational effectiveness will be adopted, which will in turn inform the criteria and standards of effectiveness evaluated. Once the criteria and standards of effectiveness are established, the criteria will be measured in the context of the program, and then analyzed through a comparison with the previously established standards. The program should then utilize the evaluative feedback as evidence of effectiveness and implications for continued improvement (Russ-Eft & Preskill, 2001).

When determining standards of effectiveness to use for evaluation, programs may adopt standards set by experts and professionals in the field, judgments of value established by constituents, or standards developed by the program’s directors. Experts and professionals in a related field may establish standards of effectiveness based on their
expert opinion and knowledge of what is effective (Chelladurai, 2001; Pace & Friedlander, 1978), focusing on specific structures, processes, and outcomes. A weakness of utilizing standards set by experts and professionals is the appropriateness of the standards.

A standard may be established within the field that is not appropriate for all programs due to the unique differences among programs or the amount of field exposure the professional is basing his/her judgment on. For example, a standard within PYD might consist of a specific amount of program dosage the youth are exposed to (e.g., duration of an intervention should be a minimum of 12 weeks). However, due to programs’ objectives and/or limited resources, a 12 week intervention may not be appropriate nor feasible. Therefore, programs may consider modifying standards, suggested by experts and professionals, to be relevant to the specific program being evaluated.

When utilizing constituents’ judgments of value, a program will be faced with the distributional issue associated with the multiple-constituency and CVF models of organizational effectiveness; essentially, whose value perspectives should receive preference. Programs should consider integrating this standard of criteria, in combination with other standards, into the evaluation process as many programs grow and thrive in large part due to the involvement of their constituent groups; therefore it is important to satisfy these groups to maintain their continued engagement.

In order to utilize effective and appropriate standards of evaluation, programs may take into consideration those set by experts, professionals, and their constituent groups.
However, many programs develop their own unique standards of evaluation, such as participation increases or the acquisition of a specific amount of funding. This standard of evaluation is developed in-house (or with assistance from external sources), by individuals intimately familiar with the program that are equipped and knowledgeable about what standards are reasonable, appropriate, and attainable. Programs likely use a combination of sources when establishing standards of evaluation that will be utilized. Each source has merit as well as weaknesses, which is why carefully considering and integrating standards of evaluation from the various sources may be the most effective means for a program to establish its standards of comparison.

**Organizational Effectiveness in Positive Youth Development Programs**

Within the context of PYD, a program is deemed effective if it is evidence-based. A PYD program is considered evidence-based if there is demonstration of the relevancy of the program, theoretical explanations for the design of the program, and program evaluation of program outcomes (Baldwin et al., 2005). An evidence-based program develops program components and content to address the malleable risk and protective factors of concern to their target population. Those program components and practices should be founded on theory and PYD best practices. Theory enables youth programs to connect program practices and curriculum to outcomes, by explaining how the specific practices/curriculum will lead to the desired outcomes of the program (which address the problem; Baldwin et al., 2005). Based on a review of 73 after-school programs, Durlak and Weissberg (2007) consistently found that programs that utilize an evidence-based approach produce multiple positive outcomes while programs that do not adopt such an
approach are less successful at producing youth outcomes, supporting evidence-based practice as a measure of effectiveness.

In evidence-based programming, logic models are often used to provide a framework for the process of connecting program components to outcomes (Baldwin et al., 2005). A logic model provides a conceptual outline of how the program components lead to proximal, distal, and ultimate outcomes (Baldwin et al., 2005). Theory is used to justify the connection between the successive steps. A model provides guidance for how the desired outcomes can be achieved.

To strengthen the evaluation of program effectiveness, a program should demonstrate both theoretical and empirical connections between the program components and program outcomes. This requires an investigation of the implementation of a program’s components as well as the resulting outcomes of the program. Evaluation of program outcomes alone is no longer an adequate measure of effectiveness (Mowbray, Holter, Teague, & Bybee, 2003). It is necessary to evaluate program components and outcomes in order to provide evidence which demonstrates the effectiveness of the program’s design and practices (Petitpas et al., 2005).

Understanding the programmatic mechanisms of PYD can help program directors more effectively design program components and strengthen the link between program components and outcomes. If a PYD program is designed in consideration of its target population and desired outcomes (relevancy), and incorporates the mechanisms that act as an agent of the intervention through which the outcomes are produced, the program has a greater likelihood of being effective and having a significant impact.
**Mechanisms of Positive Youth Development**

PYD theorists and practitioners have proposed several evidence-based practices and principles that act as mechanisms of PYD and make a significant enduring impact on youth participants. Illari and Williamson (2012) argue that a mechanism “consists of entities and activities organized in such a way that they are responsible for the phenomenon” (p. 120). In sport-based PYD, the phenomenon that programs are aiming for is PYD. Focusing on outcome evaluation to determine program effectiveness is no longer acceptable, as it merely determines the extent to which the phenomenon of interest (i.e., PYD) occurred (Mowbray et al., 2003).

Without a consideration of the mechanisms of PYD, programs have no way of knowing how they are responsible for PYD. Without knowledge of these mechanisms, a program is unable to strategically make informed decisions regarding program design and implementation, thus decreasing the likelihood that PYD will be maximized. Youth development researchers contend that an evidence-based program is one which has theoretical justifications for the program design, and evaluation of program outcomes (Baldwin et al., 2005). This argument should be extended further to incorporate evaluation of program components and outcomes in order to not only have a theoretical justification for the program design, but also empirical support for the program design. Therefore, the current study examined both the mechanisms and outcomes of the sport-based PYD program under investigation, to gain greater understanding of the mechanisms of PYD.

More specifically, the current study focused on programmatic mechanisms which
should be responsible for PYD. The comprehensive programmatic mechanism investigated was program dosage, which describes one’s degree of exposure to the program. Program dosage incorporates both quantity and quality of exposure (Nation et al., 2003). For the current study, dosage will be assessed by program implementation and program engagement. Therefore, the current study will focus on program implementation and program engagement as programmatic mechanisms of PYD within the context of a sport-based PYD program.

**Program Implementation as a Mechanism of Positive Youth Development**

Interventions have become a common practice among social programs aimed at prevention and youth development (Fraser & Galinsky, 2010). Program interventions are mechanisms intentionally designed and implemented to cause change (Fraser & Galinsky, 2010). Designing a program intervention includes the identification of principles, goals, and activities involved (Fraser & Galinsky, 2010). During the design phase of interventions, risk factors are introduced and matched with strategies that mediate the risk with the intent of producing positive outcomes (Fraser & Galinsky, 2010). For a program intervention to act as a mechanism of PYD, it must be purposefully designed and implemented. Intervention characteristics that have been found effective include programs that “(a) were comprehensive, (b) included varied teaching methods, (c) provided sufficient dosage, (d) were theory driven, and (e) provided opportunities for positive relationships” (Nation et al., 2003, p. 450).

Designing a theoretically and empirically supported program does not guarantee that the program will be implemented as intended or the desired outcomes will be
achieved (Anderson-Butcher & Case, 2010). In order for a program to have the greatest chance of reaching their desired outcomes, the program must be implemented as intended, consistently by all staff members (Fagan, Hanson, Hawkins, & Arthur, 2008). Additionally, the achievement of desired outcomes does not guarantee that the program was implemented as intended, nor provide any context for what mechanisms significantly contributed to the outcomes. An examination of program implementation fidelity can not only instill accountability, but also provide a stronger empirical connection between the program intervention and achieved outcomes.

Implementation fidelity. Program implementation fidelity refers to adherence to the program design and is reflected by staff members implementing the program as intended (Dusenbury et al., 2003; Fagan et al., 2008; Mowbray et al., 2003; Tucker & Blythe, 2008). Several dimensions of program implementation fidelity have been proposed in the literature, including adherence to the program, dosage, quality of program delivery, and participant responsiveness (Dusenbury et al., 2003; Fagan et al., 2008; Tucker & Blythe, 2008). When assessing implementation fidelity, a program should consider what elements of fidelity to measure, for which an examination of all dimensions of fidelity is encouraged for a comprehensive understanding of fidelity.

Of the fidelity dimensions, adherence to the program most closely reflects the definition of implementation fidelity and is consistently cited in the literature. Tucker and Blythe (2008) consider adherence to be “the degree to which a practitioner uses techniques and approaches prescribed by an intervention” (p. 185), while several other researchers focus on adherence to the program’s core components and content.
(Dusenbury et al., 2003; Fagan et al., 2008). A program’s critical elements may include the structure of the lesson or session, curriculum, environment, or strategies used to achieve the program’s objectives. Adherence may constitute completion of a task, directed effort towards an approach, or inclusion of the program’s key content and objectives within structured programming.

When designing programs, practitioners must be aware of the necessary dosage (i.e., intervention exposure) to produce the desired effect (Nation et al., 2003). Consideration of the number, duration, frequency, and intensity of the intervention sessions during program planning is critical as an intervention might not be enough (e.g., too brief or infrequent) to produce the desired outcomes or too much for which resources are used needlessly. A program with justifications for the number, duration, frequency, and intensity of their interventions must adhere to the intervention dosage established in order to increase the likelihood of achieving the program’s desired outcomes. Conversely, if a program establishes a specific amount of dosage necessary for subsequent outcomes and then fails to implement that degree of dosage, those outcomes are unlikely to be realized. Implementation of the program intervention, in accordance with the intended dosage, is a critical element of fidelity as a program’s intervention is one of the core components of a program which acts as a mechanism of PYD.

The quality of the delivery of the program is closely related to adherence to the program, with an emphasis placed on the effectiveness of the implementation. A practitioner may implement the program’s curriculum and teach all necessary content yet not achieve the desired outcomes of skill development, critical thinking, nor enhanced
knowledge. Practitioners may implement the intervention differently based upon their experience, strengths, and techniques, for which not all approaches will be effective and received well by the youth participants. Merely following the lesson plan is not enough to ensure the objectives of the intervention are met. There is a reactive component between the practitioner and the participants, which will determine if the intervention has the desired effect and thus was implemented effectively.

Participant responsiveness will also influence the effectiveness of an intervention. Youth may be misbehaving, unengaged, or uninterested, all of which will influence the quality of their exposure to the intervention and the resulting impact of the intervention. Participant responsiveness is considered a dimension of fidelity due to the significant influence it has on the effectiveness of a program. Practitioners can utilize strategies and techniques specific to keeping youth engaged, interested, and on-task, therefore influencing the quality of dosage the youth are exposed to and increasing the likelihood of outcome achievement.

The current study focused on the adherence to program components (i.e., program implementation) and participant responsiveness (i.e., program engagement) dimensions of implementation fidelity. These dimensions were selected based upon the primary research question concerned with how program dosage contributes to the program outcomes. The program implementation components that were assessed include intervention structure, curriculum, and climate.

**Intervention structure.** An intervention can be relatively simple or more convoluted; it may have multiple steps and/or phases, incorporate multiple components,
or focus on one primary component (Fraser & Galinsky, 2010). The intervention structure reflects the various phases and/or components within an intervention, intentionally organized in a specific order. A primary component of PYD interventions is activities, as most interventions include some form of action in order to create change and develop competencies. In addition to activities, many interventions incorporate instruction, through use of a curriculum, as well as a time for reflection or debriefing. Comprehensive programs are those that provide multiple interventions to address various risk and protective factors through increasing awareness and facilitating skill development (Nation et al., 2003).

During program design, the intervention structure is intentionally designed for the purpose of effectively developing skills. For example, introduction to the lesson should occur prior to the lesson’s activity and reflection over the lesson’s content should follow the completion of the lesson. Within each component of an intervention may be several steps that must be followed for effective implementation, such as when facilitating an activity the staff member may provide instructions, demonstrate the necessary skills, organize the youth, lead the activity, provide feedback, and then transition to the next phase of the intervention (Durlak & Weissberg, 2007). Adherence to the intervention structure is necessary as it primarily reflects implementation of the intervention.

**Experiential learning.** The sport-based PYD program under investigation adopts an experiential learning framework for their intervention. Experiential learning theory emphasizes experience as a critical part of the learning process (Kolb et al., 2001). Kolb (1984) defines experiential learning as “the process whereby knowledge is created
through the transformation of experience. Knowledge results from the combination of grasping and transforming experience” (p. 41). The experiential learning cycle contends that experiences are the foundation for observation and reflection. The learner will assimilate the observation and reflection into concepts which will inform implications for future action and experiences (Kolb et al., 2001).

The current study focused on the structure used for facilitating the experience, in relation to the intervention structure, rather than the progressive sequence of activities which speaks more to the intervention curriculum. To facilitate experiential learning, Schoel and colleagues (1988) propose a three-step process: 1) briefing, 2) activity, and 3) debriefing. During the briefing stage, the facilitator answers the questions: “What is going to happen?”, “How is it going to happen?”, and “What is expected?”. The objective of framing the experience is to provide purpose and focus, so that the participants do not engage in the experience in a random manner. During this process, the participants are thinking about and preparing for the experience, setting goals and visualizing what they are about to accomplish. Through briefing, the participants gain control over their purposeful involvement in the activity.

The next phase requires the facilitator to lead the activity (Schoel et al., 1998). Skills associated with leading an activity may include: instruction, empathy, competition, centering, conflict resolution, challenge, leadership voice, intervention, etc. The activity provides the experience during which participants observe and reflect. Throughout the activity, the facilitator should provide feedback which may include additional instruction, correction, and positive affirmation. The facilitator may need to refocus the group,
modify the activity, and/or manage difficult behaviors. Leading activities requires vigilance, as individual participants will respond to activities differently.

The final phase in facilitating an experience is debriefing, during which time the participants assimilate their observations and reflections into abstract concepts which can be transferred to future experiences and alternative contexts (Schoel et al., 1998). The debriefing should be an “experiential change process”, through which awareness of implications are reflected in action-oriented decisions and goals (Bacon, 1983; as cited in Schoel et al., 1998, p. 166). A sense of action is necessary for the learning to be transferred.

The debriefing sequence should follow three questions: “What?”, “So What?”, and “Now What?”. The “What?” begins the discussion with an easy review of what happened, which raises awareness about the behaviors developed that should be maintained or possibly changed. This descriptive question will naturally flow into interpretation, for which the “So What?” addresses what the participants gained from the experience, essentially what was learned. An evaluation of group and individual goals would occur during this part of the discussion. The “Now What?” provides the space to process how the lessons learned can be transferred to other contexts. During this time, participants should generate new goals which will complete the action-oriented debriefing process.

Assessing the effectiveness of experiential learning as a learning framework is critical to evaluate whether learning occurred and determine how the approach directly or indirectly caused the learning. A review of the scholarship on assessing experiential
learning effectiveness revealed an increasing concern for measuring and validating the learning environment (Gosen & Washbush, 2004). The article reviewed 80 studies on experiential exercises in a variety of learning environments (e.g., service learning, leisure and recreation courses, outdoor training programs, programs for problem adolescents, etc.), for which all studies found that learning took place. Experiential exercises were found associated with increased confidence, enjoyment, moral reasoning, group cohesion, self-regulation, self-esteem, and perceived skills. Overall, experiential learning outcomes are diverse as experiential exercises and environments can be adopted in a variety of settings.

Gosen and Washbush (2004) contend for the use of more rigorous research designs (e.g., pre-test and post-test, treatment and control group, random assignment) and objective measures when evaluating experiential learning. The learning measures should be defined, objective, and match the learning goals associated with the experiential exercise. Moreover, the measures must be valid, for which only five percent of the studies examined validated their research instrument (Gosen & Washbush, 2004). A challenge of evaluating the experiential learning framework is the complexity of learning. “Learning is an internal mental process, and what is learned and how it is learned is unique for each individual” (Gosen & Washbush, 2004, p. 284). Gosen and Washbush (2004) suggest that the concept of learning through experiential exercises has yet to be concretized, which highlights the need for further exploration of how experiential exercises mediate learning.
**Intervention curriculum.** The intervention curriculum outlines the content and skills the program is designed to teach. “Outcomes of [program] participation do not emerge automatically but must be nurtured through carefully designed curricula and teaching strategies” (Weiss, 2008, p. 437). The content may include life skills, athletic skills, or socio-cultural relevant topics of discussion. As previously mentioned, skills cannot merely be caught through participation, but must be explicitly taught by the instructors to ensure achievement of the desired outcomes. A curriculum provides the framework for staff members to consistently and effectively teach the program’s primary PYD constructs.

Effective PYD programs have been found to utilize a structured curriculum addressing at least five PYD constructs (Catalano et al., 2004). Programs should focus on constructs which are appropriate for their specific population, taking into consideration the age and culture of their participants as well as risk factors prevalent within the community. The program under investigation for the current study utilizes a curriculum which focuses on teaching self-control, effort, teamwork, social responsibility, and athletic competence.

A critical component of curriculum design is establishing the scope and sequence of the curriculum (ACT Department of Education and Training, 2009; Bisson, 1999). The scope of a curriculum describes the breadth and depth of the content to be delivered (ACT Department of Education and Training, 2009). For example, the sport-based PYD program under investigation delivers a curriculum that is comprised of nine sport-related
activities (i.e., breadth), for which the participants engage in five lessons, per sport activity, appropriate for a novice level (i.e., depth).

In addition to the curriculum scope, sequencing is an innate part of learning and development (Bisson, 1999). Schoel and colleagues (1988) define sequencing as “paying attention to the order of activities [content] so that the order is appropriate to the needs of the group” (p. 35). Bisson (1999) contends that there is no universal sequence, however most sequences share common developmental phases. For example, Tuckman (1965) identified four stages of group development, including: forming, storming, norming, and performing. Similarly, in sport sequencing involves linking one skill to another in a progression, resulting in the acquisition of the necessary skills for successful participation in a game-like situation (Clumpner, 2003). Clumpner (2003) proposes a sport progression of learning individual skills then partner skills, followed by small group skills and finally moving on to team skills as a means of reducing complexity and enhancing learning and development.

Sport-based PYD programs should deliver interventions that include a sequenced set of activities that facilitate active learning and focus on explicit skills and objectives to be achieved (Durlak & Weissberg, 2007). Using a sequence will allow complex skills to be broken down into smaller steps to ensure youth master the skill, enhancing development and self-efficacy. The sport-based PYD program under investigation uses an integrated sport and social skill curriculum to enhance the social and athletic competence of their youth participants, with social competence constituting the primary
objective. The curriculum is founded in cognitive behavioral theory, the TPSR model (Hellison, 2003), and the Sport Education model (Siedentop, 1998).

*Cognitive behavioral theory.* Informed by both cognitive (Dryden & Ellis, 2001; Ellis, 1962) and behavioral (Skinner, 1953; Watson, 1924) theories, the cognitive behavioral theory proposes that individual emotions and behaviors are determined by the way in which people think and appraise the world (Beck, Rush, Shaw, & Emery, 1979; Hupp, Reitman, & Jewell, 2012). According to Gonzalez-Prendes and Resko (2012), “cognitive processes, in the form of meanings, judgments, appraisals, and assumptions associated with specific life events, are the primary determinants of one’s feelings and actions in response to life events” (p. 14). Interventions which enhance social skills related to emotional regulation, problem solving, and coping can help alter emotions, change one’s thinking, and as a result change one’s behaviors. Intervention strategies focused on changing a youth’s thinking through feedback, reinforcement, and consequences can alter the youth’s behavior, thus leading to development.

Bandura (1971) extended cognitive behavioral theory by examining social learning, and in particular modeling. Social cognitive theory proposes that individuals learn through observation (Bandura, 1971). Youth who observe others modeling social skills and avoiding consequences of anti-social behaviors are likely to change their thinking on social skills and subsequent behaviors. Essentially, we learn through watching others. Additionally, social cognitive theory suggests that modeling is powerful, however a behavior is more likely to occur when there is a reward and/or reinforcement (Bandura, 1971). A program design which integrates rewards and reinforcement for
demonstrated social skills will more effectively change behaviors and ultimately influence social competence development.

The primary objective of the program under investigation is the development of social competence. Given the cognitive behavioral and social learning theoretical frameworks, there are several evidence-based strategies presented in the literature, which support the enhancement of social competence. Overall, there are four major components to social skills training, including: modeling; instructions and coaching; rehearsal and problem solving; and feedback and contingency management (Miltenberger, 2004; Pellegrini & Urbain, 1985). More specifically, Hupp and colleagues (2012) found self-monitoring, role playing, rewards, self-instruction, rational responses, positive thoughts, self-praise, mental rehearsal, and exposure to external stimuli to be effective strategies for social skill training. More specifically, effective programs have been found associated with a planned, focused sequence of activities to build social skills, which are active, engaging, and explicit (Durlack, Weissberg, & Pachan, 2010; Payton, Weissberg, Durlack, Dymnicki, Taylor, Schellinger, & Pachan, 2008). Furthermore, the most effective social skill interventions promote the “acquisition, performance, and generalization” of skills to other settings (Gresham, 1997, p. 233), which reflects the experiential learning framework (Kolb et al., 2001).

**Teaching Personal and Social Responsibility model.** The Teaching Personal and Social Responsibility (TPSR) model emerged in response to the attitudes, values, and behaviors of at-risk youth, with the purpose of developing character (Hellison, 2003). Hellison’s (2003) TPSR model is designed to develop the personal and social
responsibility of youth within the context of physical education, for which sport and social skill are taught simultaneously. Personal responsibility is reflected by two values, effort and self-direction, while social responsibility is related to the values of respecting the rights and feelings of others as well as caring and helping.

The TPSR model is divided into five developmental levels designed to foster personal and social responsibility (Hellison, 2003). Level one focuses on respecting the rights and feelings of others, incorporating lessons on self-control, peaceful conflict resolution, and including others. The second development level teaches participation and effort through lessons on self-motivation, effort, and persistence. Level three explores self-direction, teaching youth on-task independence, goal setting, and courage to resist peer pressure. During the fourth level, participants learn about helping others and leadership through lessons on caring and compassion, sensitivity and responsiveness, and inner strength. The final level provides an opportunity for participants to apply personal and social responsibility in other contexts, enhancing the transferability of learning outside of the gym.

Throughout the curriculum, the TPSR model emphasizes four themes to be applied consistently across the lessons – integration, transfer, empowerment, and teacher-student relationship (Hellison, 2003). Integration highlights the teaching of social and sport skills simultaneously, rather than separately. To successfully integrate the two sets of content, a teacher must be competent in teaching both physical activities and personal and social responsibility. In addition to integrating the curriculum, a teacher should create opportunity for youth to transfer the social skills learned to contexts outside of physical
education (highlighted by level five). Experiential learning provides an effective strategy for transfer through the use of debriefing at the completion of an activity (Schoel et al., 1998).

Within the TPSR model, Hellison (2003) integrates lessons on self-control, self-motivation, and self-direction to foster empowerment among the youth. Empowerment places the responsibility of demonstrating personal and social responsibility on the youth. A teacher can help guide the development of empowerment by gradually shifting power to the youth participant until they are taking on a leadership role. The final theme addresses the relationship between the teacher and youth participant. By respecting the strengths, individuality, knowledge, and capacity of each youth participant, a teacher promotes self-development and decency. This relationship also creates a caring climate which will entice greater engagement and empowerment. The sport-based PYD program under investigation utilizes the TPSR developmental levels and themes within their sport and social skill curricula.

A limited number of empirical studies investigating the TPSR model can be found in the literature (Hellison & Walsh, 2002). A review of the available research has found several positive outcomes associated with the TPSR model, including: motivation (i.e., attitude, positive affect; Newton, Watson, Kim, & Beacham, 2006); improved self-confidence, problem solving, interpersonal and intrapersonal skills, enthusiasm for learning, and concern for others (Cutforth & Puckett, 1999); increased commitment to the program (i.e., positive behaviors, emotional investment, attendance; Schilling, 2001); and improved self-control, effort, helping others, self-worth, self-direction, teamwork,
communication skills, interpersonal relations, sense of responsibility, sportsmanship, and physical skill (Hellison & Walsh, 2002).

Of the existing research on TPSR, even fewer empirical studies have examined implementation fidelity of the TPSR model, with Pascual and colleagues claiming to be the first to examine the connection between TPSR implementation fidelity and outcomes in 2011 (Pascual et al., 2011). Two empirical studies were found in the literature which examined TPSR implementation fidelity. Both studies adopted a qualitative approach (i.e., observations, interviews; Pascual et al., 2011; Wright & Burton, 2008), highlighting the need for a quantitative examination of implementation fidelity to provide a more rigorous analysis of the connection between process and outcome evaluation. Overall, there is a need for more research evaluating the effectiveness of the TPSR model (Hellison & Walsh, 2002; Wright, 2009).

**Sport Education model.** The Sport Education model identifies key features of sport curriculum and instruction appropriate for, but not limited to, institutional sport (Siedentop, 1998). The central features of the model include: seasons, affiliation, formal competition, record keeping, culminating events, and festivity. Seasons reflect the length of a sport unit, for which Siedentop (1998) argues that a curriculum with fewer activities and greater depth result in better outcomes. Team affiliation is also an important component of sport, contributing to a participant’s positive sport experience and personal growth.

An inherent quality of sport is competition, this sport education feature provides opportunities for participants to set goals and engage in self-directed behavior toward
desired outcomes (Siedentop, 1998). Within competition there is record keeping (e.g., points scored, rebounds, steals), a means to evaluate individual and group progress. Records provide feedback on goal progression as well as help to define measurable goals.

Closely related to formal competition is culminating events, such as tournaments, finals, and championships. A culminating event allows for celebration and recognition of accomplishments as well as an evaluation of one’s progress towards their desired outcomes (Siedentop, 1998). Often associated with culminating events are festivities, which consist of celebration for effort, sportsmanship, and personal improvement. The recognition of goal achievement can increase motivation, engagement, and persistence in sport. The Sport Education model asserts that the integration of these six critical features into physical education curriculum and instruction will result in greater educational outcomes. In consideration of the sport-based PYD program under investigation, the Sport Education model was integrated into the sport-related curricula.

The Sport Education model has primarily been studied within the context of physical education and institutional sport (Presse et al., 2011; Wallhead & O’Sullivan, 2005). Within the school-based context, the Sport Education model has been found associated with a myriad of positive outcomes, including psycho-social support and relatedness (MacPhail, Kirk, & Kinchin, 2004; Perlman, 2010; Perlman & Goc Karp, 2009), higher self-determination (Perlman & Goc Karp, 2009), team affiliation (Hastie & Sinelnikov, 2006), higher perceived competence (Hastie & Sinelnikov, 2006; Wallhead & Ntoumanis, 2004), higher enjoyment (Perlman, 2010; Wallhead & Ntoumanis, 2004), higher intrinsic motivation and task orientation (Spittle & Byrne, 2009).
Much of the research examining the Sport Education model took a comparative study approach in which two different instructional approaches (Sport Education model versus traditional instruction) were implemented in order to determine the outcomes of adopting the Sport Education curricula model (Hastie & Sinelnikov, 2006; Ko, Wallhead, & Ward, 2006; Parker & Curtner-Smith, 2005; Perlman, 2010; Perlman & Goc Karp, 2010; Spittle & Byrne, 2009). Metzler (2001) argues that empirical studies of curricular models must first verify that the model was implemented with high fidelity. In concert with Metzler’s assertion, these comparative studies have each verified implementation fidelity, strengthening the studies’ resulting implications (Hastie & Sinelnikov, 2006; Ko et al., 2006; Parker & Curtner-Smith, 2005; Perlman, 2010; Perlman & Goc Karp, 2010; Spittle & Byrne, 2009).

To date the Sport Education model has been found as an effective mechanism for youth development within the context of physical education. Wallhead and O’Sullivan (2005) argue that sport education has yet to “promote social transformation in the wider world of sport” as it has primarily been confined to institutional sport. Therefore, there is a need to not only adopt this model but also study its resulting outcomes in alternative contexts, such as sport-based PYD, to extend the body of knowledge and moreover contribute to broader social change.

**Intervention climate.** A program’s climate is formed by the structure, values, culture, and staff practices of the program. Several models have been proposed, identifying critical components of learning environments and PYD settings. Allodi’s (2007) model of quality learning environments identifies 10 key components, including:
creativity, stimulation, achievement, efficacy, safety, control, helpfulness, participation, responsibility, and influence. Eccles and Gootman (2002) proposed eight environmental features which maximize PYD, consisting of physical and psychological safety, clear and consistent structure with adult supervision, supportive relationships, opportunities to belong, positive social norms, support for efficacy and mattering, opportunities for skill building, and integration of family, school, and community efforts. Anderson-Butcher and colleagues (2011) contribute to this model with a focus on sport-based PYD, incorporating the following features: teach both sport and life skills, promote a mastery motivational climate, foster initiative, be youth-centered, use data to drive decision-making, and be enjoyable and fun.

The first consideration a program must make is in regards to youth safety. Physical and psychological safety is a necessary prerequisite for positive and healthy development (Anderson-Butcher et al., 2011; Eccles & Gootman, 2002; Weiss, 2008). Violence and unsafe conditions cannot only directly impact the physical well-being of a child but also indirectly interfere with their social, psychological, and emotional development (Eccles & Gootman, 2002). If a child feels unsafe, their attention, engagement, psychological well-being, relationships, and persistence will be negatively affected.

Sport-based PYD program settings must protect against unsafe conditions and violence (Eccles & Gootman, 2002). Facilities should be free from hazardous environments and equipment to reduce the possibility of injury or harm (Anderson-Butcher et al., 2011). Additionally, physical and psychological safety can be promoted
through consistently enforced rules and expectations, in which bullying, harassment, and violence are addressed with appropriate consequences (Anderson-Butcher, 2010; Pedersen & Seidman, 2005; Weiss, 2008).

Once a stable environment is created, programs can foster a positive psychosocial climate through the interaction of program staff and youth within the learning environment (Allodi, 2010). Through good communication, warmth, and developing authentic relationships between the staff and youth as well as among the youth, a program can create a positive, caring climate. Noddings (1984, 1992, 1995) distinguishes four dimensions of caring, including: care-giver is open and unbiased, caregiver is nonjudgmental, care-giver demonstrates concern and empathy for the care-receiver, and care-giver prioritizes the needs of the care-receiver. Caring-based programming has been found associated with perceptions of a caring climate, empathetic concern, and future expected participation, as well as lower levels of a perceived performance motivational climate (Newton et al., 2007).

Positive relationships is a critical contributor to the creation of a caring climate. Positive peer relationships have been found associated with continued sport engagement and the adoption of pro-social norms (Eccles et al., 2003; Ullrich-French & Smith, 2009). Additionally, supportive relationships with a caring adult can help youth feel greater security when taking on challenges (Larson & Walker, 2005); enhance self-concept, motivation, and performance (Allodi, 2010); develop a sense of belonging and increase engagement in positive behaviors (Anderson-Butcher, 2010). Relationships that are
youth-centered, stable, and supportive have been found to be critical mediators of PYD (Eccles & Gootman, 2002).

A sense of belonging and connectedness to one’s peers and adult leaders as well as to the program also contributes to a positive, caring climate. Perceived belonging can lead to greater investment and engagement in the program, stronger relationships, and inclusivity of diverse cultures (Anderson-Butcher, 2010). Program staff can provide social opportunities for youth to interact and develop relationships as well as foster team affiliation through enhancing a sense of identity and attachment to the group (Anderson-Butcher, 2010).

There have been inconsistent findings regarding the impact of staff practices on program outcomes. For example, a study of the influence of staff practices (i.e., emotional support, structure, autonomy support) on the outcomes of a sport-based PYD program found a non-significant predictive relationship between observed staff practices and program outcomes (Riley, 2013). To explain the non-significant association between staff practices and program outcomes, the researcher identified the limited number of observations as a major limitation in the study, suggesting the need for a more comprehensive review of staff practices. Longitudinal research of staff practices can provide a greater understanding of the environment created and resulting impact.

Another critical component of a positive environment is the motivational climate, which can significantly influence the values and outcomes of the participant. Achievement goal theory is a central motivation framework, which provides a basis for understanding an individual’s behavior and evaluation of success in an achievement
context (LeUnes, 2008; Roberts, 2001). Achievement goals have been distinguished by how competence is defined (Elliot, 2005), for which Nicholls (1984) postulated two conceptions of ability – mastery and performance. A mastery achievement goal emphasizes effort, task mastery, and personal improvement, whereas a performance achievement goal focuses on interpersonal comparison and competition (Elliot, 2005).

Motivational climate reflects the achievement goal structure of the environment of a specific achievement activity (e.g., sport; Ntoumanis & Biddle, 1999). A motivational climate is determined by situational cues (e.g., system of evaluation, basis for recognition, nature of interaction) which influence the youth’s perception of what constitutes success and failure (Duda & Balaguer, 2007; Miller, Roberts, Ommundsen, 2004). These situational cues can be a result of the program design or staff practices. For example, a program designed for all participants to equally engage in the activity, with staff who recognize and affirm hard work, personal growth, and the demonstration of competence, emphasizes a mastery-oriented motivational climate which will influence the achievement goals of the individual participants.

A mastery motivational climate is strongly supported in the literature and describes a “setting in which learning and skill development are salient and valued, and the prospect of outperforming others is not of direct interest to the individual” (Miller et al., 2004, p. 194). When mastery is emphasized, participants have been found to exert more effort and be intrinsically motivated (Ntoumanis & Biddle, 1999). Additionally, a mastery-oriented climate has been associated with perceived competence (Barkoukis et al., 2008; Boyce et al., 2009), commitment (Miller et al., 2004), self-regulatory strategies
(e.g., goal setting; Boyce, et al., 2009), respect for social conventions (Miller et al., 2004), enhanced enjoyment, (Barkoukis et al., 2008; Newton et al., 2006), continued involvement (Newton, et al., 2006), and reduced worry (Barkoukis et al., 2008).

Most research on motivational climate in sport and physical education examines the perceived climate at one point in time, with the assumption that this reflects the usual environment of a particular achievement context (Ntoumanis & Biddle, 1999). However, Ames and Archer (1988) suggest that the situational cues influencing the psychological climate are often inconsistent over time, which implies a dynamic motivational climate rather than fixed, contradicting the dominant methodological approach for measuring motivational climate. There is a need for longitudinal research which can capture changes in motivational climate over time to provide greater depth to our understanding of the impact of motivational climates (Ntoumanis & Biddle, 1999).

A caring, mastery goal-oriented climate that is physically and psychologically safe has been found to contribute to PYD (Weiss et al., 2012). Sport programs that focus on goal achievement, skill development, and personal improvement in a supportive environment can enhance a youth’s self-confidence and self-esteem, increasing the likelihood of persistence and positive development. Program staff members are a significant source of social influence (Vierimaa et al., 2012; Weiss, 2008; Weiss et al., 2012); through their leadership, practices, modeling, and feedback, staff can influence the program’s climate which directly effects a youth’s experience and outcomes (e.g., self-perceptions, behavior, connectedness).
Assessment of program implementation. It is of vital importance to assess implementation fidelity for a multitude of reasons (Dusenbury et al., 2003). First of all, a program which solely evaluates outcomes, failing to take program implementation into account, and claims effectiveness is possibly committing a Type III error in which the program is erroneously contributing the achieved outcomes to the intervention (Dobson & Cook, 1980). A program must validate that when properly implemented, the treatment will have positive effects on the participants (Dusenbury et al., 2003; Law & Shek, 2011). Moreover, the evaluation of program implementation fidelity will enable the identification of the distinct program components (e.g., instructional model, environment) which are effective or ineffective as well as highly adhered to or lack fidelity (Dusenbury et al., 2003; Law & Shek, 2011). An examination of program adherence will enable program administrators to understand how fidelity moderates the impact of the intervention. Such information will also provide insight into which program components are more feasible to implement, informing future program design, training, and support.

In order to understand whether a program is effective, why, and under what circumstances, programs must move past the traditional program evaluation approach of solely focusing on outcomes and integrate process evaluation of implementation fidelity (Harachi, Abbott, Catalano, Haggerty, & Fleming, 1999). Program implementation fidelity is critical within the field of PYD and has consistently been found associated with program effectiveness (Catalano et al., 2004). “High fidelity implementation helps increase the likelihood of desired participant outcomes, given evidence that closer adherence to core components can result in stronger participant outcomes, and that some
programs work only when implemented with a high degree of fidelity” (Fagan et al., 2008, p. 236).

In addition to ensuring the internal validity of a program, evaluation of fidelity can provide crucial information regarding which program components succeed or fail, are feasible or difficult to implement, or contribute the most to program outcomes, which in turn will inform program strategies (Dusenbury et al., 2003; Naleppa & Cagle, 2010). Programs with low implementation fidelity need to address why there is low fidelity in order for the implementation to improve. If programs are not implemented as intended, there is a reduced likelihood of outcome achievement, resulting in wasted resources and limited impact (Fagan et al., 2008). Programs with high implementation fidelity but failing to achieve the desired outcomes may need to redesign their program components in order to achieve effectiveness. Therefore, it is necessary to evaluate implementation fidelity for program accountability as well as to inform program planning.

Several methods to assess implementation fidelity have been adopted, however few standard tools have been established in the literature as measures of fidelity are often dependent on the specific program’s components, techniques, and approaches (Dusenbury et al., 2003; Fagan et al., 2008). Many fidelity assessment instruments are developed and tailored by specific programs to be appropriate for the content, activities, curriculum, and core elements of the program. The most prominent methods of assessing implementation fidelity include site observations and fidelity checklists completed by program staff members (Dusenbury et al., 2003; Fagan et al., 2008; Gottfredson et al., 2010; Mowbray et al., 2003).
Site observation and fidelity checklist tools used to measure implementation fidelity are typically tailored for specific programs and individual lessons or activities. Ideally the content should include all dimensions of implementation fidelity including adherence to program structure, dosage, quality of delivery, and participant responsiveness. Additional methods of assessment include interviews, staff surveys, participant surveys, and evaluation of project documentation and/or client records (Dusenbury et al., 2003; Fagan et al., 2008; Mowbray et al., 2003).

Existing research suggests that many PYD programs lack implementation fidelity. The National Study of Delinquency Prevention in Schools examined 3,691 school-based prevention activities and found that only half of the curriculum-based and one-fourth of the mentoring-based programs adhered to the intended dosage and only half of all programs adopted the recommended instructional strategies (Gottfredson & Gottfredson, 2002). Halfors and Godette (2002) surveyed 104 school districts across the United States and found that only 19% of schools implementing school-based prevention curricula demonstrated high fidelity.

Another example of poor fidelity within child development services was reported by the National Early Head Start Research and Evaluation Project (Paulsell et al., 2002). The project evaluated 17 funded Early Head Start research programs over a period of two years, for which five of the 17 programs were not able to reach full implementation (Paulsell et al., 2002). Dusenbury and colleagues (2003) identify a lack of resources (e.g., time, money) and an overwhelmed or unmanageable program as barriers to high fidelity. Ultimately, a lack of implementation fidelity indicates a gap between PYD research and
practice, reinforcing the need to assess and enforce fidelity (Fagan et al., 2008).

In summary, published empirical studies on process evaluation in the social sciences is limited with no standard methodology (Dusenbury et al., 2003; Harachi et al., 1999). Although there is empirical support for the benefits of participating in sport, few studies have examined both program implementation and outcomes in order to distinguish effective program strategies (Wright & Burton, 2008). Although it may be difficult to distinguish the process from the impact (e.g., separating a caring environment from perceived belonging; Hellison & Walsh, 2002), through careful and rigorous research design, researchers can gain a greater understanding of the programmatic mechanisms of PYD through sport.

**Program Engagement as a Mechanism of Positive Youth Development**

In addition to program structure, curriculum, and environment, a programmatic mechanism of great importance is that of engagement. Broadly defined, the term “engage” has several meanings that pertain to both the participant and program facilitator, including: 1) to mesh with; 2) to bring together; 3) to hold the attention of or give attention to; 4) to induce to participate; and 5) to enter into and carry on an activity (“Engage”, n.d.). Program engagement involves an interconnection between the intervention, participant, and program facilitator. The program facilitator brings together the participants in a group, inviting them to enter into an activity, for which the participants respond by exerting time and energy into the activity in a cooperative effort.

Program engagement places multiple responsibilities on the program facilitator, which include providing and facilitating the activity (i.e., intervention), developing
psychological safety, group cohesion, and a sense of belonging, as well holding the attention of the participants and inducing their participation in the activity. In relation to the role of the participants, they respond to the invitation from the program facilitator through their behavior. Essentially, the participants exert energy and effort into the activity, interacting with their peers and the leader, and giving their attention to the program facilitator.

In order for youth to experience positive change from a program intervention, they must be exposed to the intervention (Anderson-Butcher et al., 2003). Program engagement provides exposure to the program’s intervention and consequently develops the participant as a result of that exposure. “Organized activities are unique learning structures that provide opportunities for growth and development and that more time spent in activities provides an index of socialization experiences as well as greater absorption of skills” (Bohnert et al., 2010, p. 585). However, merely attending a program does not guarantee quality exposure to the program intervention. Youth who are disengaged, disruptive, aloof, disinterested, or resistant to participate will not receive the full measure of the intervention. For a comprehensive understanding of program engagement, both participant and staff behaviors should be considered.

**Participant engagement.** Youth participation in a program can be conceptualized by quantity and quality. Quantity of engagement can be measured by breadth (total number of different activities engaged in), intensity (average amount of time spent engaged in the activities), and duration (length of time engaged in the program; Bohnert et al., 2010). While quality of engagement includes three primary dimensions, including
behavioral, emotional, and cognitive engagement in the program, for which the current study focused on behavioral engagement (Bohnert et al., 2010; Fredricks, Blumenfeld, & Paris, 2004; Mahoney, Parente, & Lord, 2007; Roth et al., 2010; Weiss et al., 2005). “Behavioral engagement is defined as active participation and includes factors such as effort, concentration, attendance, following the rules, and avoiding trouble” (Bohnert et al., 2010, p. 593). Emotional engagement reflects one’s reaction to the staff, peers, and program activities, while cognitive engagement refers to one’s investment in learning (Bohnert et al., 2010).

Empirical research examining the developmental outcomes of participant involvement in organized activities initially treated involvement as a dichotomous variable by comparing participants and non-participants (Bohnert et al., 2010; Weiss et al., 2005). This framework is based on an assumption of homogeneity of involvement across participants, which fails to address the variation of engagement and individual differences in involvement. For a better understanding of participant engagement, an examination of the dimensions is warranted.

Within the PYD context, participant engagement has predominately been assessed utilizing the quantitative conceptualization of engagement, indicating a need for examination of quality of engagement within this setting. Measurements of quantity of engagement at the individual level have included: self-reported attendance (e.g., attendance during past seven days; Anderson-Butcher et al., 2003; Anderson-Butcher et al., 2004; Anderson-Butcher & Cash, 2010; Anderson-Butcher & Fink, 2005), attendance recorded by the program (Anderson-Butcher & Cash, 2010; Anderson-Butcher & Fink,
There have been inconsistent findings in the literature in relation to outcomes of engagement in a PYD program. Positive outcomes found to be associated with quantitative engagement include: self-concept (Anderson-Butcher & Case, 2010), increased altruism (Bohnert et al., 2010), enjoyment, effort, and motivation in school (Anderson-Butcher et al., 2003; Bohnert et al., 2010), academic performance (Bohnert et al., 2010; Roth et al., 2010; Weiss et al., 2005), self-esteem and emotional well-being (Bohnert et al., 2010), psychological and social adjustment (Fredricks & Simpkins, 2012), self-regulation (Mueller, Phelps, Bowers, Agans, Urban, & Lerner, 2011), leader support and a sense of belonging (Walker & Arbreton, 2004), decreased truancy (Anderson-Butcher et al., 2003), decreased drug and alcohol use (Fredricks & Simpkins, 2012), decreased antisocial behavior (Fredricks & Simpkins, 2010), and decreased depressive symptoms (Bohnert et al., 2010). In light of the positive outcomes found related to PYD program engagement, several studies and reviews have found engagement unrelated to program outcomes (Anderson-Butcher et al., 2004; Anderson-Butcher & Fink, 2005; Bohnert et al., 2010).

Focusing on time spent engaged in an activity fails to address the quality of that engagement. Fiester (2004) describes participation as “active enrollment”, which suggests additional responsibility on the part of the participant (i.e., behavior, emotion), more than just showing up. In relation to out-of-school-time programs, Weiss and
colleagues (2005) argue that “being there keeps youth safe, but being engaged enables them to grow” (p. 20). Scholars are finding that that attendance alone is insufficient for developmental outcomes, suggesting that benefits are dependent on level of engagement (Mahoney et al., 2007). Although most participation research has focused on attendance as the singular dimension of youth involvement, to be able to fully understand participation in order to intervene and improve said participation, the quality of engagement must be considered.

Of all the dimensions of participant engagement, quality of engagement (behavioral, emotional, cognitive) is the least researched (Roth et al., 2010; Weiss et al., 2005). Roth and colleagues (2010) who conducted a review on the relationship between program participation and developmental outcomes described measurement on the quality/level of engagement a rarity. The review analyzed 35 studies related to participation and developmental outcomes, for which only two studies investigated quality of engagement, finding a significant association between engagement and academic performance as well as peer relations. The researchers suggest a linkage between participant engagement and program quality (e.g., social climate, leader support), contending for longitudinal research of program engagement and quality to better understand this linkage. As program quality and staff practices is suggested to influence and interact with participant engagement, program features which engage participants should be considered.

**Program features that engage participants.** Within PYD literature, there are several evidence-based program design features suggested to enhance the effectiveness of
a PYD program. Several of these key program features are critical for youth engagement, including physical and psychological safety, relationships with caring adults, and pro-social opportunities (Weiss et al., 2005). “Social learning theory, social control theory, and social identity theory, point to the importance of belonging, commitment, bonding, and identification within social structures and how these factors impact opportunities to engage in, promote actual involvement in, and provide perceived rewards for engaging in pro-social or anti-social behaviors” (Anderson-Butcher & Fink, 2005, p. 11).

Through engagement in a program, youth interact with the program staff and their peers during activities and unstructured play. Youth often develop a sense of belonging to the program and within their peer group as a part of continued engagement in the program (Anderson-Butcher & Fink, 2005). Perceived belonging, relatedness, and leader support not only relate to program quality but also emotional engagement. Positive reactions to program staff, peers, and activities are indicators of emotional engagement which consequently influence developmental outcomes (Bohnert et al., 2010).

**Belonging**. Developing a sense of belonging within a program is a PYD evidence-based practice frequently cited and supported in the literature (Anderson-Butcher & Fink, 2005). Belonging, attachment, bonding, and commitment are constructs often used interchangeably or in congruence with each other within the PYD domain as the constructs are somewhat similar and related (Anderson-Butcher & Conroy, 2002). Belonging and relatedness are innate human needs, reflecting a youth’s connection or bond with the program, where youth participants feel they are valuable members of the program (Anderson-Butcher & Conroy, 2002; Anderson-Butcher & Fink, 2005; Deci &
Ryan, 2000). Program staff that are inclusive, welcoming, equitable, and youth-centered will create an environment in which youth feel comfortable and supported which leads to attachment (Eccles & Gootman, 2002). Anderson-Butcher and Conroy (2002) suggest that belonging incorporates elements of commitment, engagement, and connectedness. Moreover, belonging and relatedness produce a positive emotional response which contributes to motivation to engage (Martin & Dowson, 2009).

Belonging has been found associated with several positive outcomes, supporting its inclusion as an effective PYD feature. Youth have been found to benefit more from their engagement in a program when feeling a sense of belonging with the program (Anderson-Butcher & Fink, 2005). Continued participation becomes increasingly important as youth mature, due to trends of diminished engagement as youth age. Belonging has also been found to predict perceptions of opportunity for prosocial engagement and rewards for prosocial behavior (Anderson-Butcher & Fink, 2005). Furthermore, perceived belonging has been found a significant predictor of social skill development and athletic outcomes within the context of a sport-based PYD program (Anderson-Butcher, Riley, Amorose, Iachini, & Wade-Mdivanian, 2014; McDonough et al., 2013). Youth who feel a sense of peer group belonging have been found to demonstrate fewer behavior problems (Newman, Lohman, & Newman, 2007). Past research suggests that perceived belonging fosters greater engagement, which enhances developmental outcomes (Anderson-Butcher et al., 2014).

**Leader support.** Closely related to a sense of belonging is perceived leader support. Positive, supportive relationships with adults and peers, along with a sense of
belonging, are critical contributors of youth development (Ullrich-French & McDonough, 2013). Supportive relationships can be described by care, warmth, acceptance, connectedness, support, and good communication (Bowers et al., 2012; Eccles & Gootman, 2002). Quality relationships include both emotional (e.g., care, warmth) and instrumental support (e.g., mentorship). Adult leaders at youth development organizations are given the responsibility of guiding, supporting, caring for, and mentoring youth participants. Leader support consists of a complexity of relational attributes and behaviors, however there is no ideal set of qualities. Research suggests that perceptions of leader support is more critical than specific supportive qualities, as youth often have different responses to an adult’s relational qualities (Eccles & Gootman, 2002).

Supportive staff-youth relationships are a key program feature, influencing participant engagement and resulting outcomes, thus acting as a mechanism of PYD. Positive relationships with a caring adult have been found associated with increased prosocial behaviors and decreased problem behaviors (Anderson-Butcher et al., 2004), as well as continued participation in a program (Ullrich-French & McDonough, 2013). In relation to emotional support, DuBois & Silverthorn (2005) found warmth, acceptance, and closeness related to psychological well-being. Additionally, leader emotional support was found to moderate the relationship between autonomy support and social responsibility (McDonough et al., 2013). Overall, by fostering supportive relationships, program staff can enhance youth participants’ sense of belonging and create an environment which invites youth to engage.
In light of the limited research on engagement within the PYD context, more specifically the absence of measurements of quality of engagement, there is a need for a better understanding of youth engagement and program features which engage youth participants. While it is important to collect attendance data for program evaluation purposes, youth engagement can significantly vary among those who are in attendance. The interaction of youth participants’ engagement, perceived belonging and leader support will influence the quality of exposure to the program intervention, which consequently influences their developmental outcomes. In summary, the current study will investigate program implementation fidelity and engagement in congruence with outcome evaluation, in order to gain a greater understanding of the interaction of program quality and engagement and how these constructs mediate PYD, which has both practical and theoretical implications.
Chapter 3: Methods

Overview

The current study was designed to assess program dosage as a mediator of change in participant outcomes, within the context of a sports-based PYD program. To measure both program dosage and participant outcomes, two separate populations were studied through use of survey research. An instrument was developed to evaluate the program implementation by program staff, as a measure of program dosage. A second instrument was developed to evaluate perceived social and athletic competencies among youth participants, as the PYD program is designed to enhance these competencies over the course of the summer camp. The youth instrument also assessed mechanisms of PYD which contribute to program dosage, such as behavioral engagement and sense of belonging.

The instruments were administered multiple times throughout the summer program to gain an accurate assessment of program dosage and change in youth outcomes. Following data collection, the survey responses were entered, screened, treated, and reduced to allow for statistical analysis. Structural equation modeling (SEM) was conducted to test the hypothesized relationships among the variables of interest, resulting in theoretically and empirically supported mediated models.
Research Context

Sport-Based Positive Youth Development Program

The sport-based PYD program that was investigated in the proposed study was a 19 day summer camp designed to develop social and athletic competence, among disadvantaged youth, through intentional, structured programming (Anderson-Butcher et al., 2011). The summer program is free of charge, for which approximately 600-800 youth, ages 9-15, register every year from the local community. The primary objectives of the program are to increase social competence, self-control, effort, teamwork and social responsibility among the youth, as well as increase the youths’ connectedness to the program and staff (Anderson-Butcher et al., 2011). In addition, the program’s secondary objectives include increasing the youths’ athletic competence and exposure to college as well as referring youth to other opportunities in the community (Anderson-Butcher et al., 2011).

To realize the program objectives, the intervention has a dual focus for which sport and social skills are taught simultaneously through sport and play-based educational activities (Anderson-Butcher et al., 2011). Every lesson and activity has a specific curriculum which outlines the skills to be taught and the process through which to teach those skills. The program adopts Durlak and Weissburg’s (2007) principles of youth development skills by designing the curriculum to teach the sport and social skills in small, sequential steps. The lesson plans are divided into three progressive phases, which include framing of the sport and social skills, followed by facilitation of an activity which allows youth to practice the skills and receive positive, detailed feedback, and ending
with a debriefing to review the skills and discuss how to transfer the skills to other contexts.

The four primary social competence skills taught include self-control, effort, teamwork, and social responsibility, which have been collectively coined “S.E.T.S.”. Self-control can be defined as one’s ability to control their behavior (e.g., choosing to follow the rules; Gresham & Elliott, 1990). The program teaches youth participants to demonstrate positive behavior, listen and communicate effectively, understand the consequences of their actions, and refrain from poor behavior. Effort can be described as one’s exertion in an activity; the youth are taught to work hard and try to achieve their goals, which are vital skills for future success. Teamwork is an inherent quality in group activities and team sports and involves working collectively towards a common goal. The youth are taught to value the contribution of others, problem solve, and understand the different roles within a team. The final social skill taught at camp is social responsibility which entails one’s contribution to society and adherence to social norms and expectations (Wentzel, 1991). The program teaches the importance of helping others and building friendships and helps youth acknowledge their contribution to the community.

Athletic competence, which reflects one’s sport abilities, is enhanced through sport instruction in nine different sport-related activities, including: basketball, football, lacrosse, soccer, social dance, softball, swimming, volleyball, and health and fitness. The social skill curriculum is embedded within the sport curriculum to further the primary and secondary objectives of the program which are to enhance the social and athletic competence of youth participants.
Camp Experience

The summer program lasted 19 days during the month of June, for which camp was delivered Monday through Friday from 8:00 a.m. until 2:15 p.m. The youth participants were bused to the host institution’s campus every morning from various locations in the community and bused back to those locations in the afternoon. The youth were provided breakfast and lunch and were able to access the resources and facilities of the host institution throughout the day. Each day the youth engaged in six hours of programming with their group, for which groups were comprised of approximately 21-28 youth of the same age. A counselor was assigned to each group to supervise and mentor the group through the duration of camp.

The program intervention consisted of three, five-day rotations, for which each rotation included three, one-hour sport sessions facilitated by recreation leaders (two recreation leaders per sport, with the exception of swimming which included six recreation leaders for safety purposes) and a one-hour education session facilitated by an education leader. The youth participated in the same three sport-related activities for five days and then rotated to a different three sports and completed the final rotation with the last three sports in order to be exposed to all nine sport-related activities. Each day focused on specific sport and social skills, for which the first five days emphasized self-control, the next three days focused on teaching effort, the following four days provided instruction on teamwork, and the last four days of curriculum emphasized social responsibility.

The intervention curriculum was primarily designed to enhance social
competence, with a secondary objective of developing athletic competence. Based upon the cognitive behavioral and social learning theoretical frameworks, the education curriculum consisted of explicit instruction on the S.E.T.S. skills using sequenced activities and social cues. The social skills were reinforced through positive feedback and rewards, such as a button awarded to youth demonstrating S.E.T.S. Additionally, the social skills were reinforced through consequences for anti-social behaviors, such as removal from a session. The youth were provided opportunities to model and practice the social skills within the context of sport through rehearsal and problem solving challenges. Furthermore, the youth were able to transfer these learned behaviors to other contexts, such as the program’s sport clinics facilitated throughout the calendar year.

In addition to the curriculum, which consisted of sport and education lessons designed to provide sport and social skill instruction, the program facilitated a College Access Career Day and culminating Olympics event. On College Access Career Day, the youth attended vocational presentations of various available professions (e.g., nursing, fire department, teaching) and went on a tour of the host institution’s campus. The purpose of Career Day was to prepare the youth participants for their transition into adulthood. The Olympics event was a three-day culminating event where youth were given the opportunity to demonstrate their sport and social skills developed throughout camp. The event was youth-centered, for which the youth created team flags, determined what roles they would fulfill in the competitions (e.g., coach, athlete, referee), and invited their family and friends to attend (Anderson-Butcher et al., 2011). Ultimately, the
summer program is designed to be fun, and yet intentional, in order to attract youth participants and enhance positive development.

**Staff Training**

The program staff consisted of 77 individuals fulfilling one of six positions, described in Table 1. A comprehensive, two-day training, one week prior to the beginning of the summer program, was required for all staff members. The primary objectives of the training were to increase staff understanding of camp goals, objectives, and logistics, knowledge of youth development principles, ability to transfer youth development principles into camp, understanding of how to implement the curriculum, and ability to effectively manage camper behavior. The training lasted seven hours each day covering the diverse array of topics organized into two overarching themes – program and youth development.

**Table 1**

*Descriptions of program staff positions*

<table>
<thead>
<tr>
<th>Position</th>
<th>Primary Responsibilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Director</td>
<td>Oversee coordinators</td>
</tr>
<tr>
<td></td>
<td>Address logistical issues</td>
</tr>
<tr>
<td></td>
<td>Address major staff issues and child removal</td>
</tr>
<tr>
<td>Coordinator</td>
<td>Oversee staff within a specific area</td>
</tr>
<tr>
<td></td>
<td>Responsible for day-to-day responsibilities of staff</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Table 1 continued</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Administrator</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Recreation Leader</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Education Leader</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Counselor</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

On the first day of training, the program director introduced the camp mission and objectives, as well as outlined the roles and responsibilities of camp staff. The staff fulfilled one of three primary roles, including recreation leader, education leader, and counselor. The recreation leaders were responsible for coaching one of the nine sport-related activities, delivering sport and social skill instruction simultaneously. In comparison, the education leaders focused on coaching the social competencies through play-based activities. The counselor acted as mentor and supervisor for a group of youth,
directing the group’s schedule and enhancing the youths’ connectedness to the program. Following the camp introduction, the lead researcher provided an in-depth curriculum demonstration. The demonstration simulated a camp sport session and included elements of the sport and life skill instruction found in the curriculum (i.e., framing, facilitating, debriefing). The lead researcher facilitating the demonstration played the role of a recreation leader while the staff played the role of the youth in order to get the full experience of what camp should look like.

Following the camp curriculum demonstration, the counselor and youth leader coordinators reviewed key youth development principles to the staff, focusing on understanding individual difference, providing feedback, creating belonging and relationships, and motivating and engaging youth. As a majority of camp participants are disadvantaged youth from the local community, the staff were trained on protecting youth by the assistant director. This training included how to detect and report signs of abuse, how to address safety issues, and what safety guidelines to follow.

At the end of day one, the staff received information regarding program evaluation and research from the research coordinator. To ensure understanding of what participation in the research study entails, the research coordinator outlined the primary research agenda for the program, which included collecting data from the staff and the youth. The objectives of the research were explained as well as the role of the staff within the research study. The staff were provided an opportunity to review the staff instrument they were required to complete throughout the duration of camp. The research coordinator explained the individual items and scaling through relevant examples and
scenarios and answered any questions in order to increase the reliability of the tool and efficiency of data collection. In regards to the youth instrument, only education leaders received an opportunity to review the survey and ask questions as they were responsible for administering the tool.

The beginning of training day two consisted of a review of training day one and an opportunity for staff to ask questions, as well as a walking tour of camp facilities. Following the walking tour, the staff received training on positive behavior management and safety strategies and procedures from the behavioral support coordinator and assistant director. The staff were trained on how to manage youth behavior, which included reinforcing positive behavior and addressing poor behavior. Safety was another critical concern as the youth were required to travel across the host institution’s campus throughout the day, cross streets, and engage in physical activities with equipment which could potentially be dangerous (e.g., lacrosse stick). The staff learned appropriate safety procedures so that they could take proactive measures to protect the safety of the youth.

To further solidify staff knowledge and understanding of how to implement camp curriculum, the staff were divided by role (i.e., counselor, recreation leader, education leader) and engaged in a second curriculum demonstration (specific to their role) lead by the curriculum support staff. Similar to the first demonstration, the staff played the role of the youth and observed the curriculum support staff play the staff role. At the end of the demonstration, the staff evaluated the simulated camp session through completion of the staff instrument. The research coordinator facilitated a review of the staff’s evaluative responses, providing clarification regarding the interpretation and scaling of different
items to enhance the reliability of the tool. Towards the end of training day two, the staff were instructed on the role of the youth leader by the youth leadership coordinators and received an opportunity to meet the youth leaders who provided assistance throughout camp. The closing element of training included final details about camp logistics, such as attendance procedures, bus procedures, and where to park.

**Research Design**

To investigate the effectiveness and impact of a sport-based PYD program, the current study examined program outcomes, as well as the implementation of the program intervention, as a means of assessing program dosage youth are exposed. To adequately assess the outcomes and implementation of the program, a developmental, longitudinal research design was employed. Developmental research is designed to assess change over time, for which longitudinal research investigates the change or development of the same group of subjects by collecting data from those subjects over an extended period of time (Berg & Latin, 2004). Although there are several burdens and threats to internal validity associated with longitudinal research (Berg & Latin, 2004), this type of developmental design is most appropriate for assessing the impact of the program intervention.

As the program only lasted 19 days, many of the burdens and threats to internal validity were minimized, while still providing sufficient time for the intervention to produce change in the youth participants. The longitudinal research design provided valuable data regarding youth development over the course of the program. In regards to investigation of program implementation fidelity and program dosage, a longitudinal research design was utilized as data were collected from the same group of subjects (i.e.,
staff) multiple times throughout the course of the program. However, the primary purpose of this particular investigation was not to assess change in implementation over time, but rather to provide an accurate assessment of program implementation in order to determine degree of program dosage the youth were exposed to.

The primary variables under investigation included time, program implementation, youth outcomes, behavioral engagement, sense of belonging, and leader support. Based upon the longitudinal research design, the independent within-subjects factor for both the staff and youth studies was time. The youth instrument was administered at two time periods (i.e., pre-test and post-test), while the staff instrument was completed four times daily (total of 60 measures) for the 15 days of curriculum.

A critical, independent, between-subjects factor was program implementation. Average program implementation variables (i.e., structure – S.E.T.S. coaching, structure – sports coaching, curriculum, and environment) were created from the staff implementation fidelity data, which took into account days youth attended the program. In addition to the program implementation variables, behavioral engagement, sense of belonging, and leader support, were considered contributors of program dosage. Degree of exposure to the program intervention (i.e., program dosage) should influence the youth outcome change that is a result of participation in the program; therefore differences in program dosage were accounted for. The youth outcomes were the dependent variables as they were expected to change over time as a result of exposure to the program. The youth outcomes included social competence, effort, teamwork, self-control, and athletic competence.
Samples

Program Staff

The summer staff hired by the sport-based PYD program under investigation consisted of hourly employees and interns recruited through faculty and website postings at the host university, as well as through a career fair and sport camp fair hosted at the university, to work the summer camp. Each staff member fulfilled one of three primary roles: recreation leader, education leader, and counselor. The recreation leader was responsible for providing sport instruction for one particular sport. The sport instruction was designed to enhance the athletic and social competence of the youth. Recreation leaders were provided curriculum, which consisted of five progressive lesson plans. Facilitation of the lesson plans included framing the sport and social skills emphasized in the lesson, facilitating activities for youth to practice the learned skills, and leading a debriefing session to discuss how the skills can be used inside and outside of sport. The youth participated in three sport-based lessons per day with the recreation leaders and one education-based lesson facilitated by an education leader.

The education leader was primarily responsible for enhancing the social competence of youth through play-based educational activities. An education curriculum was provided which follows a similar lesson plan structure in that the education leader framed the social skill emphasized in the lesson, facilitated relevant activities to enhance learning, and reviewed the lesson through leading discussion. In comparison to the recreation curriculum, the education curriculum consisted of 15 lesson plans for which each social skill was taught over several sequential lessons. The counselor assisted the
recreation and education leader with facilitating the lessons and framing the social skills, but was primarily responsible for supervising and mentoring a group of youth. Counselors were assigned one group of youth (approximately 21-28 youth) for the entire summer program and provided constant supervision, guiding the youth through their daily schedule, building connectedness, reinforcing the social and sport skills taught, and following safety guidelines.

The staff population consisted of 52 individuals, of which 22 individuals fulfilled the recreation leader role, six fulfilled the education leader role, and 24 fulfilled the counselor role. The staff were hired as either interns or employees, for which 17.3% were interns \( (n = 9) \), 50.0% were new employees \( (n = 26) \), and 32.7% were returning employees \( (n = 17) \). Due to the limited number of staff members hired and the need to obtain an accurate account of the program dosage youth were exposed to, a census of the staff population was performed for which all staff were requested to participate in the study.

The majority of staff members were college students or recent college graduates (86.5%) as the summer program is hosted at the local university. Of the staff, 63.4% were college students \( (n = 33) \), 23.1% were recent 2014 college graduates \( (n = 12) \), and 13.5% graduated from college prior to 2014 \( (n = 7) \). The staff included 28 females (53.8%) and 24 males (46.2%), for which 84.6% ranged from 18 to 24 years of age \( (n = 44) \), with 15.4% between the ages of 26 to 50 \( (n = 8) \). The majority of staff identified themselves as White/Caucasian (63.5%; \( n = 33 \) ), 32.7% Black or African American \( (n = 17) \), and 3.8% Multiple Races \( (n = 2) \).
The program hired staff that were capable of implementing the dual-focused intervention which emphasizes teaching sport and social skills simultaneously by screening for applicants with both experience working with children and sport experience. Almost all staff had previous high school athletic experience (92.3%; \( n = 48 \)) and 34.6% of staff were current collegiate student-athletes (\( n = 18 \)). All staff were required to complete a background check and concussion in sports certification to ensure the safety of the youth.

**Youth Participants**

The sport-based PYD program recruited approximately 600-800 disadvantaged youth, ages 9-15, from the local community, to participate in the summer program. As research and evaluation is a critical element of the sport-based PYD program, a census of the youth registered for the summer program was performed, for which parents and guardians were asked to consent to their child’s participation in the study during registration for the program and youth ages 13-15 were asked to assent to their participation in the study on the first day of camp. Registration for the summer program was not contingent upon participation in the study.

Program registration consisted of 621 youth, for which 539 received consent from a parent/guardian to participate in the study. Of the 539 with parental consent, 417 completed both the pre-test and post-test survey (77.4% response rate). Data analysis was limited to those 417 youth who completed both surveys, as this study was designed to assess change over time. Of the 417 youth, 46.5% were returning campers (\( n = 194 \)) and only 41.0% reported a household income above the national poverty guideline (\( n = 171 \);
USDHHS, 2014), with 62.1% indicating eligibility for a free or reduced school lunch \((n = 259)\).

The majority of youth self-identified as Black or African-American \((81.1\%; n = 338)\), 11.0% Multiple Races \((n = 46)\), 4.3% White/Caucasian \((n = 18)\), and 0.7% “Some Other Race” \((n = 3)\), with 2.9% choosing not to identify their race \((n = 12)\). There were a greater proportion of males \((62.6\%; n = 261)\) than females \((36.9\%; n = 154)\), with three individuals not reporting gender \((0.5\%)\). With respect to age, 13.2% identified as 9 years of age or younger \((n = 55)\), 17.0% age 10 \((n = 71)\), 17.5% age 11 \((n = 73)\), 18.7% age 12 \((n = 78)\), 18.7% age 13 \((n = 78)\), 11.3% age 14 \((n = 47)\), and 3.1% age 15 \((n = 13)\), with two individuals not identifying their age. Out of the 19 days of camp, 88.0% of youth attended at least 16 days \((n = 367)\), with 37.4% attending all 19 days \((n = 156)\).

Validity and Reliability

It was critical to address threats to the internal and external validity of this study in order to minimize those threats prior to the administration of the study. Internal validity is necessary for the results of a study to be relied upon (Ary, Jacobs, & Razavieh, 2006). In regards to the current study, it was imperative to reduce threats to internal validity to strengthen the connection between the program intervention and resulting youth outcomes.

As the intervention lasted 19 days, for which the youth data were collected at the beginning and end of the intervention, several conditions and events may have threatened the internal validity of the study. During the intervention, the youth were taught sport and social skills to enhance their competencies, for which the youth outcomes reflect these
competencies. A history effect may have occurred for which events external to the intervention, such as engaging in alternative PYD programs or sport activities, influenced the youth outcomes associated with the program under investigation. Maturation may also have influenced youth outcomes, as physical or mental changes in the participant occur as a function of the passage of time. The length of the intervention was moderate therefore the influence of a history effect or maturation should be minimal (Baumgartner & Hensley, 2006).

The use of multiple measures may have influenced the internal validity of the study due to a possible testing effect for which taking one test can affect a participant’s scores on subsequent tests. The Hawthorne effect provides an example of this testing threat as participants may respond differently, such as providing socially acceptable responses, based on the knowledge that they are being tested; this was more likely to occur with the staff population who self-reported their degree of program implementation (Baumgartner & Hensley, 2006; Dunning, Heath, & Suls, 2004). Training the staff on the implementation tool, protecting their confidentiality through the use of identification numbers, and communicating the purposes of the study (e.g., the data will not be used for staff evaluation) should have reduced a possible Hawthorne effect.

Another example of a testing effect is in relation to practice or fatigue, for which the participants may respond differently in subsequent tests out of greater familiarity and comfortability with the test or out of fatigue from participating in multiple tests. The three week time period separating the pre-test and post-test for the youth population should have reduced the testing effects described. Although the use of a pre-test created
the potential for a testing effect, the pre-test was necessary for the developmental research design.

Another critical internal validity threat in longitudinal research is that of mortality, which reflects participant attrition. In regards to the program under investigation, it is common for several youth to drop-out throughout the summer program, with a few individuals removed for behavioral issues. Mortality typically results in youth failing to complete all measures (e.g., pre-test, post-test), which is particularly problematic if there is differential mortality in which the individuals from one group have greater attrition than other groups. Program administrators encouraged persistence and commitment to the program, addressing behavioral issues and concerns immediately to reduce drop-outs and ejections from both the staff and youth populations.

Non-response error is a threat similar to mortality in that data is not collected from all participants for all measures due to inability to locate subjects (e.g., subject is absent) or subjects failing to respond (e.g., staff failing to submit their responses at the end of the day; Lavrakas, 2008). Once again, the program administrators and researchers were persistent in trying to collect data from all subjects, which included monitoring data collection and following up on non-responses for three days past the original data collection time period. These actions moderately helped to control for the mortality and non-response error threats to internal validity, for which 67.2% of youth registered for the program completed all measures ($n = 417$).

Throughout the program, the recreation leaders facilitated lessons to all of the groups of youth; however, each group of youth only engaged with one counselor and one
education leader which could result in an implementer effect (Jackson, 2008). An implementation threat suggests that different individuals will implement the intervention differently, which will influence the youth outcomes of the program. As the program was implemented by different staff members, for which different groups of youth were exposed to different staff members, the program administrators tried to control for an implementer effect by thoroughly training all staff on the program intervention and monitoring program implementation throughout the summer program.

A final threat to the internal validity of the proposed study was in relation to problems with the data which could produce invalid results, affecting the statistical conclusion validity. The problems could be due to unreliable measures, low statistical power, or a violation of statistical assumptions. To reduce this threat, it was important to establish the reliability of the instruments that were used and check all assumptions when conducting the analysis. By addressing the threats to internal validity, potential extraneous variables were controlled, systematic error was reduced, and the soundness of the results were strengthened.

In comparison to internal validity, external validity reflects the degree to which the results of a study can be generalized outside of the study, to other populations and settings (Ary et al., 2006). Similar to the testing effect, the interaction of the testing and treatment and reactive effects of the testing posed a threat to the external validity of the study. The pre-test and subsequent measures or testing conditions may have influenced the youth’s reaction to the intervention or the staff’s implementation of the intervention. These threats to external validity suggest that the testing influences the outcomes of the
intervention; therefore the results of the intervention cannot be generalized to individuals who are not tested. As discussed previously, the testing effect was reduced through appropriate timing, communication, and confidentiality.

A study cannot be valid if it is not reliable (Li, Pitts, & Quarterman, 2008); therefore it was important to establish the reliability of the instruments, prior to data collection, to enhance the validity of the results. Many scales established in the literature have evidence of reliability through test-retest reliability, item-total correlations, split-half correlation, or Cronbach’s alpha. To ensure the reliability of the scales used in the current study, results of previous reliability procedures were evaluated and a pilot test was conducted, upon which the data were used for further reliability and validity testing.

Pilot Test

The staff and youth scales used for the current study were piloted during the 2013 summer program facilitated by the organization under investigation. The program staff consisted of 22 recreation leaders, 6 education leaders, and 24 counselors, all of whom were required to participate in the pilot study as part of the overall program evaluation agenda. The staff instrument was designed to be completed at the end of every session, for which the staff facilitated 4 sessions per day for 15 days, resulting in a total of 60 measures. The staff received four blank hard-copies of the instrument, printed on 5x8 index cards, at the beginning of every day and were required to turn in the four completed copies at the end of every day. To test the validity and reliability of the tool, each individual session evaluation for all staff members were analyzed as unique cases in order to run exploratory factor analysis and reliability testing.
The youth instrument was administered to 476 youth \((N = 542)\) for the belonging scale; \(N = 534\) for the leader support scale) participating in the sport-based PYD program under investigation during the 2013 summer program. Parental/guardian consent was obtained during registration for the summer program and assent was obtained from youth 13-15 years of age on the first day of the program. Registration for the program was not contingent upon participation in the pilot study. The hard-copy instrument was administered to the youth at two separate times (i.e., pre-test and post-test) to allow for longitudinal confirmatory factor analysis (LCFA). The social competence, effort, teamwork, self-control, and athletic competence measures were collected at pre-test and post-test while the belonging and leader support measures were only collected at post-test. The behavioral engagement measure was a new construct to be introduced to the program and therefore was not included in the pilot test. The results of the validity and reliability analyses from the pilot test are presented within the following instruments section.

**Instruments**

**Staff Survey**

The staff instrument, titled “Session Log” (Appendix A), was designed specifically for the program under investigation to provide a more objective, in-depth assessment of the program implementation fidelity of the staff over the course of the summer program. Faculty and staff within the Community and Youth Collaborative Institute (CAYCI) housed at The Ohio State University reviewed the research and evaluation needs of the program under investigation to determine what questions and
constructs were most critical for the program, taking into consideration the PYD principles and strategies the program is founded on. To further establish the program design and curriculum, an examination of the implementation of the program and curriculum was needed. The CACYI faculty and staff consulted the literature related to sport-based PYD curriculum, curriculum implementation, intervention fidelity and treatment integrity, and implementation stages and examined available curriculum implementation fidelity assessments to develop an implementation fidelity instrument. The tool was reviewed by six individuals, of which four have expertise in research methodology and play a critical role in the program and two have a minimum of one year experience fulfilling the summer staff role for the program.

The original instrument consisted of 39 items broken down into 6 primary sections, which consisted of adherence to program structure (14 items), ease of implementation (7 items), use of curriculum (6 items), establishment of a caring, mastery goal-oriented climate (7 items), leadership of staff (3 items), and presence of supervision (2 items). Based upon the staff role and program curriculum, only the recreation leaders were required to complete the entire battery of items. The education leaders completed 32 of the 39 items and the counselors completed 25 of the 39 items. The staff were asked to evaluate the implementation of each item for each session (four sessions per day of curriculum), using a 5-point scale (0 = No implementation, 5 = Total implementation).

The original tool was pilot tested in 2013 to establish the validity and reliability of the instrument. As this instrument was created, rather than established in the literature, principal component analysis (PCA) was employed to examine the factor structure of the
instrument, for which individual session evaluations were used as unique cases. Since only the recreation leaders completed the entire battery of items, the PCA was limited to these cases \((N = 1440)\). Oblimin rotation was used based upon the component correlations which ranged from .211 to .405, suggesting that the relationship among the components should be taken into consideration (Cohen, 1988).

The results identified 3 predominant components which account for 54.4% of the variance. The components consisted of program environment (13 items), program structure (9 items), and program curriculum (5 items). The three components demonstrated strong reliability with Cronbach’s alphas ranging from .875 to .928 (Hair, Anderson, Tathan, & Black, 1998). The items that either loaded onto an ambiguous component or had a factor loading less than .400 were removed (Hair et al., 1998).

Based upon the results, the ease of implementation, leadership of staff, and presence of supervision sections were removed, along with four items within the adherence to program structure section, one item within use of curriculum, and one item within establishment of a caring, mastery goal-oriented climate. Additionally, one item within the use of the curriculum section was moved to the adherence to the program structure based upon empirical and theoretical justification. The remaining items, which loaded onto one of the three components, were slightly modified and six new items added to accommodate changes to the program model and curriculum. The scale was also revised to a 10-point scale \((1 = \text{No implementation}, 10 = \text{Total implementation})\) to allow for more variance and sensitivity.

The revised staff instrument administered in the current study consisted of 30
items broken into 4 sections (Appendix A), which included: adherence to program structure (13 items), use of curriculum (4 items), establishment of a caring, mastery goal-oriented climate (7 items), and descriptive questions (6 items). As stated previously, the instrument used a revised 10-point Likert type scale, assessing perceived degree of implementation, which ranged from 1 (none) to 10 (total). The first section (i.e., adherence to program structure) consisted of a lesson plan divided into three progressive phases which outlined how each lesson should be facilitated.

The lesson plan was designed to begin with framing of the sport and social skills, followed by facilitation of an activity which allows the youth opportunities to practice the sport and social skills, and ending with a debriefing dialogue to provide time to process the skills and enhance transfer of learning. The framing phase consisted of five items: (1) Staff overviewed today’s S.E.T.S., (2) Staff overviewed today’s sport skill, (3) Staff gave instructions of sport cues, (4) Staff taught how to use S.E.T.S. in activity, (5) Staff gave instructions for activity; the facilitating phase included four items: (6) Staff provided opportunity for youth to practice sport skill, (7) Staff praised campers when they demonstrated good sport skill, (8) Staff praised campers when they demonstrated good S.E.T.S., (9) Staff pointed out use of S.E.T.S. in activity. The last phase was debriefing, which also consisted of four items: (10) Staff reviewed today’s S.E.T.S., (11) Staff reviewed today’s sport skill, (12) Staff discussed using S.E.T.S. outside of camp, (13) Staff engaged youth in debriefing process.

The second section of the instrument examined overall implementation of the program curriculum which focused on the social skill content to be taught. The
The curriculum section consisted of the following four items: (14) **Staff emphasized self-control**, (15) **Staff emphasized effort**, (16) **Staff emphasized teamwork**, (17) **Staff emphasized social responsibility**. The final section focused on the climate created by the staff members; the program was designed to facilitate lessons in a caring, mastery goal-oriented climate based on evidence-based PYD principles. The climate section included seven items: (18) **Staff created an emotionally safe climate**, (19) **Staff greeted youth as they arrived**, (20) **Staff used warm tone of voice**, (21) **Staff encouraged youth to try new skills**, (22) **Staff encouraged for participation by all youth**, (23) **Staff provided opportunity for youth to make at least one choice**, (24) **Staff kept youth on task**.

In addition to the implementation fidelity items, the staff instrument included six descriptive questions for the purpose of data reduction: staff identification number, day number (1-15), session number (1-4), session type (recreation, chalk talk), group numbers (1-24), and group organization (combined, separated, rotated). The staff identification number was used to connect the longitudinal data while protecting the confidentiality of the staff. Day and session information was acquired in order to coordinate the staff implementation data with the youth who were present and exposed to the program implementation, for the purpose of creating program implementation variables. The nine sport-related sessions were categorized as recreation, while the education session was labeled “Chalk Talk” for the purpose of deterring youth from associating this type of session with a traditional school class.

In regards to group number and organization, the two groups of youth that traveled together could either be combined, kept separate, or rotated during recreation.
sessions. A combined session would consist of two recreation leaders (with the exception of swimming) and two counselors implementing the session for the two groups of youth together. For a session in which the groups were organized separately, one recreation/education leader and one counselor would implement the session with one group of youth, while the other recreation/education leader and counselor would implement the session with the other group of youth separately. The groups were always organized separately for education sessions. When groups were rotated, the counselors would stay with their corresponding group, for which each group would spend half of the session with each recreation leader separately. When groups were kept separate, the staff were instructed to write down the number of the group they were instructing. For sessions in which groups were combined or rotated, the staff were instructed to record the numbers of both groups. Group organization informed data reduction procedures, which will be outlined in the statistical analysis section.

The revised instrument is comprehensive and reflects the program structure, curriculum, and environment which was to be implemented during every session. As such, all staff members (i.e., counselors, recreation leaders, education leaders) completed the entire battery of items; unlike the original instrument which consisted of three versions corresponding with the unique staff roles. Education leaders were instructed to record a “1” for all sport-related items, since the education curriculum does not include the intentional instruction of sport skills. These responses were later changed to systematic missing data in order for the implementation tool to accurately reflect the program’s curriculum. This process is further outlined within the statistical analysis.
In addition to the implementation fidelity questions within the session log, the staff were asked demographic questions within the staff application. The demographic questions used for the current study include: college graduation date, gender, age, race, previous high school athletic experience, and current collegiate athletic experience. The classification of intern versus employee was determined by the application submitted as there were two distinct versions of the staff application (i.e., internship, employment).

**Youth Survey**

The youth instrument titled “LiFE Sports Social Competence Education Survey” (Pre-test: Appendix B; Post-test: Appendix C), which was administered for the current study, examined eight primary constructs. The youth were asked questions about their perceived actions, behaviors, and beliefs in relation to social competence, effort in sport, teamwork in sport, self-control in sport, and general sport ability in order to assess the program’s realization of its primary and secondary objectives. In addition to questions measuring youth outcomes, the youth responded to questions regarding their perceptions of belonging, leader support, and behavioral engagement in order to gain a better understanding of the mechanisms that may influence program outcomes. The instrument consisted of two separate tools, for which six constructs were measured at the pre- and post-test and three constructs were measured at post-test only. The pre-test consisted of 31 items (see Appendix B), while the post-test included 50 items (see Appendix C).

**Social competence.** A modified version of the Perceived Social Competence Scale (PSCS) was used to assess social competence of the youth (Anderson-Butcher,
Iachini, & Amorose, 2008). The original scale consists of four items, which include the following: (1) *I help other people*, (2) *I ask others if I can be of help*, (3) *I am good at making friends*, (4) *I get along well with others*. A 5-point Likert type scale is utilized by the PSCS, ranging from 1 (*not at all*) to 5 (*very much*), asking how true each statement is of them. The PSCS was found to demonstrate factorial validity within the PYD context \[\chi^2(2) = .30, \ p = .86; \ \text{RMSEA} < .001; \ \text{NNFI} = 1.00; \ \text{CFI} = 1.00; \ \text{GFI} = 1.00\], acceptable internal consistency (\(\alpha = .81\)), gender invariance \[\chi^2(12) = 23.65, \ p = .02; \ \text{RMSEA} = .07; \ \text{NNFI} = .98; \ \text{CFI} = .98; \ \text{GFI} = .98\], and predictive validity (i.e., significant positive correlation between social competence and perceived belongingness; \(r = .41\); Anderson-Butcher et al., 2008).

The PSCS was originally a 6-item scale reduced to four items based on longitudinal confirmatory factor analysis (CFA). The 4-item scale has been associated with limitations, such as a ceiling effect (Anderson-Butcher, Iachini, Boester, Wade-Mdivanian, Davis, & Amorose, 2009) and limited variability in the scores (Anderson-Butcher, Amorose, Iachini, & Ball, 2013), which led to the development of four additional items, including: (5) *I show concern for others*, (6) *I am a good friend*, (7) *I show care for others*, (8) *I give support to others*. The scaling was also modified to consist of more explicit response options, ranging from 1 (*not at all true*) to 5 (*really true*).

The 8-item modified PSCS has since been tested to determine the validity and reliability of the modified scale (Anderson-Butcher et al., in press). Anderson-Butcher and colleagues (2014) conducted confirmatory factor analyses (CFA) to examine the
factor structure and invariance of the scale across gender and time. When testing the basic factor structure of the scale, a 5-item version of the scale (labeled PSCS-II) provided the best fit for the male \( \chi^2(5) = 11.30, p = .05; \text{RMSEA} = .04 (90\% \text{CI} = .00-.11); \text{CFI} = 1.00; \text{TLI} = .99 \) and female \( \chi^2(5) = 2.28, p = .81; \text{RMSEA} = .00 (90\% \text{CI} = .00-.04); \text{CFI} = 1.00; \text{TLI} = 1.02 \) populations and was found to function similarly for both groups.

The PSCS-II was then tested for longitudinal invariance across three time points, with increasingly restrictive constraints imposed on the model, for which there was evidence to support strong invariance of the measure over time (change in CFI < .01).

The PSCS-II consists of the following five items: (1) *I help other people*, (2) *I ask others if I can be of help*, (5) *I show concern for others*, (7) *I show care for others*, and (8) *I give support to others*. The five item scale was found to demonstrate moderately strong predictive validity with the Social Skills Improvement System measure of perceived social skills \((r = .597)\).

**Effort.** Effort in sport was measured using a modified version of the Multidimensional Sportspersonship Orientations Scale’s (MSOS) commitment subscale (Vallerand, Briere, Blanchard, & Provencher, 1997). The MSOS was developed to assess individual differences in sportspersonship orientations, with an emphasis on commitment, social conventions, rules and officials, opponent, and negative approach. The MSOS commitment subscale consists of five items, which include the following: (1) *Important to be at all practices*, (2) *Give maximum effort*, (3) *Think how to improve*, (4) *Do not give up after mistakes*, and (5) *More effort even if certain of losing*. The instrument provides
the subjects with scenarios, for which the subjects are asked to complete related items, answering whether they intend to behave in accordance with or against the sportsmanship orientation presented in each item, for which a 5-point Likert type scale is used, ranging from 1 (no) to 5 (yes).

The MSOS was found to demonstrate factorial validity \([\chi^2(265) = 532.53, p < .001; \text{CFI} = .90; \text{BBNNFI} = .89]\), acceptable internal consistency for each subscale (commitment subscale \(\alpha = .71\)), and temporal stability (commitment subscale test re-test \(r = .76\); Vallerand et al., 1997). The five items were modified to be complete statements not contingent upon the provision of sportspersonship related scenarios. The modified items include the following: (1) *In competition, I go all out even if I’m sure to lose*, (2) *I don’t give up even after making many mistakes*, (3) *I think about ways to improve my weaknesses*, (4) *It is important to me to be present at all practices*, and (5) *During practices, I go all out*, using a 5-point Likert type scale ranging from 1 (doesn’t correspond to me at all) to 5 (corresponds to me exactly).

The pilot test data of the modified five items was used to conduct CFAs and reliability testing of the modified scale. To examine the factor structure, a CFA was conducted resulting in a good fitting model \([\chi^2(29) = 94.275, p < .001; \text{CFI} = .979; \text{GFI} = .963; \text{RMSEA} = .069; \text{SRMR} = .042]\). The five item scale was tested for factorial invariance across time and demonstrated evidence of strong invariance over time \([\chi^2(37) = 116.913, p < .001; \text{CFI} = .974; \text{GFI} = .961; \text{RMSEA} = .067; \text{SRMR} = .043; \Delta \text{CFI} = .05]\) with four of the five global fit indices suggesting a good fitting model. In addition to factorial validity and invariance, the scale demonstrated moderate reliability with an internal
consistency of $\alpha = .744$ at pre-test and $\alpha = .855$ at post-test. Therefore, the 5-item modified scale, with the 5-point Likert type scale, was used for the current study.

**Teamwork.** Teamwork in sport was measured by a modified version of the Teamwork Scale (CACYI, 2010). CAYCI, housed at The Ohio State University, developed the Teamwork Scale for the purpose of measuring teamwork within the context of sport-based PYD programs (CAYCI, 2010). CACYI faculty experienced with research methodology consulted the literature, examined alternative teamwork measures, and gathered feedback from social workers to develop the Teamwork Scale for Youth (Lower, Anderson-Butcher, & Newman, in review). The measure consists of 10 items, which include the following: (1) *I think that teamwork is important*, (2) *People who work in teams can learn more than if they work by themselves*, (3) *I feel confident in my ability to work in a team*, (4) *I know how to give my group members feedback that will not hurt their feelings*, (5) *I ask others for feedback*, (6) *I make an effort to include other members of my group*, (7) *I value the contributions of my group members*, (8) *I treat my group members as equal members of the team*, (9) *I am good at communicating with my group members*, and (10) *I feel confident in my ability to be a leader*.

The Teamwork Scale for Youth has not been tested for validity and reliability; therefore the pilot test data were used to conduct CFAs and reliability testing to provide evidence of the quality of the measure. Initial testing for factorial validity suggested a good fitting model [$\chi^2(372) = 836.960, p < .001; \text{CFI}: .976; \text{GFI}: .894; \text{RMSEA}: .052; \text{SRMR}: .046$], however the MIs, standardized residuals, and squared multiple correlations ($R^2$) suggested that items one and two were potentially problematic. Conceptually, items
one and two focus on teamwork attitudes while the remaining six items examine teamwork behaviors. Items one and two were eliminated which resulted in a better fitting model $[\chi^2(225) = 455.609, p < .001; \text{CFI:.984; GFI:.925; RMSEA:.047; SRMR:.042}]$. The eight item modified Teamwork Scale for Youth was tested for factorial invariance across time, for which the scale demonstrated weak invariance [non-significant change in $\chi^2$, change in CFI < .01], which is the minimum requirement to support longitudinal invariance (Little, 2013). The modified scale also demonstrated an acceptable internal consistency (pre-test $\alpha = .768$; post-test $\alpha = .878$), supporting the reliability of the scale. The final teamwork scale consists of eight items: (3) I feel confident in my ability to work in a team, (4) I know how to give my group members feedback that will not hurt their feelings, (5) I ask others for feedback, (6) I make an effort to include other members of my group, (7) I value the contributions of my group members, (8) I treat my group members as equal members of the team, (9) I am good at communicating with my group members, and (10) I feel confident in my ability to be a leader, using a 5-point Likert type scale ranging from 1 (not at all true) to 5 (really true).

**Self-control.** Self-control in sport was assessed through the Social Sports Experiences Scale developed by CACYI (Anderson-Butcher, Wade-Mdivanian, Riley, & Davis, 2010). The Social Sports Experiences Scale is designed to examine actions and behaviors in sport which reflect social responsibility and self-control. The scale consists of the following eight items: (1) I respect others when playing sports, (2) I work well with my group members when playing sports, (3) I control my temper when playing sports, (4) I play sports fairly even when an adult is not around, (5) I am good at working together
with my group members when playing sports, (6) I take responsibility for my actions when playing sports, (7) I act responsibly when playing sports, and (8) I cooperate with others when playing sports. Validity and reliability of the Social Sports Experiences Scale has been supported by factorial validity (S-B $\chi^2 = 45.47$, d = 20, $p < .01$; RMSEA = .05; SRMR = .04; CFI = .99; TLI = .98), an acceptable internal consistency at pre and post-test ($\alpha = .85-.91$), and a statistically significant intraclass correlation between the pre and post-test ($r = .62$; McDonough et al., 2013). As the scale was not modified for the current study, the validity and reliability evidence established in the literature serves as demonstration of the quality of the measure (Anderson-Butcher et al., 2010).

**Athletic competence.** The Perceived Athletic Competence scale by Amorose (2002) was used to assess the youth’s perception of their general sport ability. The scale consists of three items each associated with a unique Likert type scale, including: (1) *How good do you think you are at sports?* (1 = not good at all … 5 = very good), (2) *How skilled do you think you are at sports?* (1 = not skilled at all … 5 = very skilled), and (3) *How much ability do you think you have when it comes to sports?* (1 = not much ability at all … 5 = a whole lot of ability). Amorose (2002) confirmed the scale had an acceptable internal consistency ($\alpha = .76$) and suggested strong face validity of the items.

To provide further evidence of the validity and reliability of the scale, the pilot test data were used to conduct CFAs and reliability testing. CFA upheld the three item factor structure of the scale [$\chi^2(5) = 3.284$, $p = .656$; CFI: 1.000; GFI: .998; RMSEA < .001; SRMR: .010], with all five global fit indices indicating a good fitting model. When imposing constraints on the model to test the factorial invariance across time, the model
demonstrated strong invariance by both a non-significant change in $\chi^2$ and a change in CFI < .001 [$\chi^2(9) = 10.239, p = .332; \text{CFI} = .999; \text{GFI} = .994; \text{RMSEA} = .017; \text{SRMR} = .027; \Delta \chi^2(2) = 1.24, p = .538; \Delta \text{CFI} = .001$], supporting the factorial invariance of the scale. Reliability testing supported the internal consistency of the scale (pre-test $\alpha = .847$, post-test $\alpha = .883$), confirming the reliability suggested by Amorose (2002). As the Perceived Athletic Competence scale was found to be valid and reliable, the full three item scale was used for the current study.

**Belonging.** To measure the youth’s sense of belonging to the program, a modified version of the sense of belonging scale was used (Anderson-Butcher & Conroy, 2002). The original 10-item scale was designed to measure youth and adolescents’ perceived commitment, engagement, and connectedness to a program. The scale was found to be poor fitting [$\chi^2(35) = 177.83; \text{RMSEA} = .13; \text{CFI} = .92; \text{SRMR} = .06; \text{GFI} = .88; \text{NNFI} = .90$], for which five items were removed to improve the fit of the measurement model (Anderson-Butcher & Conroy, 2002). The resulting modified scale consists of the following five items: (1) *I feel comfortable at the program*, (2) *I am a part of the program*, (3) *I am committed to the program*, (4) *I am supported at the program*, and (5) *I am accepted at the program*, using a 4-point Likert type scale ranging from 1 (*NO!* to 4 (*YES!*). The modified 5-item scale was found to demonstrate strong internal consistency ($\alpha = .93$) and construct validity [$\chi^2(5) = 11.74; \text{RMSEA} = .07; \text{CFI} = .99; \text{SRMR} = .01; \text{GFI} = .98; \text{NNFI} = .99$], tight cross validity [$\chi^2(20) = 42.29; \text{RMSEA} = .07; \text{CFI} = .99; \text{SRMR} = .11; \text{GFI} = .93; \text{NNFI} = .99; \Delta \chi^2(5) = 11.72, p > .05$], gender invariance at the levels of item pattern/structure coefficients and factor invariance [$\chi^2(15)]
and predictive validity related to self-reported 7-day attendance [Spearman’s 
\( \rho(272) = .59, p < .01 \)], 6-month attendance [\( r(277) = .37, p < .01 \)], community rewards for prosocial involvement [\( r(275) = .18, p < .01 \)], neighborhood attachment [\( r(276) = .24, p < .01 \)], community disorganization [\( r(275) = -.31, p < .01 \)], and laws and norms favorable to drug use [\( r(275) = -.32, p < .01 \); Anderson-Butcher & Conroy, 2002].

CAYCI modified the 5-item scale used by Anderson-Butcher and Conroy (2002) by adjusting the language to reflect the title of the program under investigation, incorporating the identification of comfortability with people in item one, as well as adding an item investigating the youth’s sense of safety at the program. The modifications resulted in the following six items: (1) I feel comfortable with people at LiFE Sports, (2) I am part of LiFE Sports, (3) I am committed to LiFE Sports, (4) I am supported at LiFE Sports, (5) I am accepted at LiFE Sports, (6) I feel safe at LiFE Sports. The modified tool uses a 5-point Likert type scale ranging from 1 (Not at all true) to 5 (Really true) for consistency throughout the instrument. To examine the validity and reliability of CAYCI’s modified scale, the pilot test data were used to run CFA and Cronbach’s alpha. The initial run of the measurement model resulted in a good fitting model [\( \chi^2(9) = 36.292, p < .001; \) RMSEA = .075; CFI = .989; GFI = .978; SRMR = .027], supporting the inclusion of all six items. Cronbach’s alpha of the 6-item scale demonstrated strong reliability (\( \alpha = .878 \)).

**Leader support.** The youths’ perception of their group leader’s support was measured by a modified version of the Modified Psychological Sense of School Membership
(PSSM) teacher support subscale (Cox & Williams, 2008). The PSSM scale was developed by Goodenow (1993) as a unidimensional measure to assess belonging among middle school students. Cox and Williams (2008) conducted principal axis factor analysis and SEM on the full scale to investigate any underlying dimensions. The PSSM was found to have three primary factors, including teacher support (six items), alienation (three items), and relatedness (five items).

The Modified PSSM teacher support subscale consisted of the following items:
(1) My P.E. teacher is not interested in people like me, (2) My P.E. teacher respects me, (3) I can talk to my P.E. teacher if I have a problem, (4) My P.E. teacher is interested in me, (5) I wish I were in a different P.E. class, (6) I am treated with as much respect as the other students in my P.E. class, using a 5-point scale ranging from 1 (Not at all true of me) to 5 (Very true of me). All teacher support factor loadings from the principal axis factor analysis were greater than .40. Reliability testing supports the internal consistency of the teacher support subscale, with a Cronbach’s alpha of .82.

The Modified PSSM teacher support subscale was further modified to be appropriate for the current program under investigation, rather than the physical education context. CAYCI slightly adjusted the language of the items to reflect the different context, these modified items include: (1) My LiFE Sports group leader respects me, (2) I wish I were in a different LiFE Sports group, (3) My LiFE Sports group leader is not interested in people like me, (4) My LiFE Sports group leader thinks I am interesting, (5) I can talk to my LiFE Sports group leader if I have a problem, (6) I am treated with as much respect as the other kids in my LiFE Sports group. The 5-point scale
was also slightly adjusted to 1 (Not at all true) to 5 (Really true). The modified subscale was pilot tested to establish validity and reliability of the measure.

CFA of the pilot test data were conducted to examine the basic factor structure of the modified subscale. The 6-item measurement model was found to have a poor fit [$\chi^2(9) = 141.980, p < .001; \text{RMSEA} = .166; \text{CFI} = .871; \text{GFI} = .925; \text{SRMR} = .093$]. The multiple correlation coefficients, standardized residuals, and suggested error covariances were examined to determine model modification, for which the negatively worded items (item 2, item 3) were problematic. By removing the negatively worded items, the model fit significantly improved [$\chi^2(2) = 2.234, p = .3272; \text{RMSEA} = .015; \text{CFI} = 1.00; \text{GFI} = .998; \text{SRMR} = .0107$]. Reliability testing of the modified 6-item subscale resulted in a Cronbach’s alpha of .721, for which the 4-item scale (with the negatively worded items removed) only increased the alpha by .07 ($\alpha = .791$).

To determine the most valid and reliable scale for the program under investigation, items 2 and 3 were retained and two additional items created to reflect the positive conceptualization of the items. The new items include: I do not wish I were in a different LiFE Sports group (item 2 reversed) and My Group Coach is interested in people like me (item 3 reversed). The eight items were reorganized to balance the positively and negatively worded items, for which the following order was used: (1) I do not wish I were in a different LiFE Sports group, (2) My Group Coach is interested in people like me, (3) My Group Coach respects me, (4) My Group Coach thinks I am interesting, (5) I can talk to my Group Coach if I have a problem, (6) I wish I were in a different LiFE Sports group, (7) My Group Coach is not interested in people like me, (8) I
am treated with as much respect as the other kids in my LiFE Sports group. The language was slightly adjusted to reflect revisions to the program’s curriculum (i.e., LiFE Sports group leaders are now called “Group Coach”) and the modified 5-point scaling previously noted was upheld.

**Engagement.** To assess quality of engagement, the current study focused on behavioral engagement, as opposed to emotional or cognitive engagement (Fredricks et al., 2004). A modified version of the Engagement versus Disaffection with Learning behavioral engagement subscale was used to measure quality of engagement in the sport-based PYD program under investigation (Skinner, Kindermann, & Furrer, 2009). The 5-item behavioral engagement subscale is part of a larger 27 item self-report measure of engagement and disaffection in school designed for the elementary school student population. The instrument is an updated version of the original engagement scale developed as part of the Rochester Assessment Package for Schools, which has been heavily utilized in the academic context [Fredricks et al., 2004; Institute for Research and Reform in Education (IRRE), 1998]. The behavioral engagement subscale consists of five items that include: (1) *I try hard to do well in school*, (2) *In class, I work as hard as I can*, (3) *When I’m in class, I participate in class discussions*, (4) *I pay attention in class*, and (5) *When I’m in class, I listen very carefully*, using a 4-point Likert type scale ranging from 1(*Not At All True*) to 4(*Very True*). The subscale has demonstrated acceptable construct validity ($\chi^2 = 6.59, p < .05$; CFI = 1.00; TLI = 1.00; RMSEA = .05) and moderate reliability over two time-periods ($\alpha = .61-.72$; Skinner et al., 2009).

The behavioral engagement subscale was modified to be appropriate for the
context of the current study, reflecting the PYD program rather than class setting. The tense of the scale was also modified to be past-tense as the tool was used within the post-test to measure the youths’ perceived engagement during their time in the program. The five modified items include: (1) *I tried hard to do well in LiFE Sports*, (2) *In LiFE Sports, I worked as hard as I could*, (3) *When I was in LiFE Sports, I participated in the discussions*, (4) *I paid attention in LiFE Sports*, and (5) *When I was in LiFE Sports, I listened very carefully*, and use a 5-point Likert type scale ranging from 1 (*Not at all true*) to 5 (*Really true*) to remain consistent with the other tools.

**Demographics.** To gain a more in-depth understanding of the youth outcomes and outcome growth, demographic information was collected within the youth registration application. The demographic questions used for the current study included: previous program participation, household income, number of household residents, eligibility for free or reduced lunch at school, gender, race, and age. The demographic items were collected to provide valuable contextual information.

**Attendance.** Youth attendance data were collected daily by program administration staff as a means of maintaining documentation of the youth that were present and which group they were traveling with. The attendance records indicated youth who were present, youth with an excused absence (i.e., parent/guardian provided notice), and youth with an unexcused absence (i.e., no notice provided). The data of youth present were used as secondary data for the current study to determine amount of program exposure.
Data Collection

Program Staff

Research participation. Staff were required to attend a two day training one week prior to camp. At the beginning of staff training, the staff were provided an overview of the program evaluation and research agenda of the program under investigation. All staff were required to participate in program evaluation as part of their role, as the session log tool was used to both assess program implementation fidelity and manage quality control. The purpose of quality control was based upon the belief that self-evaluation of one’s implementation fidelity, consisting of review and reflection upon the program’s intended implementation, should enhance one’s implementation fidelity. As the session log was used for program evaluation purposes, the staff were not asked to sign a research consent form. The researchers of the current study received approval from the Institutional Review Board of the governing institution to use the de-identified session log program evaluation data, as well as the demographic data outlined within the instruments section, as secondary data for the research study.

Training. To enhance the reliability of the staff instrument (a.k.a., “session log”), approximately 20 minutes of staff training was delegated for training the staff on the session log tool. Staff were provided a template of the instrument during staff training to examine the items and ask any questions regarding the interpretation of the items and scale or data collection procedures. The staff also received a two page handout of tips for how to fill out the session log appropriately, to enhance the reliability of the tool (Appendix D). At this time, the staff had already been exposed to a camp demonstration,
for which the research director demonstrated the appropriate facilitation of a sport session. Therefore, when training the staff on the session log tool, the research coordinator used relevant examples and scenarios related to the camp demonstration to help explain the items and the scaling (e.g., what constitutes no implementation versus total implementation). For example, total implementation of the item “Preview of tomorrow’s social skill” requires the explicit identification of the specific social skill to be emphasized the next day, at the end of the session during debriefing.

To reinforce the training on the session log tool and further enhance the reliability of the staff instrument, the staff participated in a second camp demonstration with the objective of practicing using the tool. The staff were separated by role (i.e., counselor, education leader, recreation leader), for which the staff received a camp demonstration tailored to their specific role. At the completion of the second camp demonstration, the staff spent approximately five to eight minutes completing the practice session log. Staff responses to each item were reviewed and discussed as a reliability check, to ensure accurate interpretation of the items and consistent application of the scaling. The staff asked any remaining questions regarding implementation fidelity evaluation as well as data collection procedures.

To conclude the training related to the session log tool, the staff were provided the context that rarely is it possible for a program to be implemented with perfect fidelity due to environmental, program, and participant factors. The research coordinator reiterated the objectives of the study, which focus on program evaluation and effectiveness for the purpose of program improvement, and conveyed that the subject’s responses would not
be used for performance evaluations nor impact employment status. Lastly, the research coordinator expressed the importance of honest responses, as the subjects’ responses will be used for future program decision-making.

**Data collection.** An identification number was assigned to each staff member for the purpose of connecting the longitudinal data while protecting the confidentiality of the subjects. The instrument identified the specific staff identification number, date, and session number to allow for longitudinal analysis. The session log tool was printed on 5x8 index cards, for which four hard-copy instruments (one for each session) were attached to a loose-leaf ring for distribution to each individual staff member daily.

At the beginning of each day of camp, with the exception of Career Day and Olympics, the loose-leaf rings with four blank session logs were distributed to the subjects based on staff identification number. The staff were instructed to complete one session log at the conclusion of each session, rather than complete all session logs at end of the day, to enhance the accuracy of their responses. The staff submitted their four completed session logs at the end of each day, upon which the completed logs were reviewed for missing data by the research coordinator and any missing data filled in by the staff prior to their departure. Completion of the 30 item staff instrument for the current study took less than five minutes, as there were only 15 minutes available in between each session.

**Youth Participants**

**Recruitment, consent, and assent.** Parents and guardians were recruited to consent to their child’s participation in the research study when completing registration
for the summer program. A parental consent form was embedded within the registration packet for the summer program (Appendix E). The parental consent form consisted of a description of the research study, which included information regarding the purpose of the study, voluntary nature of participation, procedures, task, and duration associated with participation, participant rights, confidentiality protection, and the risks and benefits of the study. Registration for the summer program was not contingent upon parental/guardian consent for their child to participate in the research study.

During registration, the research coordinator was present to answer any questions about the research study as well as collect and witness (through a signature) the signed parental consent forms. Only children with parental consent participated in the research study. Youth 13-15 years old, with parental consent, were recruited to participate in the research study on the first day of camp prior to data collection. During the youth’s education session, they were given the opportunity to assent to participation in the research study by reviewing and signing the youth assent form (Appendix F).

Similar to the parental consent form, the youth assent form provided information about the purpose of the study, procedures, tasks, and duration of participation, participant rights, and risks and benefits utilizing language that is appropriate and can be comprehended by adolescents. A research assistant, approved by the university’s institutional review board (IRB), was present in every education session to answer any questions about the research study as well as collect and witness all signed youth assent forms. For youth 13-15 years old, only those with parental consent and youth assent participated in the research study. The sport-based PYD program has an established
parental consent form and youth assent form that were appropriate for the current study. Therefore, the established forms were utilized, with no alterations made.

**Identification and group assignment.** Once parents and guardians registered their child for the summer program and a master registration list had been finalized, the youth were assigned an identification number which was used throughout the research study to protect their confidentiality. The youth were also randomly assigned into groups consisting of approximately 21-28 youth. A stratified random sampling technique was employed in which youth were stratified by age and then within each age group (e.g., 9, 10, 11 … 15) the youth were randomly assigned a group for the duration of camp. A total of 24 groups were created, for which the gender distribution was evaluated for each group and a few individuals were randomly re-assigned to another group to ensure a relatively even distribution of males and females in each group. Each group was assigned a counselor and along with one other group (e.g., groups 1 & 2, groups 3 & 4, etc.) followed a unique schedule throughout the duration of camp.

**Data collection.** The demographic data were collected within the registration application and attendance data were collected three times daily - in the morning, at lunch time, and in the afternoon by program administration staff. The governing institution’s IRB approved the use of both demographic and attendance data as secondary data for the current study. The youth surveys were administered to youth with parental consent (and youth assent if applicable) during the education session on the first day of camp (i.e., pre-test) and last day of curriculum (i.e., post-test).

Of the tools that were used, the social competence, effort, teamwork, self-control,
and athletic competence scales were administered at pre-test and post-test to allow for the change in youth outcomes to be analyzed. The belonging, leader support, and behavioral engagement scales were only administered at the post-test as the items are based on the youth’s experience at camp. Each youth survey had the youth’s identification number listed for the purpose of longitudinal analysis. A title page with the youth’s name and group number was attached to the survey for distribution purposes.

One research assistant was present in all education sessions to help facilitate the data collection process. All education leaders and research assistants were approved by the institution’s IRB to assist with the research study. The education leader and research assistant distributed the youth surveys, based on youth identification, and collected the surveys once completed or at the completion of the education session. At the end of the data collection day, the researcher reviewed all youth surveys for missing data and set aside incomplete surveys from complete surveys for follow-up procedures. The title page was removed from completed surveys for confidentiality purposes. For youth who were absent during data collection, unable to complete the full survey, or had missing data, the researcher provided opportunities during the youth’s education session and lunch to complete the survey for up to three days past the data collection period. These data collection procedures were repeated for the post-test.

The 31 item pre-test and 50 item post-test youth surveys for the current study were part of larger instruments. The comprehensive pre-test survey consisted of 52 items and the post-test survey 87 items. The survey items used for the current study should take no more than 30 to 40 minutes to complete; therefore, the total time (pre-test and post-
test) required for participation in the research study consisted of approximately 60-80 minutes.

Data Analysis

The data were imported into SPSS Statistics 22 software for initial review, treatment, and analysis. Data screening was employed to examine the extent of missing data. The youth data were limited to subjects who completed both the pre-test and the post-test for the purpose of longitudinal analysis; therefore, all youth who completed only one measure (pre-test or post-test) were removed from the analysis. In regards to the staff data, there were instances where a staff member was absent during camp, however that staff member was temporarily replaced by a substitute who completed the program implementation tool. In order to create accurate program implementation variables, no staff members (regardless of missing data) were removed from the analysis.

As the session log tool was designed to be comprehensive and completed by all staff roles, items unrelated to specific staff roles were treated as systematic missing data. For example, the curriculum designed for the education session does not incorporate the intentional instruction of sport skills, therefore all sport-related items are not relevant to education leaders. Education leaders were instructed to record a “1” for all sport-related items, for which that data were later changed to be systematically missing in order to accurately measure program implementation and not skew the data.

Once the data sets were established (e.g., removal of any youth who did not complete one of the measures, systematic missing data accounted for), missing data procedures, such as multiple imputation and mean replacement, were conducted to reduce
the amount of random missing data (Little, 2013; Schumacker & Lomax, 2010). For the staff data set, only random missing data due to partially completed measures were treated; therefore if a staff member was absent and did not complete a measure, that missing data were not treated.

Validity

Once missing data were treated, the validity and reliability of the individual youth scales and staff instrument were examined. CFA using SEM were conducted to examine the factor structure of the social competence, effort, teamwork, self-control, athletic competence, belonging, leader support, and behavioral engagement scales and factorial invariance of the social competence, effort, teamwork, self-control, and athletic competence scales over two time periods (Schumacker & Lomax, 2010). As the session log instrument was developed to assess implementation fidelity for the program under investigation and is not an established tool, PCA was conducted to explore the factor structure of the instrument (Stevens, 2009).

Confirmatory factor analysis. When conducting CFA to examine the factor structure of the selected scales, the following steps outline the procedures that were employed. The treated data were imported into LISREL 9.1 software to produce a covariance matrix of the observed variables for each scale. The covariance matrix, model relationships, and specified parameters were inputted into LISREL 9.1 for CFA. For models with data from two time periods (i.e., social competence, effort, teamwork, self-control, athletic competence), the individual scale items were specified to load onto the latent factor at the corresponding time point. As the latent factor for both time periods
represent the same construct, the factors were permitted to covary (Little, 2013).

Similarly, as the same items were used for both time periods, the corresponding item-specific errors were permitted to correlate (Little, 2013).

For models with data from one time period (i.e., belonging, leader support, behavioral engagement), the items were specified to load onto one latent factor, as the scales are designed to be unidimensional. The item-specific errors were not permitted to correlate as each item is unique. For each latent variable, the item with the largest factor loading was fixed to one to act as the reference variable and allow for factor loadings to be compared (Schumacker & Lomax, 2010).

When evaluating the model, the statistical significance of the $t$-values for the individual paths were examined first by comparing the observed $t$-values to the critical $t$-value for a two-tailed test at the $\alpha < .05$ level of significance (Schumacker & Lomax, 2010). Once the statistical significance of the parameter estimates were determined, the fit of the model was assessed through five global fit indices, which include: chi-square ($\chi^2$), normed chi-square (NC), root-mean-square error of approximation (RMSEA), comparative fit index (CFI), goodness-of-fit index (GFI), and standardized root-mean square residual (SRMR). The following standards of model fit were adopted, which are the following: a non-significant $\chi^2$, a NC less than 5.0, a RMSEA less than .100, a CFI and GFI greater than .95 (greater than .90 is considered acceptable), and a SRMR less than .05 (Kline, 2005; Schumacker & Lomax, 2010).

Once the global fit indices were examined, the standardized residuals, MI, EPC statistics, and square multiple correlations ($R^2$) were evaluated. Justifications for model
modifications were based on large standardized residuals (greater than 1.96 for a two-tailed test with an $\alpha < .05$), MIs and EFC statistics associated with a large anticipated decrease in $\chi^2$, and low parameter $R^2$ values (Schumacker & Lomax, 2010). The model was considered to be theoretically and empirically supported if the majority of global fit indices indicated a good fitting model and there were no theoretically justifiable modification suggestions (Schumacker & Lomax, 2010).

Once the structure of each youth scale was established, the social competence, effort, teamwork, self-control, and athletic competence scales were tested for factorial invariance across time using progressively restrictive longitudinal CFA models. The models imposed specific invariance constraints to determine whether the construct was maintained over time (Little, 2013). The first model tested was a configural invariance model for which no constraints were placed on any of the parameter estimates (Little, 2013). If the majority of global fit indices indicated an acceptable model fit, the more restrictive weak factorial invariance model was tested. For the weak factorial invariance model, the corresponding factor loadings were constrained to equal (Little, 2013).

The weak factorial invariance model was considered to be upheld if the change in CFI was less than .01 and the change in $\chi^2$ was non-significant, per Little’s (2013) recommendations. The $\chi^2$ test is suggested to be too sensitive (Little, 2013), therefore the CFI standard was considered the minimum requirement for model evaluation. If the weak factorial invariance model was upheld and the majority of fit indices indicated a good fitting model, the more restrictive strong factorial invariance model was tested. For the strong factorial invariance model, the corresponding intercepts were constrained to equal
(Little, 2013). The model evaluation utilized the standards previously noted for which the model was considered to demonstrate strong evidence of factorial invariance across time if the change in CFI was less than .01, change in $\chi^2$ was non-significant, and majority of fit indices indicated an acceptable model fit. Weak factorial invariance is the minimum requirement to suggest invariance over time, however strong factorial invariance is the desired outcome as it is a harder test to pass (Little, 2013).

**Principal component analysis.** To explore the factor structure of the session log instrument, PCA was conducted, for which each session log evaluation (60 per staff member) was treated as a unique case. Orthogonal rotation, due to uncorrelated factors ($r < .30$; Cohen, 1988), was used to assist with the interpretation of the components (Stevens, 2009). Components with an eigenvalue greater than one and practical significance were retained, based upon research which has found the Kaiser criterion to be accurate when the number of variables is less than 30 and sample size is greater than 250 (Stevens, 2009). Reliable components consisted of factors with four or more loadings that averaged to be greater than .60, for which loadings less than .40 were removed based upon statistical and practical significance (Stevens, 2009). Factors that loaded onto multiple components were reviewed and the loading with greater theoretical support selected (Stevens, 2009).

**Reliability**

Following validity testing, the reliability of each confirmed factor (i.e., social competence, effort, teamwork, self-control, athletic competence, belonging, leader support, behavioral engagement, program structure – S.E.T.S. coaching, program 135
structure – sports coaching, program environment, and program curriculum) was tested. Cronbach’s alpha was calculated for each confirmed scale, for which scaling if item deleted was considered. A scale was judged reliable based upon the recommended standard of a Cronbach’s alpha of .70 or greater (Hair et al., 1998) and item deletion resulting in a similar or smaller Cronbach’s alpha. Psychometric literature suggests a sample size of at least 300 when calculating Cronbach’s alpha, as the accuracy of the coefficient is improved with a larger sample size (Yurdugul, 2008). As previously outlined, each individual session log evaluation was treated as a unique case in order to have the sample size necessary for validity and reliability testing.

Data Reduction

Based upon the results of the validity and reliability testing, the youth data were reduced from multiple item scales to singular variables. The weighted mean of the item scores for each confirmed scale (i.e., social competence, effort, teamwork, self-control, athletic competence, belonging, leader support, behavioral engagement) was calculated for the corresponding time periods (i.e., pre-test and post-test, post-test only), based upon standardized factor loadings. The resulting mean scores were used for subsequent analysis.

Program implementation. To calculate the program implementation variables (i.e., adherence to program structure, use of curriculum, establishment of caring, mastery goal-oriented climate) and allow for an integrated analysis of staff and youth data, the staff schedule was corresponded with the youth group schedules to determine which implementation scores corresponded with which groups of youth. At each session during
camp, there were approximately two to seven staff present who evaluated the program implementation during the session; there were also two groups of youth engaging in the session with whom the implementation scores are associated. Therefore, each group of youth was associated with the implementation scores of the two to seven staff facilitates their respective sessions. The staff’s individual item responses associated with a specific group of youth were averaged across the staff members for the individual session in order to reduce the data and get a comprehensive assessment of the program implementation.

The staff scores associated with each group of youth were based upon whether the two groups of youth traveling together were separated, combined, or rotated during the session. If the groups were separated, each group was associated with the implementation scores of their counselor and the education leader or recreation leader(s) designated for their group. Groups that were combined during the session received the implementation scores of all staff present (counselors and recreation leaders), as the point of reference for the evaluation involved both groups. Sessions utilizing group rotation consisted of the separate groups spending half of the session with each recreation leader. Groups engaging in a rotated session were associated with the implementation scores of their counselor and all recreation leaders, based upon whose implementation they were exposed to.

The program implementation variables were calculated based upon the results of the validity and reliability testing. All items found to contribute to a program implementation component were averaged across all sessions the respective youth were present at, based upon attendance data. By averaging the data, rather than using
summation, the implementation data were normalized, reducing possible multicollinearity between the attendance variable and program implementation variables. Program attendance was integrated into the data reduction procedure in order to account for individual differences in program exposure within the group. The program implementation variables were used in subsequent analysis.

**Normality**

Once the data were reduced, descriptive statistics of the variable scores were calculated to assess normality and inform the structural model. Normality is an assumption of SEM as data that is skewed or peaked will affect the variance and covariance among the variables (Schumacker & Lomax, 2010). The assumption of univariate normality was considered to be met if all skewness scores fell within the standard range of +2.00 to -2.00 and all kurtosis scores fell within the standard range of +5.00 to -5.00, based on Kendall & Stuart’s (1958) recommendation. The assumption of multivariate normality was examined through conducting normality tests in LISREL 9.1. Multivariate normality was assumed to be met if Mardia’s measure of relative multivariate kurtosis was less than 3.0 (Siekpe, 2005). For data found to be non-normal, transformations were conducted (e.g., squared or cubed power transformation) to attempt to improve the normality (Bruce, Pope, & Stanistreet, 2008; von Hippel, 2004). To further account for non-normal data, an asymptotic covariance matrix was included along with the sample covariance matrix in the SEM input (Schumacker & Lomax, 2010).
Structural Equation Modeling

There are three primary approaches to testing SEM models, the one-step, two-step (Anderson & Gerbin, 1988), and four-step approach (Mulaik & Millsap, 2000). These approaches adopt different methods of identifying the source of specification error in a poor fitting structural model (Kline, 2005). The one-step approach combines the measurement and structural components of a SEM model into a single analysis. In comparison, the two-step approach first tests the measurement models separately from the structural model. Once the measurement models demonstrate an acceptable fit, the measurement and structural components will be combined into a structural regression model for SEM analysis. The original structural model will be compared to alternative structural models, using the $\chi^2$ difference test, to determine the best fitting structural model.

The final approach to testing SEM is an extension of the two-step approach and considered the most rigorous (Kline, 2005). The four-step approach involves testing a sequence of four hierarchical models, which requires at least four indicators for each factor. The four models include progressively restrictive constraints, for which if a model with fewer constraints fails to be upheld, then the more restrictive models should not be tested. The least restrictive model will allow all indicators to load onto all factors, with the purpose of determining the appropriate number of factors. The second step involves testing the measurement models to determine good fitting CFA models (similar to the first step in the two-step approach). The third step tests the structural model, combining the measurement and structural components, however “at least one unanalyzed
association between two factors is respecified as a direct effect or reciprocal effect and some of the factors are respecified as endogenous” (Kline, 2005, p. 218). The fourth step involves testing the a-priori hypotheses about parameters which typically includes the use of parameter constraints. The third and fourth step of the this approach is a more specific means of fulfilling step two in the two-step approach.

In comparison of the three primary approaches, two-step and four-step modeling is considered superior to the one-step approach due to the separation of measurement and structural issues by testing the CFA models prior to the SEM models (Kline, 2005). Although the four-step approach is an extension of two-step modeling, the complexity of the steps and requirement of at least four indicators per factor are disadvantages to consider. Based upon the limited number of indicators per factor in the current models under investigation (e.g., three indicators hypothesized to load onto the athletic competence factors), and the simplicity yet advancement of two-step modeling, this approach was selected for SEM analysis.

**Measurement models.** Following the two-step approach, the first step consists of establishing the measurement models prior to combining the measurement and structural components into a structural model. Testing the measurement models utilized similar steps as outlined for CFA. The asymptotic and sample covariance matrices of the observed variables were produced and inputted into LISREL 9.1 software. Additionally, the parameter relationships and constraints were specified and then SEM was run. To evaluate the measurement model, the observed t-values were first compared to the critical t-value of a two-tailed test at the α < .05 level of significance to determine the
significance of the parameter estimates. Once the significance of the observed $t$-values was determined, the goodness of fit indices (i.e., $\chi^2$, NC, RMSEA, CFI, GFI, and SRMR) were reviewed to examine model fit. The following standards of model fit were applied, including: a non-significant $\chi^2$, a NC less than 5.0, a RMSEA less than .100, a CFI and GFI greater than .95 (greater than .90 is considered acceptable), and a SRMR less than .05 (Kline, 2005; Schumacker & Lomax, 2010).

Upon determining the fit of the measurement model, the standardized residuals, MIs, EPC statistics, and $R^2$ were reviewed to establish whether any model modifications were theoretically justifiable and empirically supported. Model modifications were based upon results demonstrating large standardized residuals (outside the +/- 1.96 threshold), MIs associated with large EPC statistics, and/or low $R^2$ values. The measurement model was considered theoretically and empirically sound if the majority of global fit indices suggested a good fitting model and there were no theoretically justifiable modifications to apply to the model (Schumacker & Lomax, 2010).

**Structural equation models.** To examine the mediating relationship between program dosage and program outcomes, SEM was conducted. The analysis procedures closely follow the steps previously outlined. An asymptotic covariance matrix and the sample covariance matrix of the observed variables were produced and then inputted into LISREL 9.1 along with the model relationships and specified parameters. The model was first evaluated by the significance of the parameter estimates, based on the critical $t$-value for a two-tailed test at the $\alpha < .05$ level of significance. Once the significance of the parameter estimates was determined, the global fit indices previously identified were
examined to determine the fit of the model ($\chi^2$, NC, RMSEA, CFI, GFI, and SRMR). The following standards were utilized to determine goodness of fit, including: a non-significant $\chi^2$, a NC less than 5.0, a RMSEA less than .100, a CFI and GFI greater than .95 (greater than .90 is considered acceptable), and a SRMR less than .05 (Kline, 2005; Schumacker & Lomax, 2010). Upon examination of the global fit indices, the standardized residuals, MIs, EPC statistics, and $R^2$ values were reviewed to determine if any model modifications were needed. If the model produced large standardized residuals (greater than 1.96), MIs and EPC statistics suggested to substantially decrease the $\chi^2$, or low $R^2$ values that could be theoretically justified, the model was modified to improve the fit. The model was found to be empirically and theoretically sound if the majority of global fit indices indicated an acceptable model fit and there were no theoretically justifiable modifications to apply (Schumacker & Lomax, 2010).
Chapter 4: Results

Overview

To test the developed models of the current study, SEM was employed. Prior to SEM, the data and variables of interest were examined for validity, reliability, and normality to establish the credibility of the resulting analysis and findings. Results of the statistical analyses are presented within seven sections, including: demographics, data treatment, validity, reliability, normality, measurement models, and structural models.

Demographics

Staff Participants

To assess program dosage as a mediator of change in youth outcomes, the target populations studied included program staff and youth participants of the program under investigation. The staff population consisted of 22 recreation leaders, six education leaders, and 24 counselors. The majority of staff members were college students or recent graduates (86.5%), with 84.6% of staff between the ages of 18 to 24. The gender breakdown was roughly even, with 53.8% females and 46.2% males. With respect to race, 63.5% of staff identified themselves as White/Caucasian, 32.7% Black/African, and 3.8% Multiple Races. Of the 52 staff members, 48 reported previous high school athletic experience and 18 were current collegiate student-athletes. All staff were required to
complete a background check, concussion in sports certification, and protecting youth training prior to camp.

**Youth Participants**

The sport-based PYD program is designed to serve 600-800 economically disadvantaged youth in the local community. During registration, 621 youth enrolled in the summer program (302 returning participants), for which 539 parents/guardians consented to their child’s participation in the research study. Of the 539 youth with parental consent, 417 completed both measures (pre-test and post-test), resulting in a 77.4% response rate. As the data were limited to the 417 youth who completed both measures for statistical analysis purposes, the demographic information will reflect this subset of those registered for the program.

When considering the socio-economic status of the youth population, the majority of youth were associated with a household income at or below the national poverty guideline (59.0%; USDHHS, 2014) and eligible for a free or reduced school lunch (62.1%), as reported by their parent/guardian. The youth were 9 to 15 years of age, with 13.2% age 9, 17.0% age 10, 17.5% age 11, 18.7% age 12, 18.7% age 13, 11.3% age 14, and 3.1% age 15. The majority of youth were males (62.6%), with 36.9% females. Of the 417 youth who participated in the study, 81.1% self-identified as Black/African American, 11.0% Multiple Races, 4.3% White/Caucasian, and 0.7% “Some Other Race”. The program experienced high attendance, for which 88.0% of youth attended at least 16 of the 19 days of camp and 37.4% attended all 19 days.
Data Treatment

Prior to statistical analysis, the staff and youth data were reviewed to determine the amount of missing data. With a longitudinal research design, attrition is an anticipated challenge. Staff may be absent due to illness, family emergency, or professional development opportunities, but are incentivized to attend based upon the requirements of their employment or internship. Therefore, staff attrition and/or absence was minimal. Youth participants may also be absent due to illness or family emergency, but may have discontinued participation due to lack of incentive, as the program is free of charge, and limited motivation to attend the program for 19 days. A few missing scores will not significantly impact statistical analysis, however considerable amounts of missing data can create serious issues during statistical analysis (Kline, 2005; Schumacker & Lomax, 2010), therefore missing data must be addressed.

To address missing data, mean-replacement and multiple imputation procedures were conducted. For nominal amount of missing data (less than 1%), a mean-replacement technique was employed to treat the missing data. Mean replacement is a simple procedure for which the missing score is replaced by the sample average (Kline, 2005). This technique has been found effective when a small amount of values are missing (Schumacker & Lomax, 2010). With a larger amount of missing data (greater than 1%), mean replacement may reduce the variability of the scale by distorting the distribution of the data (Kline, 2005). A recommended technique to address larger amounts of missing data is regression imputation, which involves a more robust statistical procedure (Schumacker & Lomax, 2010). Regression imputation replaces the missing score with a
predicted score generated from conducting regression analysis on the nonmissing data (Kline, 2005). In the current study, multiple imputation procedures were employed to address data with large amount of missing values.

**Staff Survey**

The staff survey was developed for all staff members to complete at the conclusion of each session of curriculum, for which the program consisted of 15 days of curriculum with four sessions every day (60 measures total). Out of the 26 survey items assessing implementation of the program design, curriculum, and environment, 5 items were sport-specific (e.g., “Staff provided opportunity for youth to practice sport skill”). The education leaders were instructed not to complete the sport-specific items as their role was to teach and reinforce the social skills. The education curriculum did not include sport instruction, facilitation, or reinforcement; therefore, these items were not relevant to this subset of the staff population. With 12 education leaders intentionally not reporting 5 of the 26 survey items, for 60 measures, approximately 4.44% of staff data (3600 values) were systematically missing and not treated. As all youth were exposed to the same education curriculum (no sport instruction), the systematic missing data should not cause problems in the statistical analysis.

Upon review of the staff data, 0.1% of data were randomly missing (75 values). Due to the small amount of missing data, mean replacement was conducted, for which the missing score was replaced by the respondent’s mean variable score (Note: factors were determined through exploratory factor analysis, which will be later discussed). The resulting data set consisted of 100% complete cases.
Youth Survey

The youth survey was administered at two separate times, on the first and last day of curriculum (pre-test and post-test, respectively). As the data were limited to the 417 youth participants who completed both survey measures, to allow for statistical analysis which accounts for change in youth outcomes, missing data screening and treatment were limited to this sample. When examining the comprehensive youth data set (pre-test and post-test), 2.7% of data were found missing (1,044 values). Missing data scores within the variables under investigation were also examined, for which 3.3% of social competence scores were missing (136 values), 3.2% of effort scores (133 values), 3.3% of teamwork scores (217 values), 3.4% of self-control scores (224 values), 3.3% of athletic competence scores (82 values), 0.2% of belonging scores (6 values), 0.1% of behavioral engagement scores (1 value), and 0.3% of leader support scores were missing (11 values).

The Markov chain Monte Carlo multiple imputation technique was employed to treat the missing data for the youth survey, as there was a moderate amount of missing data (2.7%). The data were treated one construct at a time, rather than include all constructs within the same MCMC run. For example, missing data for the social competence pre-test and post-test scores were imputed separately from the other youth outcomes. The scales were treated separately based upon the supposition that not all constructs will be good predictors of each other, however responses to a scale developed to be internally consistent should highly correlate and therefore act as good predictors for the regression imputation. At the completion of the multiple imputation, the data were
once again screened for missingness, for which less than 0.1% of values were found missing (4 values). With such a small amount of remaining missing values, the missing data should not influence the statistical analysis.

**Validity and Reliability**

It is critical to establish the validity and reliability of the scale scores to make sure the scales are measuring what they are supposed to measure consistently (Kline, 2005; Schumacker & Lomax, 2010). An assumption of SEM is that all observed variables are valid and reliable (Schumacker & Lomax, 2010), therefore factor analysis and reliability testing were employed to examine the validity and reliability of the scale scores. For instruments developed for the purpose of the current study, PCA was conducted to examine the underlying factor structure of the new instrument (Stevens, 2009). CFA was conducted for all established scales to confirm the factor structure previously identified in the literature (Kline, 2005; Schumacker & Lomax, 2010). The factor structure information not only established the validity of the scales, but also informed data reduction procedures.

In addition to conducting CFA to examine the factor structure of the scales under investigation, longitudinal CFA was employed to assess measurement invariance across time for all longitudinal scales. It is critical to establish factorial stability across time when using longitudinal measures to ensure that any changes in scores are due to a change in the construct information rather than a change in how the subjects responded to the items (Little, 2013). Once the factor structure and stability of the scales were established, internal consistency of the measures was examined through reliability
testing, for which a Cronbach's alpha was calculated. Validity and reliability of all measures was established prior to SEM.

**Staff Survey**

The staff survey consisted of 24 items designed to measure 3 constructs: adherence to program structure, use of curriculum, and creation of a caring, mastery-oriented environment. All staff members completed the entire battery of items for each session of curriculum (60 sessions total). However, the education leaders were instructed to not complete the five sport-specific items (items 2, 3, 6, 7, 11) that measured adherence to program structure of sports coaching, as those items were not appropriate for their role. Although the 24 items were purposefully designed to measure 3 distinct constructs, the survey instrument had never been tested for validity and reliability. Therefore, PCA was conducted to examine the underlying factor structure of the tool.

Prior to conducting PCA, each completed survey was organized as an individual case, resulting in 3111 cases (please note that for 9 sessions, absent staff were not replaced with a substitute). As the education leaders were instructed to only complete 19 of the 24 items (5 items of systematically missing data), while the recreation leaders and counselors were required to complete the entire battery, a separate PCA was conducted on all cases ($N = 3111$), and recreation leaders and counselors only ($n = 2391$) to determine which analysis was easier to interpret. PCA using only complete cases (excluding data from education leaders) was found to be easier to interpret as well as conceptually supported.
The PCA results for both data sets were quite similar with the only major differences associated with item 13 and 21. The findings associated with the comprehensive data set indicated that item 13 ("Staff engaged youth in debriefing process") loaded with the adherence to program structure for the S.E.T.S. coaching component and item 21 ("Staff encouraged youth to try new skills") loaded with the ambiguous component. In comparison, the findings associated with the condensed data set (recreation leaders and counselors only), eliminating all systematic missing data, found that item 13 loaded with the ambiguous component and item 21 loaded with both the adherence to program structure for sports coaching component and creation of a caring, mastery-oriented climate component. The latter analysis is more conceptually sound as the debriefing process includes both sport and social skills (does not fit with adherence to program structure - S.E.T.S. or sport coaching separately) and encouragement to try new skills is primarily found during recreation sessions but is also a significant indicator of the environment created. Therefore, the PCA using only complete data (recreation leaders and counselors) was maintained.

PCA with orthogonal rotation was employed to ease interpretation of the components. An orthogonal rotation was used as the correlations among the components ranged from -.020 to .368, with the majority of correlations less than .200. The weak correlations among the components suggest that the relationships do not need to be accounted for in the analysis; therefore, an orthogonal rotation is more appropriate (Stevens, 2009). The initial PCA with orthogonal rotation did not constrain the analysis to a certain number of components, for which 5 components were extracted, accounting
for 60.6% of the variance. The five components had an eigenvalue greater than one, which is why they were retained. When examining the rotated component matrix, the five components were found to represent the following constructs: adherence to program structure - S.E.T.S. coaching; adherence to program structure - sports coaching; use of curriculum; creation of a caring, mastery-oriented environment; and ambiguous component (Table 2).

Table 2

*Factor loadings and Cronbach alphas for session log observed variables*

<table>
<thead>
<tr>
<th>Factors and Items</th>
<th>β</th>
<th>α</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adherence to Program Structure - Sports Coaching</td>
<td></td>
<td>0.880</td>
</tr>
<tr>
<td>Item 2: Staff overviewed today's sport skill</td>
<td>0.853</td>
<td></td>
</tr>
<tr>
<td>Item 3: Staff gave instructions of sport cues</td>
<td>0.774</td>
<td></td>
</tr>
<tr>
<td>Item 6: Staff provided opportunity for youth to practice sport skill</td>
<td>0.879</td>
<td></td>
</tr>
<tr>
<td>Item 7: Staff praised campers when they demonstrated good sport skill</td>
<td>0.833</td>
<td></td>
</tr>
<tr>
<td>Item 11: Staff reviewed today's sport skill</td>
<td>0.647</td>
<td></td>
</tr>
<tr>
<td>Adherence to Program Structure - S.E.T.S. Coaching</td>
<td></td>
<td>0.850</td>
</tr>
<tr>
<td>Item 1: Staff overviewed today's S.E.T.S.</td>
<td>0.822</td>
<td></td>
</tr>
<tr>
<td>Item 4: Staff taught how to use S.E.T.S. in activity</td>
<td>0.805</td>
<td></td>
</tr>
<tr>
<td>Item 8: Staff praised campers when they demonstrated good S.E.T.S.</td>
<td>0.711</td>
<td></td>
</tr>
</tbody>
</table>

Continued
### Table 2 continued

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>Staff pointed out use of S.E.T.S. in activity</td>
<td>0.772</td>
</tr>
<tr>
<td>10</td>
<td>Staff reviewed today's S.E.T.S.</td>
<td>0.583</td>
</tr>
<tr>
<td>12</td>
<td>Staff discussed using S.E.T.S. outside of camp</td>
<td>0.576</td>
</tr>
</tbody>
</table>

**Creation of a Caring, Mastery-Oriented Climate**

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Staff gave instructions for activity</td>
<td>0.560</td>
</tr>
<tr>
<td>18</td>
<td>Staff created an emotionally safe climate</td>
<td>0.494*</td>
</tr>
<tr>
<td>19</td>
<td>Staff greeted youth as they arrived</td>
<td>0.772</td>
</tr>
<tr>
<td>20</td>
<td>Staff used warm tone of voice</td>
<td>0.787</td>
</tr>
<tr>
<td>21</td>
<td>Staff encouraged youth to try new skills</td>
<td>0.435</td>
</tr>
<tr>
<td>22</td>
<td>Staff encouraged participation by all youth</td>
<td>0.712*</td>
</tr>
</tbody>
</table>

**Use of Curriculum**

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>14</td>
<td>Staff emphasized self-control</td>
<td>0.488</td>
</tr>
<tr>
<td>15</td>
<td>Staff emphasized effort</td>
<td>0.718</td>
</tr>
<tr>
<td>16</td>
<td>Staff emphasized teamwork</td>
<td>0.815</td>
</tr>
<tr>
<td>17</td>
<td>Staff emphasized social responsibility</td>
<td>0.690</td>
</tr>
</tbody>
</table>

**Ambiguous Factor**

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>Staff engaged youth in debriefing process</td>
<td>0.679**</td>
</tr>
<tr>
<td>23</td>
<td>Staff provided opportunity for youth to make at least one choice</td>
<td>-0.359**</td>
</tr>
</tbody>
</table>

*Note. *Item reduced scale reliability and was removed. ** Item loaded onto ambiguous factor with less than three loadings and was removed.*
Of the 24 items, 4 items loaded onto multiple components (loadings greater than .400; Stevens, 2009). Item 10 ("Staff reviewed today's S.E.T.S.") loaded onto adherence to program structure - S.E.T.S coaching (.583) and the ambiguous component (.469), item 11 ("Staff reviewed today's sport skill") loaded onto adherence to program structure - sports coaching (.647) and the ambiguous component (.482), item 12 ("Staff discussed using S.E.T.S. outside of camp") loaded onto adherence to program structure - S.E.T.S. coaching (.576) and the ambiguous component (.505), and item 21 ("Staff encouraged youth to try new skills") loaded onto adherence to program structure - sports coaching (.577) and creation of a caring, mastery-oriented environment (.435). In relation to items 10, 11, and 12, the items conceptually fit better with the component in which they had a larger loading score; therefore, these items were maintained with their largest component scores. When considering item 21, "new skills" was associated with both sport and social skills, therefore organizing this item within the adherence to program structure - sports coaching would be inappropriate as it would neglect the reinforcement of social skills. Item 21 was designed to assess the staff's creation of a mastery-oriented climate, therefore it conceptually fits better with this component. Refer to Table 2 for a complete breakdown of the final factor structure of the staff survey.

The components were considered empirically and theoretically sound if a minimum of three items loaded onto the components, with a component score greater than .400, and the items conceptually fit together (Stevens, 2009). The fifth component consisted of items 13 ("Staff engaged youth in debriefing process") and 23 ("Staff provided youth opportunity to make at least one choice"), which fall below the three item
threshold and are not conceptually related, therefore this component was removed from the analysis. Once the four components with empirical and theoretical support were established, reliability testing was conducted to examine the internal consistency of the items.

Reliability testing involved the calculation of Cronbach's alpha for each component, for which Cronbach's alpha were calculated based upon item deletion as well (see Table 2). The adherence to program structure components demonstrated strong internal consistency with an $\alpha = .880$ for sports coaching and $\alpha = .850$ for S.E.T.S. coaching, for which the Cronbach's alpha would not increase if any item was removed.

The creation of a caring, mastery-oriented climate component initially had an unacceptable Cronbach's alpha ($\alpha = .593$), for which the results suggested the removal of items 18 and 22 to increase the reliability of the component. Item 22 ("Staff encouraged participation by all youth") was removed, which increased the reliability to $\alpha = .675$ still below the .700 threshold (Hair et al., 1998). Therefore, item 18 ("Staff created an emotionally safe climate") was removed to improve the reliability of the component, which resulted in an $\alpha = .728$. The four item creation of a caring, mastery-oriented climate component was considered reliable, meeting the standard previously noted. The final component, use of curriculum, had a Cronbach's alpha slightly under the .700 threshold ($\alpha = .693$), however the results did not support the removal of any item. As the component's Cronbach's alpha is extremely close to the reliability standard adopted for the current study, the component will be maintained.
Upon PCA and reliability testing, the staff survey was found to have four underlying components that are both valid and reliable (see Table 2). The four components included: adherence to program structure - sports coaching (items 2, 3, 6, 7, 11); adherence to program structure - S.E.T.S. coaching (items 1, 4, 8, 9, 10, 12); creation of a caring, mastery-oriented climate (items 5, 19, 20, 21); and use of curriculum (items 14, 15, 16, 17). Items 13, 18, 22, and 23 were removed from the analysis. The four components reflected the staff’s program implementation fidelity which will be transposed to youth program dosage variables (which will be discussed within the data reduction section).

Youth Survey

The youth survey consisted of eight scales designed to measure youth outcomes associated with participation in the sport-based PYD summer camp. The scale constructs include: social competence, effort, teamwork, self-control, athletic competence, leader support, belonging, and engagement. The youth outcome scales used have been previously established in the literature, with validity and reliability evidence reviewed within the methods section. Although the validity and reliability of the scales have been previously tested, it is critical to reexamine the scales within the current study as valid and reliable scores is a necessary prerequisite for SEM (Kline, 2005; Schumacker & Lomax, 2010). CFA were conducted to confirm the validity of the scale scores, as the factor structure and stability have previously been established and therefore need to be confirmed rather than explored (Schumacker & Lomax). To establish the reliability of the scale scores, Cronbach’s alpha was calculated for the confirmed factors. The following
section demonstrates validity and reliability evidence for each youth outcome scale used in the current study.

**Social competence.** CFA of the 5-item PSCS-II was conducted to examine the factor structure of the social competence scale (Anderson-Butcher et al., in press). The initial CFA, with all items loading onto a single latent, resulted in a good fitting model, with all six global fit indices indicating an acceptable model fit [$\chi^2(5) = 8.70, p = .122$; NC = 1.74; RMSEA = .04; CFI = .99; GFI = .99; SRMR = .01]. With no theoretically justifiable modification suggestions, the factor structure of the 5-item scale was found to be empirically and theoretically sound.

The social competence youth outcome was measured at two separate times to capture change over time, therefore the factor stability of the scale scores were examined through LCFA. The confirmed 5-item scale was used as the measurement model for LCFA model testing. A series of progressively restrictive measurement models were tested to determine the factorial invariance of the measurement model (Table 3). The first model tested examined configural invariance, for which five of the six global fit indices confirmed the factor structure of the model, indicating a good fitting model. The weak invariance model, testing equal factor loadings across time, also resulted in an acceptable model fit with the majority of fit indices indicating a good fitting model. When examining the change in $\chi^2$ and CFI global fit indices for the weak invariance model, the change in CFI less than .01 supports the tenability of the factor structure and factor loading invariance constraints imposed. The final model (i.e., strong invariance) tested equal intercepts across time, which is the most restrictive invariance constraint imposed.
The strong invariance model was not upheld, as none of the six global fit indices met the standards previously noted, including the significant change in $\chi^2$ and .25 change in CFI. Weak invariance is the minimum requirement to suggest factorial invariance across time, therefore the results moderately support the tenability of factorial invariance across two time points.
Table 3

*Model fit statistics for the tests of invariance in social competence scale scores across two time points*

<table>
<thead>
<tr>
<th>Model Tested</th>
<th>$\chi^2$</th>
<th>$Df$</th>
<th>$p$</th>
<th>$\Delta \chi^2$</th>
<th>$\Delta df$</th>
<th>$P$</th>
<th>$\Delta$ CFI</th>
<th>$GFI$</th>
<th>RMSEA</th>
<th>NC</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Competence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>75.32</td>
<td>29</td>
<td>&lt;.001</td>
<td>--</td>
<td>--</td>
<td>.99</td>
<td>--</td>
<td>.97</td>
<td>.05</td>
<td>2.60</td>
<td>.03</td>
</tr>
<tr>
<td>2</td>
<td>89.94</td>
<td>33</td>
<td>&lt;.001</td>
<td>7.63</td>
<td>4</td>
<td>.106</td>
<td>.99</td>
<td>&lt;.01</td>
<td>.97</td>
<td>.05</td>
<td>2.73</td>
</tr>
<tr>
<td>3</td>
<td>1895.82</td>
<td>33</td>
<td>&lt;.001</td>
<td>1805.88</td>
<td>0</td>
<td>&lt;.001</td>
<td>.66</td>
<td>.25</td>
<td>.54</td>
<td>.32</td>
<td>57.45</td>
</tr>
</tbody>
</table>

*Note.* Model 1 tested for configural invariance, Model 2 tested for weak invariance, and Model 3 tested for strong invariance. $\chi^2 =$ chi-square; $Df =$ degrees of freedom; CFI = comparative fit index; GFI = goodness-of-fit index; RMSEA = root mean square error of approximation; NC = normed chi-square; SRMR = standardized root-mean square residual.
Having confirmed the factor structure and stability of the 5-item social competence measurement model, the reliability of the scale was tested using Cronbach’s alpha. Reliability testing was conducted at both the pre-test and post-test. At the pre-test, the 5-item scale demonstrated strong reliability ($\alpha = .846$); however, the results suggested an alpha increase of .053 if item two was removed. The scale showed strong reliability at the post-test as well, with an $\alpha = .899$ and no item deletion supported. As the 5-item scale was found valid and demonstrates a strong reliability with all items included, the complete scale will be maintained and is found to be valid and reliable.

**Effort.** To confirm the validity of the 5-item modified version of the MSOS commitment subscale, used to measure perceived effort, CFA were conducted (Vallerand et al., 1997). The initial CFA run resulted in an acceptable fitting model [$\chi^2(5) = 44.38, p < .001; \text{NC} = 8.88; \text{RMSEA} = .12; \text{CFI} = .96; \text{GFI} = .97; \text{SRMR} = .04$], with no theoretically justifiable modification suggestions indicating that a more parsimonious model would offer a better fit. Therefore, the factor structure of the 5-item measurement model was found to be empirically and theoretically supported.

The confirmed 5-item measurement model was tested for factorial invariance across time, as the scale was used at two separate time points. LCFA was employed to test the factor stability of the measurement model across time, using progressively restrictive models (i.e., configural invariance, weak invariance, strong invariance; Table 4). The configural and weak invariance models were found to have an acceptable fit, with four of the six global fit indices meeting the standards of model fit. Upon examination of the change in $\chi^2$ and CFI, the tenability of the factor structure and factor loading
invariance constraints were supported, therefore the configural and weak invariance models were upheld. LCFA of the strong invariance model resulted in a poor fitting model, with no global fit indices meeting the standards of model fit, including the significant change in $\chi^2$ and .28 change in CFI. As the weak invariance model was upheld, the LCFA provides moderate evidence of the factor stability of the measurement model across time.
Table 4

Model fit statistics for the tests of invariance in effort scale scores across two time points

<table>
<thead>
<tr>
<th>Model</th>
<th>Tested</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>( p )</th>
<th>( \Delta \chi^2 )</th>
<th>( \Delta df )</th>
<th>( P )</th>
<th>CFI</th>
<th>( \Delta CFI )</th>
<th>GFI</th>
<th>RMSEA</th>
<th>NC</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effort</td>
<td>1</td>
<td>161.37</td>
<td>28</td>
<td>&lt;.001</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.97</td>
<td>--</td>
<td>.95</td>
<td>.09</td>
<td>5.76</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>169.57</td>
<td>32</td>
<td>&lt;.001</td>
<td>8.20</td>
<td>4</td>
<td>.085</td>
<td>.97</td>
<td>&lt;.01</td>
<td>.95</td>
<td>.09</td>
<td>5.30</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1319.79</td>
<td>32</td>
<td>&lt;.001</td>
<td>1150.22</td>
<td>0</td>
<td>&lt;.001</td>
<td>.69</td>
<td>.28</td>
<td>.62</td>
<td>.27</td>
<td>41.24</td>
<td>.30</td>
</tr>
</tbody>
</table>

Note. Model 1 tested for configural invariance, Model 2 tested for weak invariance, and Model 3 tested for strong invariance. 

\( \chi^2 \) = chi-square; df = degrees of freedom; CFI = comparative fit index; GFI = goodness-of-fit index; RMSEA = root mean square error of approximation; NC = normed chi-square; SRMR = standardized root-mean square residual.
With the factor structure and stability of the 5-item measurement model confirmed, demonstrating the validity of the scale, reliability testing was subsequently conducted to examine the internal consistency of the scale. Cronbach’s alpha was calculated at the pre-test \((\alpha = .782)\) and post-test \((\alpha = .878)\), for which the scale showed strong reliability. The reliability testing did not support the removal of any item, therefore the 5-item measurement model assessing perceived effort was maintained and found to be valid and reliable.

**Teamwork.** CFA was conducted to confirm the factor structure and stability of the 8-item Teamwork Scale for Youth (Lower et al., in review). The initial CFA run demonstrated a good fitting model, with four of the six global fit indices indicating an acceptable model fit \([\chi^2(20) = 129.19, p < .001; \text{NC} = 6.46; \text{RMSEA} = .09; \text{CFI} = .97; \text{GFI} = .94; \text{SRMR} = .04]\). After reviewing the modification suggestions, there were no theoretically justifiable modifications which would significantly improve the fit of the model. Therefore, the 8-item teamwork measurement model was found empirically and theoretically sound.

As the Teamwork Scale for Youth was used as a longitudinal measure of the teamwork youth outcome, LCFA was conducted to examine the factor stability of the scale over time. Progressively restrictive invariance constraints were imposed on the model to test its stability across two time points (Table 5). The configural invariance model, examining factor structure, was first tested and found to be a good fitting model, with five of the six global fit indices meeting the standards of model fit. The next invariance constraint imposed was equal factor loadings (weak invariance model). The next
more restrictive model was found to have an acceptable fit and not significantly deteriorate from the configural invariance model. As the weak invariance model supported the factorial invariance of the measurement model across time, a final invariance constraint was imposed to test the measure’s factor stability. The strong invariance model significantly deteriorated, for which all global fit indices indicated a poor model fit, with a considerable change in $\chi^2$ and CFI. With the configural and weak invariance models supporting the factorial stability of the measurement model, there is enough evidence to confirm that the measure remains invariant across time.
Table 5

*Model fit statistics for the tests of invariance in teamwork scale scores across two time points*

<table>
<thead>
<tr>
<th>Model Tested</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>( p )</th>
<th>( \Delta \chi^2 )</th>
<th>( \Delta df )</th>
<th>( p )</th>
<th>CFI</th>
<th>( \Delta \text{CFI} )</th>
<th>GFI</th>
<th>RMSEA</th>
<th>NC</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teamwork</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>333.52</td>
<td>95</td>
<td>&lt;.001</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.97</td>
<td>--</td>
<td>.93</td>
<td>.07</td>
<td>3.51</td>
<td>.05</td>
</tr>
<tr>
<td>2</td>
<td>354.87</td>
<td>102</td>
<td>&lt;.001</td>
<td>21.35</td>
<td>7</td>
<td>.003</td>
<td>.97</td>
<td>&lt;.01</td>
<td>.93</td>
<td>.07</td>
<td>3.48</td>
<td>.05</td>
</tr>
<tr>
<td>3</td>
<td>2273.16</td>
<td>102</td>
<td>&lt;.001</td>
<td>1918.29</td>
<td>0</td>
<td>&lt;.001</td>
<td>.73</td>
<td>.24</td>
<td>.54</td>
<td>.200</td>
<td>22.29</td>
<td>.29</td>
</tr>
</tbody>
</table>

*Note.* Model 1 tested for configural invariance, Model 2 tested for weak invariance, and Model 3 tested for strong invariance. \( \chi^2 \) = chi-square; df = degrees of freedom; CFI = comparative fit index; GFI = goodness-of-fit index; RMSEA = root mean square error of approximation; NC = normed chi-square; SRMR = standardized root-mean square residual.
To test the reliability of the 8-item teamwork measurement model, Cronbach’s alpha was calculated at the pre-test and post-test. The measure demonstrated strong reliability at both the pre-test ($\alpha = .812$) and post-test ($\alpha = .873$). However, at the post-test, the results suggested the removal of item 3 to increase the reliability of the measure. As the 8-item scale demonstrated a strong reliability with all items maintained, and the suggested increase in alpha with the removal of item 3 was nominal (less than .01), the complete battery of items was maintained. Ultimately, the 8-item Teamwork Scale for Youth was found to be a valid and reliable measure.

**Self-control.** The self-control youth outcome was assessed through the 8-item Social Sports Experiences Scale (Anderson-Butcher et al., 2010). CFA were conducted to confirm the factor structure and stability of the scale prior to data reduction. The initial CFA tested the factor structure of the 8-item measurement model, resulting in an acceptable fitting model $[\chi^2(20) = 156.64, p < .001; \text{NC} = 7.832; \text{RMSEA} = .11; \text{CFI} = .97; \text{GFI} = .93; \text{SRMR} = .04]$. Upon examination of the model modification suggestions, no theoretically justifiable changes to the model were warranted. Therefore, the 8-item self-control measurement model was found to have a theoretically and empirically sound factor structure.

Similar to the additional social youth outcome measures, the Social Sports Experiences Scale was administered at two time points to measure change over time. Due to the longitudinal design of the scale, LCFA was conducted to confirm the factor stability of the scale scores (Table 6). To test the factor stability of the scale, progressively restrictive invariance constraints were imposed on the measurement model,
for which a non-significant change in $\chi^2$ and change in CFI less than .01 was indicative of model stability (i.e., the model did not deteriorate due to the added invariance constraint). The first model tested equal factor structure across time, for which the model was found to have an acceptable fit, with five of the six global fit indices indicating a good fitting model. The weak invariance model imposed an additional constraint of equal factor loadings, resulting in a good fitting model that did not significantly deteriorate (non-significant change in $\chi^2$ and change in CFI < .01). Lastly, the most restrictive model was tested (strong invariance), with the imposed constraint of equal intercepts. Unlike the first two models, the strong invariance model was not upheld, for which the global fit indices suggested a poor fitting model and the change in $\chi^2$ and CFI indicated significant model deterioration. As the minimum requirements for factor stability were met (i.e., weak invariance), there is reasonable evidence of invariance across time.
Table 6

Model fit statistics for the tests of invariance in self-control scale scores across two time points

<table>
<thead>
<tr>
<th>Model Tested</th>
<th>$\chi^2$</th>
<th>$df$</th>
<th>$p$</th>
<th>$\Delta \chi^2$</th>
<th>$\Delta df$</th>
<th>$p$</th>
<th>CFI</th>
<th>$\Delta$ CFI</th>
<th>GFI</th>
<th>RMSEA</th>
<th>NC</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>431.60</td>
<td>95</td>
<td>&lt;.001</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>.97</td>
<td>--</td>
<td>.90</td>
<td>.08</td>
<td>4.54</td>
<td>.05</td>
</tr>
<tr>
<td>2</td>
<td>442.35</td>
<td>102</td>
<td>&lt;.001</td>
<td>10.75</td>
<td>7</td>
<td>.150</td>
<td>.97</td>
<td>&lt;.01</td>
<td>.90</td>
<td>.08</td>
<td>4.34</td>
<td>.05</td>
</tr>
<tr>
<td>3</td>
<td>2608.30</td>
<td>102</td>
<td>&lt;.001</td>
<td>2165.95</td>
<td>0</td>
<td>&lt;.001</td>
<td>.78</td>
<td>.19</td>
<td>.50</td>
<td>.21</td>
<td>25.57</td>
<td>.32</td>
</tr>
</tbody>
</table>

Note. Model 1 tested for configural invariance, Model 2 tested for weak invariance, and Model 3 tested for strong invariance. $\chi^2 = \text{chi-square}; df = \text{degrees of freedom}; CFI = \text{comparative fit index}; GFI = \text{goodness-of-fit index}; RMSEA = \text{root mean square error of approximation}; NC = \text{normed chi-square}; SRMR = \text{standardized root-mean square residual}.$
To test the reliability of the 8-item self-control measurement model, Cronbach’s alpha was calculated at the pre-test and post-test. Results of the analysis revealed strong reliability at both time points (pre-test: $\alpha = .865$; post-test: $\alpha = .905$). The reliability testing did not suggest item deletion to strengthen the internal consistency of the scale, therefore all eight items were maintained. The results support the validity and reliability of the self-control measurement model.

**Athletic competence.** To assess the validity and reliability of the 3-item Perceived Athletic Competence scale, CFA and reliability testing were employed (Amorose, 2002). CFA of the factor structure was the first model tested, for which the initial run resulted in a perfect fit as the model was saturated [$\chi^2(0) = 0, p = 1; \text{RMSEA} = 0$]. With these results there was no need to consider model modifications.

To confirm the factor stability of the 3-item scale, LCFA was conducted using progressively restrictive invariance constraints (Table 7). The first model tested the stability of the factor structure across time, for which the model was found to be an acceptable fit with all global fit indices indicating a good fitting model. The weak invariance model imposed equal factor loadings, and was upheld, demonstrated by five of the six global fit indices and the non-significant change in $\chi^2$ and change in CFI less than .01. The final model tested was the strong invariance model, constraining the intercepts to be equal. The strong invariance model failed to be upheld as all global fit indices indicated a poor fitting model and the change in $\chi^2$ and CFI did not meet the standards previously outlined. Although the strong invariance model was not upheld, the configural and weak invariance models support the tenability of factor stability across time.
Table 7

Model fit statistics for the tests of invariance in athletic competence scale scores across two time points

<table>
<thead>
<tr>
<th>Model</th>
<th>Tested</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p</th>
<th>$\Delta \chi^2$</th>
<th>$\Delta df$</th>
<th>P</th>
<th>CFI</th>
<th>$\Delta$ CFI</th>
<th>GFI</th>
<th>RMSEA</th>
<th>NC</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>7.49</td>
<td>4</td>
<td>.112</td>
<td>--</td>
<td>--</td>
<td>.99</td>
<td>--</td>
<td>.99</td>
<td>.04</td>
<td>1.87</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>9.12</td>
<td>6</td>
<td>&lt;.001</td>
<td>1.62</td>
<td>2</td>
<td>.444</td>
<td>.99</td>
<td>&lt;.01</td>
<td>.99</td>
<td>.03</td>
<td>1.52</td>
<td>.02</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>478.08</td>
<td>6</td>
<td>&lt;.001</td>
<td>468.96</td>
<td>0</td>
<td>&lt;.001</td>
<td>.83</td>
<td>.17</td>
<td>.80</td>
<td>.39</td>
<td>79.68</td>
<td>.28</td>
</tr>
</tbody>
</table>

Note. Model 1 tested for configural invariance, Model 2 tested for weak invariance, and Model 3 tested for strong invariance. $\chi^2$ = chi-square; df = degrees of freedom; CFI = comparative fit index; GFI = goodness-of-fit index; RMSEA = root mean square error of approximation; NC = normed chi-square; SRMR = standardized root-mean square residual.
The confirmed 3-item athletic competence measurement model was tested for reliability at the pre-test and post-test. The results showed strong reliability at both time points (pre-test: $\alpha = .854$; post-test: $\alpha = .874$), supporting the internal consistency of the scale. The reliability testing did not suggest the removal of any item, therefore all three items were maintained, meeting the minimum factor requirement of three items. Overall, the results provide substantial evidence of the validity and reliability of the 3-item athletic competence measurement model.

**Belonging.** A modified version of the 5-item sense of belonging scale was used to assess perceived belonging to the program under investigation and was administered at post-test only (Anderson-Butcher & Conroy, 2002). To confirm the factor structure of the scale, CFA was employed. The initial CFA run resulted in a good fitting model, with five of the six global fit indices indicating an acceptable model fit [$\chi^2(9) = 33.92, p < .001$; NC = 3.77; RMSEA = .08; CFI = .99; GFI = .97; SRMR = .02]. There were no theoretically justifiable modification suggestions to warrant the removal of any items, therefore the 5-item measurement model was maintained and considered empirically and theoretically sound. To examine the reliability of the confirmed measurement model, Cronbach’s alpha was calculated at the post-test, demonstrating strong internal consistency ($\alpha = .908$). Results of the reliability testing did not suggest the removal of any items, supporting the reliability of the full measure. Based upon these results, the 5-item belonging measurement model was found valid and reliable.

**Leader support.** Validity and reliability of the 8-item modified PSSM teacher support subscale was examined through CFA and reliability testing (Cox & Williams,
Previous analysis had suggested the removal of the negatively worded items as they were problematic to the measurement model, therefore the current study added two items to reflect the positive conceptualization of the negatively worded items with the purpose of finding the best fitting model. The 8-item model contains both the negatively and positively worded corresponding items.

CFA of the 8-item belonging measurement model resulted in a poor fitting model, with only one of the six global fit indices indicated an acceptable model fit [$\chi^2(20) = 214.66, p < .001; NC = 10.73; RMSEA = .15; CFI = .88; GFI = .91; SRMR = .09$]. Upon examination of the results, item 6 and 7 (the negatively worded items) were found to be problematic. The $R^2$ of items 6 ($R^2 = .073$) and 7 ($R^2 = .0663$) explained a nominal amount of the variance in belonging scores. Additionally, the standardized residual between items 6 and 7 was large (9.942), suggesting that the covariance between the items is not well explained by the model. Lastly, when examining the modification suggestions, the output suggests the addition of an error covariance between items 6 and 7, with an anticipated decrease in $\chi^2$ by 98.6.

Based upon this preponderance of evidence, the two negatively worded items (items 6 and 7) were removed and the positively worded items (items 1 and 2) were maintained. CFA of the 6-item measurement model resulted in a good fitting model, with five of the six global fit indices indicating an acceptable model fit [$\chi^2(9) = 21.92, p < .001; NC = 2.44; RMSEA = .06; CFI = .99; GFI = .98; SRMR = .03$]. Review of the modification suggestions revealed no theoretically justifiable modifications to adopt to
improve the fit of the model, therefore the 6-item measurement model was maintained and found valid.

To establish the reliability of the 6-item leader support measurement model, Cronbach’s alpha was calculated at the post-test. The results demonstrated strong internal consistency ($\alpha = .822$), however suggested the removal of item 1 for a .02 increase in alpha. As the 6-item model was found valid and reliable with all items maintained, the full measure was maintained. Overall, the 6-item measurement model was found to be the best fitting model, with strong internal consistently, supporting the validity and reliability of the modified scale.

**Engagement.** Perceived quality of youth engagement was measured using a 5-item behavioral engagement subscale (Fredricks et al., 2004; IRRE, 1998). To confirm the validity of the measure, CFA was conducted. The initial CFA run resulted in an acceptable model fit [$\chi^2(5) = 48.942, p < .001; \text{NC} = 9.79; \text{RMSEA} = .15; \text{CFI} = .96; \text{GFI} = .95; \text{SRMR} = .04$]. When reviewing the results, there were no theoretically justifiable modification suggestions that warranted consideration, therefore the 5-item measurement model was maintained and supported as valid.

Reliability testing of the 5-item engagement measurement model was employed at the post-test. Results of the analysis demonstrated strong internal consistency ($\alpha = .849$), with no item deletion suggested. Based upon the results of the CFA and reliability testing, the 5-item engagement measurement model has been found valid and reliable.

In regards to quantity of engagement, attendance data were collected as part of the program evaluation and used as secondary data to measure quantity of engagement in the
program under investigation. Content validity of the measure is evident based upon the use of attendance data to reflect engagement in previous studies (e.g., Anderson-Butcher & Cash, 2010; Anderson-Butcher & Fink, 2005; Gottfredson et al., 2010). As the data consisted of a single variable, construct validity and reliability testing is not appropriate.

**Data Reduction**

To prepare the data for SEM, the number of items from the questionnaire were reduced to a smaller number of components, through the creation of factor scores. Calculating a mean or sum score based upon the dimensionality of the data (which variables load together) is a common data reduction practice (DiStefano, Zhu, & Mindrila, 2009). Factor scores for the social competence constructs (i.e., general social competence, effort, teamwork, self-control), belonging, leader support, and engagement were created by calculating the weighted mean of the confirmed factor using factor loadings. By taking factor loadings into consideration, the items with the largest factor loadings (explaining the most amount of variance in the construct) had the greatest effect on the factor score. A factor score was not created for the athletic competence construct as the individual items were used as observed variables in the structural model. In relation to youth attendance, reflecting quantity of engagement, the data were aggregated into a sum score to reflect total attendance throughout the 19 day summer camp.

To create program dosage variables at the youth individual level, the staff program implementation data were first organized by which youth were exposed to the program implementation. By examining the camp schedule (which groups of youth were with which staff members during which sessions) and youth attendance data (who
attended which sessions), the researcher was able to correspond the specific youth
exposed to specific staff member’s program implementation during specific sessions. In
Microsoft Excel 2013, the youth who were in attendance for a particular session were
associated with the program implementation data of the specific staff members
(recreation leaders, education leaders, counselors) facilitating the session. Youth who
were not in attendance for a particular session were not associated with program
implementation data for that session as they were not exposed to the program at that time.

Once the staff program implementation data were corresponded with the specific
youth exposed to the program implementation, the large data set (upwards of 4,320 data
points per youth) was first reduced by calculating a mean score across the staff
implementing the program. For example, youth who attended a swimming session would
receive program implementation data for that session (24 variables) from the six
swimming recreation leaders and their counselor (168 data points); the seven staff
members’ data were averaged to create a mean program dosage score (program
implementation the youth were exposed to) for each of the 24 variables.

Following the creation of mean program dosage scores for every program session
the youth attended, mean program dosage factor scores were calculated for each of the
four confirmed program implementation components: adherence to program structure –
sports coaching (i.e., ‘sports coaching’); adherence to program structure – S.E.T.S.
coaching (i.e., ‘S.E.T.S. coaching’); creation of a caring, mastery-oriented climate (i.e.,
‘environment’); use of curriculum (i.e., ‘curriculum’). The mean factor scores were
calculated by taking the average of the variables, which loaded onto the corresponding
components and were found reliable, across all 60 sessions (4 sessions per day for 15
days of curriculum). An average across all 60 sessions was calculated, rather than a
summation, to eliminate the influence of attendance, therefore controlling for
multicollinearity between the program dosage and attendance factors. Therefore, each
youth had a mean sports coaching, S.E.T.S. coaching, environment, and curriculum factor
score reflecting their degree of exposure to the primary program components. At the
completion of data reduction, the data set was reduced to 16 factor scores [i.e., social
competence (pre- and post-test), effort (pre- and post-test), teamwork (pre- and post-test),
self-control (pre- and post-test), belonging, leader support, engagement, attendance,
sports coaching, S.E.T.S. coaching, environment, curriculum], with 6 variables (i.e.,
athletic competence items at the pre- and post-test) maintained.

**Normality**

Prior to conducting SEM, the data were screened for univariate and multivariate
normality. Normality is an assumption of SEM, as data that deviate from a normal
distribution will influence the variance-covariance among variables, parameter estimates,
and standard errors (Schumacker & Lomax, 2010). To establish univariate normality for
each variable, the skewness and kurtosis scores were reviewed. Skewness scores within
the range of +2.00 to -2.00 and kurtosis scores within the range of +5.00 to -5.00, were
considered evidence of univariate normality (Kendall & Stuart, 1958). Table 8 provides
an overview of the descriptive statistics, including skewness and kurtosis scores, for all
observed variables under investigation.
Table 8

*Descriptive statistics of observed variables*

<table>
<thead>
<tr>
<th>Observed Variables</th>
<th>$N$</th>
<th>$M$</th>
<th>$SD$</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Competence (Pre)</td>
<td>417</td>
<td>3.73</td>
<td>0.81</td>
<td>-0.478</td>
<td>-0.259</td>
</tr>
<tr>
<td>Social Competence (Post)</td>
<td>417</td>
<td>3.95</td>
<td>0.91</td>
<td>-0.778</td>
<td>0.093</td>
</tr>
<tr>
<td>Effort (Pre)</td>
<td>417</td>
<td>4.06</td>
<td>0.76</td>
<td>-0.819</td>
<td>0.262</td>
</tr>
<tr>
<td>Effort (Post)</td>
<td>417</td>
<td>4.20</td>
<td>0.80</td>
<td>-1.037</td>
<td>0.574</td>
</tr>
<tr>
<td>Teamwork (Pre)</td>
<td>417</td>
<td>3.88</td>
<td>0.70</td>
<td>-0.617</td>
<td>0.326</td>
</tr>
<tr>
<td>Teamwork (Post)</td>
<td>417</td>
<td>4.00</td>
<td>0.78</td>
<td>-0.730</td>
<td>0.255</td>
</tr>
<tr>
<td>Self-Control (Pre)</td>
<td>417</td>
<td>4.22</td>
<td>0.64</td>
<td>-0.897</td>
<td>0.394</td>
</tr>
<tr>
<td>Self-Control (Post)</td>
<td>417</td>
<td>4.19</td>
<td>0.73</td>
<td>-0.926</td>
<td>0.420</td>
</tr>
<tr>
<td>Athletic Competence Item 1 (Pre)</td>
<td>417</td>
<td>4.26</td>
<td>0.80</td>
<td>-1.121</td>
<td>1.660</td>
</tr>
<tr>
<td>Athletic Competence Item 1 (Post)</td>
<td>417</td>
<td>4.36</td>
<td>0.82</td>
<td>-1.331</td>
<td>1.865</td>
</tr>
<tr>
<td>Athletic Competence Item 2 (Pre)</td>
<td>417</td>
<td>4.07</td>
<td>0.85</td>
<td>-0.795</td>
<td>0.640</td>
</tr>
<tr>
<td>Athletic Competence Item 2 (Post)</td>
<td>417</td>
<td>4.26</td>
<td>0.84</td>
<td>-1.104</td>
<td>1.288</td>
</tr>
<tr>
<td>Athletic Competence Item 3 (Pre)</td>
<td>417</td>
<td>4.19</td>
<td>0.87</td>
<td>-0.899</td>
<td>0.391</td>
</tr>
<tr>
<td>Athletic Competence Item 3 (Post)</td>
<td>417</td>
<td>4.29</td>
<td>0.88</td>
<td>-1.320</td>
<td>1.718</td>
</tr>
<tr>
<td>Attendance*</td>
<td>417</td>
<td>17.47</td>
<td>1.95</td>
<td>-2.504</td>
<td>10.638</td>
</tr>
<tr>
<td>Behavioral Engagement</td>
<td>417</td>
<td>3.37</td>
<td>0.59</td>
<td>-0.992</td>
<td>1.039</td>
</tr>
<tr>
<td>Belonging</td>
<td>417</td>
<td>3.98</td>
<td>0.93</td>
<td>-1.083</td>
<td>0.820</td>
</tr>
</tbody>
</table>

Continued
Table 8 continued

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leader Support</td>
<td>417</td>
<td>3.82</td>
<td>0.93</td>
<td>-0.823</td>
<td>0.291</td>
</tr>
<tr>
<td>Program Structure (Sports Coaching)*</td>
<td>417</td>
<td>7.73</td>
<td>0.34</td>
<td>-3.474</td>
<td>38.335</td>
</tr>
<tr>
<td>Program Structure (S.E.T.S. Coaching)</td>
<td>417</td>
<td>6.79</td>
<td>0.54</td>
<td>-1.185</td>
<td>2.989</td>
</tr>
<tr>
<td>Program Environment*</td>
<td>417</td>
<td>8.58</td>
<td>0.35</td>
<td>-3.146</td>
<td>35.356</td>
</tr>
<tr>
<td>Program Curriculum*</td>
<td>417</td>
<td>6.61</td>
<td>0.49</td>
<td>-1.207</td>
<td>6.447</td>
</tr>
</tbody>
</table>

Note. *Observed variables do not meet the standards of univariate normality and were transformed to achieve a normal distribution.

The univariate normality assumption was met for 18 of the 22 observed variables, for which the attendance, program structure (sports coaching), program structure (S.E.T.S. coaching), program environment, and program curriculum observed variables were found nonnormal. To address the univariate nonnormality, transformations were conducted on the nonnormal observed variables to convert the original scores to new scores that are more normally distributed (Kline, 2005). For negatively skewed data, a power transformation, using a power greater than one, can reduce the skewness (Bruce et al., 2008; von Hippel, 2004). As the attendance and program curriculum observed variables were only moderately nonnormal, with a negative skew and leptokurtic distribution, a quadratic power transformation was conducted to address the nonnormality (Table 9; Bruce et al., 2008). In relation to the program structure (sports coaching) and program environment observed variables, the data were negatively skewed with a severe
leptokurtic distribution, therefore a cubic power transformation was employed to address the nonnormality (Table 9; Bruce et al., 2008). The power transformations improved the normality of the observed variable distributions, with the attendance and program curriculum variables meeting the standards of univariate normality, and the program structure (sports coaching) and program environment variables achieving an acceptable distribution (Kendall & Stuart, 1998; Kline, 2005). Upon completion of the transformations, the observed variables have met the assumption of univariate normality.

Table 9

*Descriptive statistics of transformed observed variables*

<table>
<thead>
<tr>
<th>Transformed Observed Variables</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance – Squared</td>
<td>417</td>
<td>16.27</td>
<td>313.00</td>
<td>-1.539</td>
<td>3.002</td>
</tr>
<tr>
<td>Program Structure (Sports Coaching) – Cubed</td>
<td>417</td>
<td>4.65</td>
<td>0.55</td>
<td>-0.773</td>
<td>7.149</td>
</tr>
<tr>
<td>Program Environment - Cubed</td>
<td>417</td>
<td>6.34</td>
<td>0.70</td>
<td>-0.742</td>
<td>6.960</td>
</tr>
<tr>
<td>Program Curriculum - Squared</td>
<td>417</td>
<td>4.39</td>
<td>0.62</td>
<td>-0.517</td>
<td>2.277</td>
</tr>
</tbody>
</table>

To establish multivariate normality, Mardia’s measure of relative multivariate kurtosis was examined, for which a score less than the 3.0 standard was considered to be evidence of multivariate normality (Siekpe, 2005). The relative multivariate kurtosis of the current transformed data set was 1.29, demonstrating that the multivariate normality assumption was met. Overall, the results provide evidence that the data has met the univariate and multivariate assumption of SEM. As a supportive measure, the asymptotic
covariance matrix was inputted into the SEM analyses, along with the sample covariance matrix, to address any remaining nonnormality (Schumacker & Lomax, 2010).

**Measurement Models**

The two-step approach to testing structural models was selected for its advancement, simplicity, and flexibility (Kline, 2005). The first step of this approach is to test the measurement models using CFA to determine the best fitting models, prior to combining the measurement models into a structural model. This section will review the results of the social competence, athletic competence, and program dosage measurement models.

**Social Competence**

Social competence is reflected by an individual’s ability to successfully engage in positive social interactions with others and includes a multitude of social skills (Anderson-Butcher et al., in press). Social skills can be any ability or behavior which enables one to be effective within social situations (Anderson-Butcher et al., in press). The current study examined general social competence, effort, teamwork, and self-control as social skills which contribute to a youth’s social competency. As such, the social competence measurement model, at pre-test and post-test, is specified by the four social skill indicators (i.e., general social competence, effort, teamwork, and self-control) loading onto the social competence latent factor.

CFA were conducted to examine the social competence measurement model at the pre-test and post-test. The social competence measurement model at the pre-test was tested first. The general social competence effort, teamwork, and self-control observed
variables were specified to load onto the social competence latent factor, for which the item-specific errors were not allowed to correlate as each item represented a unique social skill. The initial CFA was run to determine which observed variable should act as the reference variable to allow comparison among the factor loadings. The teamwork observed variable was found to explain the most variance in social competence ($R^2 = .838$, and therefore was fixed to one to serve as the reference variable. When evaluating the model (Figure 1), the statistical significance of the $t$-values for the individual paths were examined, for which all parameter estimates were found significant at the $\alpha < .05$ level of significance. The global fit indices were then reviewed to determine the fit of the model, for which five of the six indices upheld the standards of model fit previously outlined [$\chi^2(2) = 7.44, p = .024; \text{NC} = 3.72; \text{RMSEA} = .08; \text{CFI} = .99; \text{GFI} = .99; \text{SRMR} = .02$] suggesting a good fitting model.
Once the global fit indices were examined, the $R^2$, standardized residuals, MIs, and EPC statistics were evaluated. The moderate to large $R^2$ support the inclusion of all indicators (social competence $R^2 = .60$; effort $R^2 = .36$; teamwork $R^2 = .84$; self-control $R^2 = .63$). Upon review of the standardized residuals, all residuals fell within the +/- 1.96 standard, suggesting that all correlations are well accounted for in the model (Schumacker & Lomax, 2010). The final output to consider were the MIs and EPC statistics. There were no modification suggestions that were theoretically justifiable and would result in a considerable improvement in model fit. Based upon the results, the social competence measurement model was found to be theoretically and empirically sound at the pre-test.
The same CFA process was conducted for the social competence measurement model at the post-test to confirm the factor structure of the model. Upon the initial CFA run, once again the teamwork social skill accounted for the greatest amount of variance in social competence ($R^2 = .860$), and therefore was fixed to one to act as the reference variable. All parameter estimates were found statistically significant, supporting the inclusion of all indicators. When considering model fit, three of the six global fit indices suggested a good fitting model [$\chi^2(2) = 19.73, p < .001; \text{NC} = 9.86; \text{RMSEA} = .15; \text{CFI} = .99; \text{GFI} = .98; \text{SRMR} = .02$].

As three global fit indices suggested a poor model fit, the $R^2$ were examined. Similar to the pre-test, the strong $R^2$ supports the inclusion of all indicators (social competence $R^2 = .61$; effort $R^2 = .54$; teamwork $R^2 = .86$; self-control $R^2 = .75$). The standardized residuals suggested a few potential issues, with the residual between teamwork and social competence (5.19) exceeding the +/- 1.96 standard. The relationships between effort and social competence (-3.34) as well as effort and self-control (2.14) moderately exceeded the +/- 1.96 standard. These results suggest that these relationships are not well accounted for in the model. The MIs suggest the addition of an error covariance between effort and self-control and between self-control and social competence, with an anticipated decrease in $\chi^2$ by 17.1 and 29.6 respectively. As the standardized residual and MI suggested that the relationship between effort and self-control is not well accounted for in the model, the error covariance between the observed variables was allowed to correlate.
The modification to the social competence measurement model at the post-test considerably improved the model (Figure 2). All parameter estimates remained significant, supporting the inclusion of all indicators. Upon review of the global fit indices, the model was found to have an acceptable fit, with all fit indices meeting the standards of model fit [\(\chi^2(1) = 0.78, p = .378\); NC = 0.78; RMSEA < .01; CFI = 1.00; GFI = .99; SRMR < .01]. The \(R^2\), standardized residuals, and MIs were reviewed to determine if any modifications were theoretically justifiable. The \(R^2\) remained strong (social competence \(R^2 = .61\); effort \(R^2 = .48\); teamwork \(R^2 = .91\); self-control \(R^2 = .70\)) and all standardized residuals fell within the +/- 1.96 standard. In relation to the MIs, the suggested error covariance between self-control and social competence was associated with a minimal decrease in \(\chi^2\). Based upon these results, the modified social competence measurement model at the post-test was found theoretically and empirically sound.
Figure 2

Social competence measurement model at post-test with standardized solutions

![Diagram showing the measurement model of social competence with standardized solutions.](image)

*Note. Standardized solutions are reported. The fixed factor is denoted by ‘F’. *** p < .001.*

**Athletic Competence**

A common objective of physical activity and sport programs is the development of athletic competence. Perceived athletic competence refers to an individual’s belief about their ability within a physical activity or sport context (Ridgers, Fazey, & Fairclough, 2007). More specifically, competence involves one’s ability to handle the required tasks of a given achievement domain (Ridgers et al., 2007). To assess athletic competence within the current study, the 3-item Perceived Athletic Competence scale was used (Amorose, 2002). Each of the three items acts as an indicator of the latent athletic competence construct. Although the validity of the scale (factor structure and
invariance across time) was confirmed, the measurement models at the pre-test and post-test were still tested to ensure no measurement issues.

CFA of the athletic competence measurement model at the pre-test were conducted, for which the initial run was employed to determine which indicator should act as the reference variable. The factor loadings of the three indicators were quite similar, with item 2 accounting for the greatest amount of variance in the latent construct ($R^2 = .70$) and therefore was fixed to one to act as the reference variable. Results of the CFA (Figure 3) revealed statistically significant parameter estimates for the individual paths, at the $\alpha < .05$ level of significance, supporting the inclusion of all three items. Upon review of the goodness of fit statistics, the findings indicated a perfect fit, for which the model was saturated [$\chi^2(0) = 0$, $p = 1.0$], with no standardized residuals nor modification suggestions to review. Therefore the athletic competence measurement model at the pre-test was found theoretically and empirically sound.
To confirm the best fitting athletic competence measurement model at the post-test, CFA was employed. Similar to the pre-test, each of the three indicators demonstrated large factor loadings, with item 2 accounting for the greatest amount of variance in the athletic competence latent construct ($R^2 = .771$). Item 2 was fixed to one to act as the reference variable in order to allow for comparison among the factor loadings. The results were quite similar to the pre-test CFA (Figure 4), with all three indicators demonstrating statistical significance at the $\alpha < .05$ level of significance, suggesting that all items be maintained. The model was found to be saturated, thus having a perfect fit [$\chi^2(0) = 0, p = 1.0$]. As the model was saturated, there were no standardized residuals or MIs to review. Therefore, based upon the results, the athletic competence measurement model at the post-test was found theoretically and empirically sound.
Figure 4

*Athletic competence measurement model at post-test with standardized solutions*

Note. Standardized solutions are reported. The fixed factor is denoted by ‘F’. *** p < .001.

**Program Dosage**

Program dosage refers to the degree of exposure to an intervention a participant experiences (Nation et al., 2003). Those in attendance of the summer sports camp under investigation were exposed to the program intervention, thus receiving a degree of program dosage. While those absent on a given day were not exposed to the program intervention and therefore did not receive program dosage on days absent. Dosage can be measured by the quantity and quality of the exposure (Nation et al., 2003). For the current study, the quantity of program dosage the youth were exposed to was assessed through total number of days attended (out of 19 days) and the four program dosage variables reported by the staff, including adherence to program structure (sports coaching), adherence to program structure (S.E.T.S. coaching), use of curriculum, and creation of a caring, mastery-oriented environment. Quality of program dosage the youth were
exposed to was measured through perceived belonging, leader support, and engagement reported by the youth.

**Single factor measurement model.** CFA were conducted to determine the best fitting program dosage measurement model, for which program dosage was a single latent construct. The initial run was conducted to determine which indicator should act as the reference variable to allow for comparison among the factor loading. The quality of program dosage variables reported the greatest factor loadings, with leader support accounting for the greatest variance in program dosage ($R^2 = .624$) and therefore was fixed to one to act as the reference variable. The second CFA, with leader support fixed to one, resulted in a poor fitting model, with all global fit indices failing to meet the standards of model fit [$\chi^2(20) = 665.33, p < .001; \text{NC} = 33.27; \text{RMSEA} = .28; \text{CFI} = .74; \text{GFI} = .73; \text{SRMR} = .20$].

When evaluating the statistical significance of the $t$-values for the individual paths, seven parameter estimates were found statistically significant at the $\alpha < .05$ level of significance. The use of curriculum indicator was non-significant ($p = .139$) and accounted for less than 1% of the variance in the program dosage latent variable. The standardized residuals were examined, revealing potential issues with the program implementation indicators. For example, the standardized residual between the program implementation indicators primarily ranged from 3.79 to 4.97, with one residual equal to 0.24. Upon review of the major MIs, the addition of an error covariance between sports coaching and S.E.T.S. coaching (142.1 decrease in $\chi^2$), environment and S.E.T.S. coaching (88.7 decrease in $\chi^2$), environment and sports coaching (196.7 decrease in $\chi^2$),
and curriculum and sports coaching (72.2 decrease in $\chi^2$) were suggested. Based upon the results, the first modification made was the removal of the non-significant use of curriculum indicator.

The removal of the use of curriculum observed variable slightly improved the fit of the model, however the model was still a poor fit. The goodness of fit statistics fail to meet the standards of model fit $[\chi^2(14) = 463.95, p < .001; \text{NC} = 33.14; \text{RMSEA} = .28; \text{CFI} = .77; \text{GFI} = .76; \text{SRMR} = .19]$, suggesting a poor fitting model. Upon examination of the statistical significance and factor loading of the individual paths, although all parameter estimates were found significant at the $\alpha < .05$ level of significance, the program implementation observed variables and attendance variable each accounted for only 2% of the variance in the program dosage latent variable. In comparison, the quality of program dosage observed variables (i.e., belonging, leader support, engagement) each explained 50 to 63% of the variance in program dosage. The largest standardized residuals were between the adherence to program structure and environment variables (ranging 4.21 to 5.00), suggesting that the covariance between these variables is not well explained by the model. Lastly, the MIs once again suggested the addition of an error covariance between sports coaching and S.E.T.S. coaching (142.6 decrease in $\chi^2$), environment and S.E.T.S. coaching (89.1 decrease in $\chi^2$), and environment and sports coaching (196.9 decrease in $\chi^2$).

As the attendance and program implementation observed variables were equally poor indicators of the program dosage latent construct, the attendance variable was removed first to determine where the biggest measurement issues were. The modification
did not improve the fit of the model \[\chi^2(9) = 453.96, p < .001; \text{NC} = 50.44; \text{RMSEA} = .34; \text{CFI} = .81; \text{GFI} = .74; \text{SRMR} = .22\]. The large standardized residuals between the adherence to program structure and environment variables remained (ranging 4.20 to 4.99), as well as the MIs between the adherence to program structure and environment variables associated with a large anticipated decrease in \(\chi^2\) (ranging 88.9 to 196.8). Based upon these results, the environment observed variable was removed to try and improve the fit of the model.

The modified program dosage measurement model (curriculum, attendance, and environment removed) was found to be a poor fit, with the majority of global fit indices failing to uphold the standards of model fit \[\chi^2(5) = 184.35, p < .001; \text{NC} = 36.87; \text{RMSEA} = .29; \text{CFI} = .84; \text{GFI} = .87; \text{SRMR} = .15\]. When reviewing the individual paths of the parameter estimates, the sports coaching observed variable was found statistically non-significant \((p = .054)\). All other parameter estimates were found significant at the \(\alpha < .05\) level of significance. The largest standardized residual (5.04) and MI (143.2 decrease in \(\chi^2\)) was between the adherence to program structure variables. Based upon the non-significant parameter estimate and large standardized residual and EPC statistic, the sports coaching observed variable was removed to try and improve the fit of the model.

The modification, removing the sports coaching observed variable, considerably improved the fit of the model. Goodness of fit statistics reported an acceptable fitting model, with all global fit indices indicating a good fitting model \[\chi^2(2) = 3.72, p = .156; \text{NC} = 1.86; \text{RMSEA} = .05; \text{CFI} = .99; \text{GFI} = .99; \text{SRMR} = .02\]. All parameter estimates were found statistically significant, however the S.E.T.S. coaching observed variable \((p = .054)\).
.046) only explained 1.4\% of the variance in the program dosage latent construct. The standardized residuals fell within the accepted +/-1.96 range and no MIs were suggested. Although the modified program dosage measurement model is a good fit, with acceptable standardized residuals and no modification suggestions, S.E.T.S. coaching remains a poor indicator of the program dosage latent variable. Therefore, this observed variable was removed to improve the measurement model.

The final modified single factor program dosage measurement model (Figure 5; curriculum, attendance, environment, sports coaching, S.E.T.S. coaching removed) was saturated, demonstrating a perfect fit [$\chi^2(0) = 0.0, p = 1.0, \text{RMSEA} = 0.0$]. All parameter estimates were found statistically significant and accounted for 49 to 64\% of the variance in the program dosage latent construct. There were no standardized residuals nor MIs to consider, therefore the modified program dosage measurement model, defined by behavioral engagement, belonging, and leader support, was found theoretically and empirically sound.
As the observed variables assessing quantity of program dosage (attendance, sports coaching, S.E.T.S. coaching, curriculum, environment) seem to be poor indicators of the comprehensive program dosage latent construct, a multiple factor program dosage measurement model was considered. The model consists of two factors, including quantity of program dosage and quality of program dosage. The quantity of program dosage latent construct was defined by youth attendance, adherence to program structure (sports coaching), adherence to program structure (S.E.T.S. coaching), use of curriculum, and creation of a caring, mastery-oriented climate. While the quality of program dosage latent construct was defined by the belonging, leader support, and behavioral engagement observed variables.

**Multiple factor measurement model.** The quantity and quality of program dosage observed variables were divided into separate factors to determine if a multiple
factor measurement model might more accurately define program dosage. The individual measurement models were tested separately to determine the best fitting models. Once the best fitting measurement models were confirmed, a two factor measurement model was tested. Subsequently, the two-factor measurement model was compared to the modified one factor program dosage measurement model (with the curriculum observed variable removed) to determine the best fitting measurement model defining program dosage.

CFA of the quantity of program dosage measurement model were conducted to confirm this individual factor separately from quality of program dosage. The indicator variables included attendance, sports coaching, S.E.T.S. coaching, curriculum, and environment. The initial run was conducted to determine which observed variable could act as the reference variable, thus allowing for comparison across factor loadings. Results of the initial CFA indicated problems with model which hindered the selection of a reference variable. More specifically, the sports coaching observed variable was associated with a negative error variance. Prior to constraining a reference variable to one, the output was reviewed to determine what modification might eliminate the presence of a negative error variance among the observed variables.

When evaluating the model, the statistical significance of the individual paths were examined, for which the attendance observed variable was non-significant at the \( \alpha < .05 \) level of significance \( (p = .347) \). The global fit indices indicated a poor model fit, with only one of the six fit indices upholding the standards of model fit \( \chi^2(5) = 122.64, p < .001; \text{NC} = 24.53; \text{RMSEA} = .24; \text{CFI} = .81; \text{GFI} = .92; \text{SRMR} = .09 \). The final
considerations included the standardized residuals and MIs. The standardized residuals did not substantially deviate from the +/-1.96 standard previously noted, with residuals ranging from less than .01 to -2.56. Upon review of the MIs, all error covariances suggested were amongst the program implementation observed variables. Based upon the results, the first theoretically justifiable modification warranted was the removal of the attendance observed variable. As the reference variable has yet to be selected, the LISREL 9.1 software default selects a reference indicator for the latent variable.

The modified quantity of program dosage measurement model did not improve the fit of the model, suggesting continued measurement issues with the model. While all parameter estimates were found statistically significant at the $\alpha < .05$ level of significance, the negative error variance of the sports coaching indicator remained. The goodness of fit statistics demonstrated a poor fitting model, with the majority of fit indices indicating a poor model fit [$\chi^2(2) = 120.57, p < .001; NC = 60.29; RMSEA = .38; CFI = .80; GFI = .90; SRMR = .10$]. Upon review of the standardized residuals, the residuals between the curriculum observed variable and S.E.T.S. coaching (2.27) and environment variables (-2.56) were just outside the acceptable standard. Additionally, three of the four MIs suggested incorporated the curriculum observed variable, suggesting a decrease in $\chi^2$ between 35.9 to 89.1. Since removing the attendance observed variable did not improve the fit of the model and the results of the modified model suggest measurement issues with the curriculum observed variable, the attendance indicator was added back into the model and the curriculum indicator was removed.
The modifications to the measurement model (i.e., maintaining attendance and removing curriculum), considerably improved the fit of the model and addressed the negative error variance issue. When reviewing the individual paths of the parameter estimates, the attendance observed variable was once again found statistically non-significant ($p = .35$), accounting for less than 1% of the variance in the latent construct, suggesting the removal of this item. While all other parameter estimates were found statistically significant at the $\alpha < .05$ level of significance and accounted for 40 to 87% of the variance in the latent construct, with the sports coaching observed variable demonstrating a positive error variance. The model was found to be an acceptable fit, with all global fit indices indicating a good fitting model [$\chi^2(2) = 2.28, p = .320$; NC = 1.14; RMSEA = .02; CFI = 1.0; GFI = .99; SRMR = .02], and no MIs suggested. Additionally, all standardized residuals fell within the +/-1.96 standard range for a two-tailed test with an $\alpha < .05$. Although the model fit, standardized residuals, and lack of modification suggestions imply an acceptable model, the attendance observed variable continues to be a poor indicator of quantity of program dosage and therefore was removed.

The final modified measurement model defined quantity of program dosage by three program implementation observed variables, including adherence to program structure (sports coaching and S.E.T.S. coaching) and creation of a caring, mastery-oriented environment. As all indicators reflect program implementation, the latent construct was specified as ‘program implementation’, rather than ‘quantity of program dosage’. Within the modified program implementation measurement model (Figure 6),
sports coaching was fixed to one to act as the reference variable as this indicator was associated with the largest factor loading (.93) and $R^2$ (.87). Results of the CFA reveal a saturated model [$\chi^2(0) = 0.0, p = 1.0$, RMSEA = 0.0], with statistical significant parameter estimates, and no standardized residuals nor modification suggestions to report. Therefore, the modified program implementation measurement model was found theoretically and empirically sound.

Figure 6

*Program implementation measurement model with standardized solutions*

Note. Standardized solutions are reported. The fixed factor is denoted by ‘F’. *** $p < .001$.

With the measurement model defining quantity of program dosage confirmed, the next measurement model tested examined the quality of program dosage observed variables, including belonging, leader support, and engagement (Note: This model mirrors the single factor modified program dosage measurement model). The latent construct was specified as ‘program engagement’ rather than ‘quality of program dosage’ to provide a more clear distinction of the construct. The initial CFA was conducted to
determine which indicator should act as the reference variable, allowing for comparison across the factor loadings. All three indicators demonstrated large factor loadings and $R^2$, however the leader support observed variable was found to be the best indicator of the program engagement latent construct ($R^2 = .64$) and therefore was fixed to one to act as the reference variable. The resulting model (Figure 7) was found to be saturated [$\chi^2(0) = 0.0, p = 1.0, RMSEA = 0.0$], with statistically significant parameter estimates and no standardized residuals nor modification suggestions to report. Therefore the program engagement measurement model was found theoretically and empirically sound.

Figure 7

*Program engagement measurement model with standardized solutions*

![Program Engagement Model](image)

*Note.* Standardized solutions are reported. The fixed factor is denoted by ‘F’. *** $p < .001$.

To test the multiple factor program dosage measurement model, the program implementation and engagement measurement models were combined into a single CFA, for which the latent constructs were allowed to correlate as they both measure an element of program dosage (Figure 8). The multiple factor measurement model was found to be a
good fitting model, with all global fit indices indicating an acceptable model fit [$\chi^2(8) = 11.584, p = .17; NC = 1.45; RMSEA = .03; CFI = .99; GFI = .99; SRMR = .03$]. All parameter estimates were statistically significant at the $\alpha < .05$ level of significance, with large $R^2$ (ranging between .41 to .86). The standardized residuals fell within the accepted range of +/- 1.96 and the only modification suggestion was associated with a nominal decrease in $\chi^2$. Therefore, based on the results, the two-factor program dosage measurement model was found theoretically and empirically sound.
To determine the best program dosage measurement model to incorporate into the structural models, the single and multiple factor models were compared using the $\chi^2$ difference test. As the single factor measurement model was saturated [$\chi^2(0) = 0, p = 1.0$], the $\chi^2$ difference is equal to the $\chi^2$ of the multiple factor measurement model [$\chi^2(8) = 11.584, p = .17$], suggesting that there is not a statistically significant difference between...
the two models. The multiple factor program dosage measurement model was selected for use in all structural models as the model provides a more in-depth examination of program dosage and addresses more of the research questions under investigation.

**Structural Equation Models**

To test the structural models, the measurement and structural components were combined into structural regression models for analysis. As the program dosage latent construct was found to demonstrate a better fit as a two factor model (i.e., program implementation, program engagement), the original structural models were revised to reflect this change. Based upon theoretical consideration, two different models were hypothesized to reflect the interaction between the youth development and program dosage latent constructs (Figure 9). Model A specifies both program dosage factors (i.e., program implementation, program engagement) as mediators of youth development. In comparison, Model B specifies program engagement as a mediator of youth development, while program implementation acts as an antecedent of program engagement.
To test the proposed structural models, correlation and covariance matrices of the observed variables were produced (Tables 10 and 11). In addition to the covariance matrix, an asymptotic covariance matrix was created to address any remaining nonnormality (Note: The matrix is too extensive to report). The matrices reflect the transformed observed variables presented in Table 9. The covariance and asymptotic covariance matrices were inputted into LISREL 9.1 software for SEM analysis. Each proposed structural model (Models A and B), using the two factor program dosage measurement model, were tested and then compared to alternative structural models, using the $\chi^2$ difference test, to determine the best fitting structural models.
### Table 10

**Correlation matrix of observed variables**

<table>
<thead>
<tr>
<th>Observed variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC (Pre)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC (Post)</td>
<td>.53</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort (Pre)</td>
<td>.43</td>
<td>.32</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort (Post)</td>
<td>.39</td>
<td>.53</td>
<td>.53</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork (Pre)</td>
<td>.72</td>
<td>.51</td>
<td>.54</td>
<td>.44</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork (Post)</td>
<td>.47</td>
<td>.75</td>
<td>.39</td>
<td>.67</td>
<td>.55</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Con. (Pre)</td>
<td>.61</td>
<td>.49</td>
<td>.53</td>
<td>.44</td>
<td>.72</td>
<td>.52</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Con. (Post)</td>
<td>.46</td>
<td>.66</td>
<td>.43</td>
<td>.69</td>
<td>.47</td>
<td>.80</td>
<td>.55</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 1 (Pre)</td>
<td>.08</td>
<td>.15</td>
<td>.44</td>
<td>.31</td>
<td>.26</td>
<td>.24</td>
<td>.14</td>
<td>.18</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 2 (Pre)</td>
<td>.15</td>
<td>.16</td>
<td>.44</td>
<td>.35</td>
<td>.29</td>
<td>.28</td>
<td>.20</td>
<td>.20</td>
<td>.69</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 3 (Pre)</td>
<td>.25</td>
<td>.21</td>
<td>.49</td>
<td>.42</td>
<td>.40</td>
<td>.33</td>
<td>.29</td>
<td>.29</td>
<td>.66</td>
<td>.66</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 1 (Post)</td>
<td>.15</td>
<td>.25</td>
<td>.35</td>
<td>.45</td>
<td>.28</td>
<td>.39</td>
<td>.23</td>
<td>.32</td>
<td>.61</td>
<td>.54</td>
<td>.50</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 2 (Post)</td>
<td>.21</td>
<td>.30</td>
<td>.42</td>
<td>.49</td>
<td>.34</td>
<td>.42</td>
<td>.25</td>
<td>.36</td>
<td>.61</td>
<td>.52</td>
<td>.57</td>
<td>.75</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 3 (Post)</td>
<td>.19</td>
<td>.30</td>
<td>.37</td>
<td>.50</td>
<td>.31</td>
<td>.48</td>
<td>.23</td>
<td>.41</td>
<td>.54</td>
<td>.54</td>
<td>.56</td>
<td>.69</td>
<td>.71</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attend. ²</td>
<td>.03</td>
<td>.03</td>
<td>.07</td>
<td>.04</td>
<td>.08</td>
<td>.03</td>
<td>.12</td>
<td>-.01</td>
<td>-.04</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th></th>
<th>Behavioral Engage.</th>
<th>16</th>
<th>.39</th>
<th>.65</th>
<th>.38</th>
<th>.64</th>
<th>.43</th>
<th>.69</th>
<th>.44</th>
<th>.71</th>
<th>.19</th>
<th>.20</th>
<th>.26</th>
<th>.30</th>
<th>.35</th>
<th>.38</th>
<th>.10</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Belonging</td>
<td>17</td>
<td>.31</td>
<td>.57</td>
<td>.31</td>
<td>.46</td>
<td>.36</td>
<td>.62</td>
<td>.36</td>
<td>.52</td>
<td>.23</td>
<td>.25</td>
<td>.25</td>
<td>.25</td>
<td>.28</td>
<td>.36</td>
<td>.05</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td>Leader Support</td>
<td>18</td>
<td>.32</td>
<td>.57</td>
<td>.25</td>
<td>.48</td>
<td>.33</td>
<td>.60</td>
<td>.31</td>
<td>.54</td>
<td>.14</td>
<td>.15</td>
<td>.19</td>
<td>.18</td>
<td>.24</td>
<td>.29</td>
<td>.16</td>
<td>.56</td>
</tr>
<tr>
<td></td>
<td>Sports Coach.³</td>
<td>19</td>
<td>.09</td>
<td>.09</td>
<td>.00</td>
<td>.04</td>
<td>.03</td>
<td>.05</td>
<td>-.00</td>
<td>.06</td>
<td>.02</td>
<td>-.04</td>
<td>-.08</td>
<td>-.01</td>
<td>.01</td>
<td>.05</td>
<td>-.06</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td>S.E.T.S. Coach.</td>
<td>20</td>
<td>.18</td>
<td>.14</td>
<td>.07</td>
<td>.13</td>
<td>.12</td>
<td>.12</td>
<td>.13</td>
<td>.03</td>
<td>.01</td>
<td>-.06</td>
<td>.04</td>
<td>.06</td>
<td>.05</td>
<td>-.07</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>Env.³</td>
<td>21</td>
<td>.11</td>
<td>.09</td>
<td>.00</td>
<td>.04</td>
<td>.04</td>
<td>.03</td>
<td>.05</td>
<td>.05</td>
<td>.00</td>
<td>-.00</td>
<td>-.03</td>
<td>.01</td>
<td>.03</td>
<td>-.00</td>
<td>-.00</td>
<td>.06</td>
</tr>
<tr>
<td></td>
<td>Curr.²</td>
<td>22</td>
<td>-.04</td>
<td>.00</td>
<td>.05</td>
<td>-.05</td>
<td>-.01</td>
<td>-.07</td>
<td>.00</td>
<td>.04</td>
<td>.00</td>
<td>-.04</td>
<td>-.03</td>
<td>-.00</td>
<td>.06</td>
<td>-.05</td>
<td>.02</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note. SC = social competence; Self-Con. = self-control; AC = athletic competence; Attend. = attendance; Behavioral Engage. = behavioral engagement; Sports Coach. = program structure - sports coaching; S.E.T.S. Coach = program structure - S.E.T.S. coaching; Env. = program environment; Curr. = program curriculum.
Table 1

Covariance matrix of observed variables

<table>
<thead>
<tr>
<th>Observed variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC (Pre)</td>
<td>.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC (Post)</td>
<td>.40</td>
<td>.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort (Pre)</td>
<td>.27</td>
<td>.22</td>
<td>.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort (Post)</td>
<td>.25</td>
<td>.39</td>
<td>.32</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork (Pre)</td>
<td>.41</td>
<td>.33</td>
<td>.29</td>
<td>.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teamwork (Post)</td>
<td>.30</td>
<td>.53</td>
<td>.23</td>
<td>.41</td>
<td>.30</td>
<td>.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Con. (Pre)</td>
<td>.32</td>
<td>.29</td>
<td>.26</td>
<td>.22</td>
<td>.32</td>
<td>.26</td>
<td>.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Con. (Post)</td>
<td>.27</td>
<td>.44</td>
<td>.24</td>
<td>.40</td>
<td>.24</td>
<td>.45</td>
<td>.26</td>
<td>.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 1 (Pre)</td>
<td>.05</td>
<td>.11</td>
<td>.27</td>
<td>.20</td>
<td>.15</td>
<td>.15</td>
<td>.07</td>
<td>.10</td>
<td>.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 1 (Post)</td>
<td>.11</td>
<td>.12</td>
<td>.29</td>
<td>.23</td>
<td>.17</td>
<td>.18</td>
<td>.11</td>
<td>.12</td>
<td>.47</td>
<td>.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 2 (Pre)</td>
<td>.18</td>
<td>.17</td>
<td>.33</td>
<td>.29</td>
<td>.24</td>
<td>.23</td>
<td>.16</td>
<td>.19</td>
<td>.46</td>
<td>.49</td>
<td>.76</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 2 (Post)</td>
<td>.10</td>
<td>.19</td>
<td>.22</td>
<td>.30</td>
<td>.16</td>
<td>.25</td>
<td>.12</td>
<td>.19</td>
<td>.40</td>
<td>.37</td>
<td>.36</td>
<td>.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 3 (Pre)</td>
<td>.14</td>
<td>.23</td>
<td>.27</td>
<td>.33</td>
<td>.20</td>
<td>.28</td>
<td>.14</td>
<td>.22</td>
<td>.42</td>
<td>.44</td>
<td>.42</td>
<td>.52</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AC Item 3 (Post)</td>
<td>.14</td>
<td>.24</td>
<td>.25</td>
<td>.35</td>
<td>.19</td>
<td>.33</td>
<td>.13</td>
<td>.27</td>
<td>.38</td>
<td>.41</td>
<td>.43</td>
<td>.50</td>
<td>.53</td>
<td>.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attend.2</td>
<td>.08</td>
<td>.09</td>
<td>.16</td>
<td>.18</td>
<td>.08</td>
<td>.19</td>
<td>.06</td>
<td>.28</td>
<td>- .03</td>
<td>- .10</td>
<td>.04</td>
<td>.03</td>
<td>.06</td>
<td>.03</td>
<td></td>
<td>9.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th></th>
<th>Behavioral Engage.</th>
<th>SC</th>
<th>.19</th>
<th>.35</th>
<th>.17</th>
<th>.30</th>
<th>.18</th>
<th>.32</th>
<th>.17</th>
<th>.31</th>
<th>.09</th>
<th>.10</th>
<th>.13</th>
<th>.15</th>
<th>.18</th>
<th>.20</th>
<th>.19</th>
<th>.35</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td></td>
<td></td>
<td>.23</td>
<td>.48</td>
<td>.22</td>
<td>.34</td>
<td>.23</td>
<td>.45</td>
<td>.22</td>
<td>.35</td>
<td>.17</td>
<td>.20</td>
<td>.18</td>
<td>.19</td>
<td>.22</td>
<td>.30</td>
<td>.16</td>
<td>.30</td>
</tr>
<tr>
<td>17</td>
<td>Belonging</td>
<td></td>
<td>.24</td>
<td>.48</td>
<td>.17</td>
<td>.35</td>
<td>.21</td>
<td>.43</td>
<td>.18</td>
<td>.37</td>
<td>.11</td>
<td>.12</td>
<td>.15</td>
<td>.14</td>
<td>.19</td>
<td>.24</td>
<td>.45</td>
<td>.31</td>
</tr>
<tr>
<td>18</td>
<td>Leader Support</td>
<td></td>
<td>.04</td>
<td>.05</td>
<td>.00</td>
<td>.02</td>
<td>.01</td>
<td>.02</td>
<td>-.00</td>
<td>.03</td>
<td>.01</td>
<td>-.02</td>
<td>-.04</td>
<td>-.01</td>
<td>.01</td>
<td>.03</td>
<td>-.11</td>
<td>.02</td>
</tr>
<tr>
<td>19</td>
<td>Sports Coach. ¹</td>
<td></td>
<td>.08</td>
<td>.07</td>
<td>.03</td>
<td>.06</td>
<td>.04</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
<td>.01</td>
<td>.00</td>
<td>-.03</td>
<td>.02</td>
<td>.03</td>
<td>.02</td>
<td>-.11</td>
<td>.04</td>
</tr>
<tr>
<td>20</td>
<td>S.E.T.S. Coach.</td>
<td></td>
<td>.06</td>
<td>.06</td>
<td>.00</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.03</td>
<td>.00</td>
<td>.00</td>
<td>-.02</td>
<td>.00</td>
<td>.02</td>
<td>-.00</td>
<td>-.01</td>
<td>.02</td>
</tr>
<tr>
<td>21</td>
<td>Env. ³</td>
<td></td>
<td>-.02</td>
<td>.00</td>
<td>.02</td>
<td>-.02</td>
<td>-.01</td>
<td>-.03</td>
<td>.00</td>
<td>.02</td>
<td>.00</td>
<td>-.02</td>
<td>-.02</td>
<td>-.00</td>
<td>.03</td>
<td>-.09</td>
<td>.01</td>
<td>.03</td>
</tr>
</tbody>
</table>

*Note. SC = social competence; Self-Con. = self-control; AC = athletic competence; Attend. = attendance; Behavioral Engage. = behavioral engagement; Sports Coach. = program structure - sports coaching; S.E.T.S. Coach = program structure - S.E.T.S. coaching; Env. = program environment; Curr. = program curriculum.*
Social Competence Model

To test the social competence structural models, the measurement and structural components were combined into two proposed models (i.e., Model A and B). Model A specifies both program implementation and program engagement as mediators of change in social competence. Conversely, Model B specifies program implementation as an antecedent of program engagement, with program engagement acting as a mediator of change in social competence. For each model, the proposed structural models were tested, along with theoretically justifiable alternative models, to determine the best fitting model. A final model was selected from all models tested based upon theoretical justification and empirical support.

Model A. The proposed structural model (i.e., Model A1) specifies a mediating relationship between program implementation and change in social competence, as well as between program engagement and change in social competence. In addition to the proposed model, seven theoretically justifiable alternative models were tested to determine the best fitting model. Figure 10 illustrates the eight structural models tested.
Figure 10

*Path diagrams of social competence Model A structural models*

**Model A1**

- Soc. Comp. (PRE) ➔ Soc. Comp. (POST)
- Program Implement. ➔ Program Engage.

**Model A2**

- Soc. Comp. (PRE) ➔ Soc. Comp. (POST)
- Program Implement. ➔ Program Engage.

**Model A3**

- Soc. Comp. (PRE) ➔ Soc. Comp. (POST)
- Program Implement. ➔ Program Engage.

**Model A4**

- Soc. Comp. (PRE) ➔ Soc. Comp. (POST)
- Program Implement. ➔ Program Engage.

**Model A5**

- Soc. Comp. (PRE) ➔ Soc. Comp. (POST)
- Program Implement. ➔ Program Engage.

**Model A6**

- Soc. Comp. (PRE) ➔ Soc. Comp. (POST)
- Program Implement. ➔ Program Engage.

**Model A7**

- Soc. Comp. (PRE) ➔ Soc. Comp. (POST)
- Program Implement. ➔ Program Engage.

**Model A8**

- Soc. Comp. (PRE) ➔ Soc. Comp. (POST)
- Program Implement. ➔ Program Engage.

*Note.* Soc. Comp. = social competence; Program Implement. = program implementation; and Program Engage. = program engagement.
The proposed structural model (Model A1) was inputted into LISREL 9.1 software for analysis. The errors of corresponding observed variables at the pre-test and post-test (i.e., social competence, effort, teamwork, self-control) were permitted to correlate as the same scale was used at both time periods. To allow for comparison across factor loadings, the reference variables [i.e., teamwork (pre-test and post-test); adherence to program structure – sports coaching; leader support], determined when testing the measurement models, were once again fixed to one. An additional parameter specification was the inclusion of an error covariance between effort and self-control measured at the post-test, based upon the measurement model previously tested. A path between the social competence latent constructs was specified to account for the social competence pre-test scores as an antecedent of the post-test scores. Additionally, paths from social competence (pre-test) to program implementation and engagement, as well as paths from program implementation and engagement to social competence (post-test) were included to specify the mediating influence of program implementation and program engagement.

Once all parameters were specified, the initial proposed model (Model A1) was tested first (Table 12). When evaluating the model, the statistical significance of the $t$-values were first evaluated at the $\alpha < .05$ level of significance. The parameter estimates of all observed variables were found significant ($p < .001$), supporting the measurement models previously established. Upon examination of the structural equations, program implementation was found to be a non-significant mediator. The paths between social competence (pre-test) and program implementation ($p = .267$), as well as program
implementation and social competence (post-test; \( p = .650 \)), did not uphold the \( \alpha < .05 \) level of significance.

Table 12

*Structural equation modeling goodness of fit indices for social competence Model A*

<table>
<thead>
<tr>
<th>Model Tested</th>
<th>( \chi^2 )</th>
<th>df</th>
<th>( p )</th>
<th>NC</th>
<th>RMSEA</th>
<th>CFI</th>
<th>GFI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A1</td>
<td>156.87</td>
<td>66</td>
<td>&lt; .001</td>
<td>2.38</td>
<td>0.06</td>
<td>0.99</td>
<td>0.95</td>
<td>0.04</td>
</tr>
<tr>
<td>Model A2</td>
<td>158.15</td>
<td>67</td>
<td>&lt; .001</td>
<td>2.36</td>
<td>0.06</td>
<td>0.99</td>
<td>0.95</td>
<td>0.05</td>
</tr>
<tr>
<td>(SC_Pre → Implement. removed)</td>
<td>158.30</td>
<td>68</td>
<td>&lt; .001</td>
<td>2.33</td>
<td>0.06</td>
<td>0.99</td>
<td>0.95</td>
<td>0.05</td>
</tr>
<tr>
<td>Model A3</td>
<td>274.40</td>
<td>67</td>
<td>&lt; .001</td>
<td>4.10</td>
<td>0.09</td>
<td>0.98</td>
<td>0.92</td>
<td>0.18</td>
</tr>
<tr>
<td>(SC_Pre → Engage. removed)</td>
<td>575.11</td>
<td>68</td>
<td>&lt; .001</td>
<td>8.46</td>
<td>0.13</td>
<td>0.94</td>
<td>0.88</td>
<td>0.24</td>
</tr>
<tr>
<td>Model A4</td>
<td>575.11</td>
<td>68</td>
<td>&lt; .001</td>
<td>8.46</td>
<td>0.13</td>
<td>0.94</td>
<td>0.88</td>
<td>0.24</td>
</tr>
<tr>
<td>Model A5</td>
<td>575.11</td>
<td>68</td>
<td>&lt; .001</td>
<td>8.46</td>
<td>0.13</td>
<td>0.94</td>
<td>0.88</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Continued
Table 12 continued

<table>
<thead>
<tr>
<th>Model</th>
<th>274.46</th>
<th>68</th>
<th>&lt; .001</th>
<th>4.04</th>
<th>0.09</th>
<th>0.98</th>
<th>0.92</th>
<th>0.18</th>
</tr>
</thead>
<tbody>
<tr>
<td>A6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(SC_Pre → Engage. removed; SC_Pre → Implement. removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>436.16</th>
<th>68</th>
<th>&lt; .001</th>
<th>6.41</th>
<th>0.11</th>
<th>0.96</th>
<th>0.89</th>
<th>0.11</th>
</tr>
</thead>
<tbody>
<tr>
<td>A7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Engage. → SC_Post removed; Implement. → SC_Post removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>576.68</th>
<th>70</th>
<th>&lt; .001</th>
<th>8.24</th>
<th>0.13</th>
<th>0.94</th>
<th>0.88</th>
<th>0.24</th>
</tr>
</thead>
<tbody>
<tr>
<td>A8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(SC_Pre → Engage. removed; SC_Pre → Implement. removed; Engage. → SC_Post removed; Implement. → SC_Post removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** SC_Pre = social competence (pre-test); SC_Post = social competence (post-test); Implement. = program implementation; Engage. = program engagement; \( \rightarrow \) = path; \( \chi^2 \) = chi-square; df = degrees of freedom; NC = normed chi-square; RMSEA = root mean square error of approximation; CFI = comparative fit index; GFI = goodness-of-fit index; SRMR = standardized root-mean square residual.

In relation to the fit of the model (Table 12), five of the six global fit indices indicated a good fitting model. Although the model was an acceptable fit, the standardized residuals, MIs, and EPC statistics were reviewed to determine any
theoretically justifiable modifications. The standardized residuals were primarily within the \( \pm 1.96 \) threshold established. However, the largest standardized residual (4.19) was between social competence (pre-test) and adherence to program structure (S.E.T.S. coaching) observed variables, supporting program implementation as a poor outcome of social competence pre-test scores. The theoretically justifiable MIs, suggested to result in a moderate decrease in \( \chi^2 \), included the addition of an error covariance between leader support and belonging (32.8 decrease in \( \chi^2 \)) and between behavioral engagement and belonging (25.6 decrease in \( \chi^2 \)). As program implementation was found to be a non-significant mediator, the path between social competence (pre-test) and program implementation was removed in Model A2 to determine the best fitting model.

The alternative model (Model A2) was inputted, for which the path between social competence (pre-test) and program implementation was removed, and SEM conducted. Once again all observed \( t \)-values were significant at the \( \alpha < .05 \) level of significance. The path between program implementation and social competence (post-test) remained non-significant \( (p = .664) \), suggesting that program implementation is a poor antecedent of social competence at the post-test. All other paths between the structural components were found statistically significant at the \( \alpha < .05 \) level of significance. When examining the fit of the model (Table 12), five of the six global fit indices suggested an acceptable model fit. The largest standardized residual was once again between social competence (pre-test) and adherence to program structure (S.E.T.S. coaching; 4.04). With regards to the modification suggestions, the proposed error covariances between leader support and belonging (32.8 decrease in \( \chi^2 \)) and between
behavioral engagement and belonging (25.6 decrease in $\chi^2$) remained considerable. When comparing Model A1 and A2, a $\chi^2$ difference test was conducted, resulting in a non-significant $\chi^2$ ($p = .258$), suggesting equally good fitting models.

The next modification to test was the removal of program implementation as a mediating latent construct (Model A3). Specifically, the paths between social competence (pre-test) and program implementation, as well as between program implementation and social competence (post-test) were removed. When examining the factor loadings and structural equations, all parameter estimates were found significant at the $\alpha < .05$ level of significance. Examination of the global fit indices (Table 12) revealed an acceptable model fit, with five of the six indices indicating a good fitting model. The largest standardized residual remained between social competence (pre-test) and adherence to program structure (S.E.T.S. coaching; 4.04). Once again, the MIs suggested the addition of an error covariance between leader support and belonging (32.6 decrease in $\chi^2$) and between behavioral engagement and belonging (23.6 decrease in $\chi^2$). When comparing Model A1 with Model A3, the $\chi^2$ difference was non-significant ($p = .489$), suggesting equally good fitting models.

The fourth model tested (Model A4) is a theoretically justifiable alternative model, for which program implementation is maintained as a mediator of change in social competence, while the path between social competence (pre-test) and program engagement removed. Therefore, Model A4 is testing program engagement as an antecedent of social competence at the post-test. SEM of the alternative model was conducted, followed by an evaluation of the model. The observed $t$-values were found to
be statistically significant at the $\alpha < .05$ level of significance, while the paths between social competence (pre-test) and program implementation ($p = .788$), as well as program implementation and social competence (post-test; $p = .661$) were found non-significant.

In regards to the fit of Model A4 (Table 12), four of the six global fit indices indicated an acceptable model fit. When reviewing the standardized residuals, the largest residuals were found between the social competence (pre-test) observed variables and the program engagement observed variables (ranging from 4.74 to 7.162), contradicting the modification removing the path between social competence (pre-test) and program engagement. The modification suggestion with the greatest anticipated decrease in $\chi^2$ was to add the path from social competence (pre-test) to program engagement back into the model (102.2 decrease in $\chi^2$). The $\chi^2$ difference test revealed a significant difference between Model A1 and A4 ($p < .001$), with Model A1 demonstrating greater fit.

An additional alternative model considered (Model A5) removed program engagement as a mediator; therefore eliminating the paths between social competence (pre-test) and program engagement and between program engagement and social competence (post-test). Results of SEM analysis indicate significant observed $t$-values and a significant path between social competence pre-test and post-test. Program implementation was found to be a non-significant mediator of change in social competence ($p = .211; p = .927$). The global fit indices suggested a poor fitting model (Table 12), with five of the six indices indicating a poor model fit. The largest standardized residuals were between the social competence observed variables (pre-test and post-test) and the program engagement observed variables (ranging from 4.74 to
10.68), contradicting the removal of program engagement as a mediator. Upon consideration of theoretically justifiable modification suggestions, the MI’s with the greatest anticipated decrease in $\chi^2$ suggested the paths from social competence (pre-test) to program engagement (110.9 decrease in $\chi^2$) and program engagement to social competence (post-test; 157.6 decrease in $\chi^2$) be added back into the model. When comparing Model A1 and Model A5, the $\chi^2$ difference was significant ($p < .001$), for which Model A1 demonstrated better model fit.

The remaining three alternative models examine the program implementation and engagement latent constructs as outcome variables only, antecedent variables only, or were completely removed from the model. Model A6 specifies program implementation and engagement as antecedents of social competence at the post-test; the paths between social competence (pre-test) and program implementation as well as engagement were removed. The resulting analysis revealed statistically significant observed $t$-values, with only social competence (pre-test; $p < .001$) and program engagement ($p < .001$) latent constructs significantly predicting social competence post-test scores. The path between program implementation and social competence (post-test) was found non-significant ($p = .636$). In relation to the fit of the model (Table 12), four of the six fit indices indicated a good fitting model. When reviewing the standardized residuals, the largest residuals were between the social competence (pre-test) observed variables and the program engagement observed variables (ranging from 4.744 to 7.162). Similar to Models A4 and A6, the primary MI suggested that the path from social competence (pre-test) to program engagement be added back in (100.7 decrease in $\chi^2$).
Models A1 and A6 resulted in a significant $\chi^2 (p < .001)$, for which Model A1 was found to be a better fitting model.

Model A7 tested program implementation and engagement as outcomes of social competence pre-test scores. The paths between program implementation and social competence (post-test) and between program engagement and social competence (post-test) were removed from the model. SEM analysis was conducted to test the model fit. All observed t-values were found significant at the $\alpha < .05$ level of significance. In regards to the structural equations, social competence (post-test; $p < .001$) and program engagement ($p < .001$) were significant outcome variables of social competence (pre-test), while program implementation was found to be a non-significant outcome variable ($p = .241$). Model A7 demonstrated a poor fit (Table 12), with five of the six global fit indices suggesting a poor fitting model. The largest standardized residuals were found between the social competence (post) and program engagement observed variables (ranging 4.01 to 7.42), contradicting the removal of program engagement as an antecedent of social competence at the post-test. Lastly, the theoretically justifiable MIs suggested to considerably decrease $\chi^2$ included the addition of a path from program engagement to social competence (post-test; 304.4 decrease in $\chi^2$), and the addition of an error covariance between leader support and belonging (267.6 decrease in $\chi^2$). When comparing Model A7 to A1, the $\chi^2$ difference test demonstrated a significant change, suggesting that the alternative Model A7 significantly deteriorated from the proposed Model A1.
The final alternative model tested (Model A8) removed both program implementation and engagement as mediators of change in social competence. The only relationship examined was between the social competence scores at the pre-test and post-test. All observed variables and the structural equation was significant at the $\alpha < .05$ level of significance. The fit of the model was not acceptable (Table 12), with five of the six global fit indices indicating a poor fitting model. The largest standardized residuals were between program engagement and social competence (post-test) observed variables (ranging from 7.59 to 10.72), with large residuals between program engagement and social competence (pre-test) observed variables as well. These results contradict the removal of program engagement as a mediator of change in social competence. Upon review of the modification suggestions, the primary modification to consider is the addition of paths from social competence (pre-test) to program engagement and then to social competence (post-test), which would result in an anticipated decrease in $\chi^2$ by 109.7 and 157.2 respectively. The $\chi^2$ difference test resulted in a significant difference between Model A8 and Model A1 ($p < .001$), with Model A1 demonstrating a better model fit.

In summary, Models A1, A2, A3, A4, and A6 were found to have an acceptable model fit, with the majority of global fit indices suggesting a good fitting model. Comparatively, Models A5, A7, and A8 demonstrated a poor model fit based upon the majority of global fit indices failing to uphold the standards previously noted. Of the five models with an acceptable fit, Models A2 and A3 were not significantly different from the proposed Model A1, while Models A4 and A6 were significantly different from the
proposed model. The standardized residuals and MIs support the inclusion of program engagement as a mediator of change in social competence, however there were mixed results regarding program implementation as a mediator of change in social competence. When including program implementation as a mediator of change in social competence (Model A1), the program implementation paths were found non-significant at the $\alpha < .05$ level of significance, however the global fit indices indicated a good fitting model. In Model A1, the structural model demonstrated that 91.5% of the variability in social competence post-test scores are accounted for. In comparison, when removing program implementation as a mediator (Model A3), all paths in the structural equation are significant and 91.4% of the variability in social competence post-test scores is explained. Based upon the law of parsimony and the minimal change in the social competence (post-test) $R^2$, Model A3 was determined to be the best model that is both theoretically and empirically sound (Kline, 2005).

**Model B.** The proposed structural Model B1 maintains a mediating relationship between program engagement and change in social competence, while specifying program implementation as an antecedent of program engagement. Alternative models were considered, for which four theoretically justifiable alternative models were developed. Figure 11 illustrates the five structural models tested.
Figure 11

Path diagrams of social competence Model B structural models

Note. Soc. Comp. = social competence; Program Implement. = program implementation; and Program Engage. = program engagement.
Similar to Model A1, the proposed structural model (Model B1) was inputted into LISREL 9.1 software for SEM analysis. As the longitudinal observed variables (i.e., social competence, effort, teamwork, self-control) used the same scale at two separate time periods, the errors of the corresponding observed variables were permitted to correlate. The reference variables for each measurement model [i.e., teamwork; adherence to program structure – sports coaching; leader support], determined during measurement model testing, were once again fixed to one to allow for comparison across the factor loadings. Additionally, the errors associated with the effort and self-control observed variables at the pre-test were allowed to correlate, based upon the established measurement model. In relation to the structural model, a path between the social competence latent variables was specified to account for change in social competence. Furthermore, a path from social competence (pre-test) to program engagement and then to social competence (post-test) was specified to test the mediating effect of program engagement. Lastly, a path from program implementation to program engagement was specified to account for program implementation as an antecedent of engagement.

Once all parameter estimates were specified, SEM analysis was conducted to test Model B1 (Table 13). To evaluate the model, the observed $t$-values were reviewed, based on the $\alpha < .05$ level of significance, for which all observed variables were found significant. Upon review of the structural equations, program engagement was a significant mediator of change in social competence ($p < .001$), with the path between social competence latent constructs significant as well ($p = .004$). Although, the path from program implementation to engagement was found non-significant ($p = .154$).
Table 13

*Structural equation modeling goodness of fit indices for social competence Model B*

<table>
<thead>
<tr>
<th>Model Tested</th>
<th>$\chi^2$</th>
<th>Df</th>
<th>p</th>
<th>NC</th>
<th>RMSEA</th>
<th>CFI</th>
<th>GFI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model B1</td>
<td>157.04</td>
<td>67</td>
<td>&lt; .001</td>
<td>2.34</td>
<td>0.06</td>
<td>0.99</td>
<td>0.95</td>
<td>0.04</td>
</tr>
<tr>
<td>Model B2</td>
<td>273.14</td>
<td>68</td>
<td>&lt; .001</td>
<td>4.02</td>
<td>0.09</td>
<td>0.98</td>
<td>0.92</td>
<td>0.18</td>
</tr>
<tr>
<td>(SC_Pre $\rightarrow$ Engage. removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model B3</td>
<td>436.16</td>
<td>68</td>
<td>&lt; .001</td>
<td>6.41</td>
<td>0.11</td>
<td>0.96</td>
<td>0.89</td>
<td>0.11</td>
</tr>
<tr>
<td>(Engage. $\rightarrow$ SC_Post removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model B4</td>
<td>157.04</td>
<td>67</td>
<td>&lt; .001</td>
<td>2.34</td>
<td>0.06</td>
<td>0.99</td>
<td>0.95</td>
<td>0.04</td>
</tr>
<tr>
<td>(SC_Pre $\rightarrow$ Engage. removed; SC_Pre $\rightarrow$ Implement. added)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model B5</td>
<td>436.16</td>
<td>68</td>
<td>&lt; .001</td>
<td>6.41</td>
<td>0.11</td>
<td>0.96</td>
<td>0.89</td>
<td>0.11</td>
</tr>
<tr>
<td>(SC_Pre $\rightarrow$ Engage. removed; SC_Pre $\rightarrow$ Implement. added; Engage. $\rightarrow$ SC_Post removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* SC_Pre = social competence (pre-test); SC_Post = social competence (post-test); Implement. = program implementation; Engage. = program engagement; $\rightarrow$ = path; $\chi^2$ = chi-square; df = degrees of freedom; NC = normed chi-square; RMSEA = root mean square error of approximation; CFI = comparative fit index; GFI = goodness-of-fit index; SRMR = standardized root-mean square residual.
In relation to the fit of the model (Table 13), five of the six global fit indices indicated a good fitting model. When examining the standardized residuals, the values were close to the +/-1.96 threshold, with large standardized residuals between effort and social competence (post-test; -3.01) and between adherence to program structure – S.E.T.S. coaching and social competence (pre-test; 4.189). The theoretically justifiable MIs were moderate, with the largest anticipated decrease in $\chi^2$ associated with the addition of an error covariance between leader support and belonging (32.6 decrease in $\chi^2$) as well as behavioral engagement and belonging (23.7 decrease in $\chi^2$).

Once the proposed model was tested, the four alternative models were run and then compared to the proposed model (Model B1) to determine the best fitting model. The first alternative model tested (Model B2) did not include a path from social competence (pre-test) to program engagement. Therefore, program engagement acted as an antecedent of social competence scores at the post-test and an outcome of program implementation, but was not a mediator in Model B2. SEM analysis was conducted, upon which all observed t-values demonstrated significance at the $\alpha < .05$ level of significance. The structural models also showed significant paths, however the path from program implementation to program engagement was barely non-significant ($p = .050$). In relation to the goodness of fit (Table 13), four of the six global fit indices suggested an acceptable fitting model. Large standardized residuals were found between the social competence (pre-test) and program engagement observed variables (ranging from 4.69 to 7.11), contradicting the removal of the path from social competence (pre-test) to program engagement. The predominant modification suggestion is to add the path from social
competence (pre-test) to program engagement back into the model (100.5 decrease in $\chi^2$).

When comparing Model B1 to B2, the $\chi^2$ difference test demonstrated a significant
difference between the models, for which Model B1 was a better fitting model.

The next alternative model tested (Model B3) removed the path from program
engagement to social competence (post-test), therefore program engagement only acted
as an outcome of social competence at the pre-test and program implementation. All
observed variables were associated with a significant t-value, at the $\alpha < .05$ level of
significance. However, when examining the structural equations, all paths were found
statistically significant except for the path from program implementation to program
engagement ($p = .108$). In relation to model fit (Table 13), only one of the six global fit
indices indicated a good fitting model, which does not meet the standard previously
noted. The standardized residuals were reviewed, for which the largest standardized
residuals were found between the program engagement observed variables and social
competence (post-test) observed variables (ranging from 4.01 to 7.424), contradicting the
removal of the path between program engagement and social competence (post-test). The
theoretically justifiable MIs suggested to add, which would result in a considerable
anticipated decrease in $\chi^2$, included the inclusion of a path from program engagement to
social competence (post-test; 304.4 decrease in $\chi^2$) as well as the addition of an error
covariance between leader support and belonging (267.6 decrease in $\chi^2$). In comparison to
Model B1, Model B3 was found significantly different ($p < .001$), based upon the $\chi^2$
different test, for which Model B1 is a better fitting model.
Model B4 represents another alternative model, in which the path from social competence (pre-test) was modified to lead to program implementation rather than program engagement. Therefore, the path from social competence (pre-test) to program engagement was removed and a path from social competence (pre-test) to program implementation was added. The model was run and the results indicated statistically significant observed t-values, at the α < .05 level of significance. When examining the structural model, the paths from program engagement and social competence (pre-test) to social competence (post-test) were significant ($p < .001; p = .004$ respectively). Comparatively, the paths from social competence (pre-test) to program implementation ($p = .269$) and program implementation to program engagement ($p = .267$) were non-significant, contradicting the modification applied to this model. Model B4 was found to demonstrate an acceptable model fit (Table 13), with five of the six global fit indices suggesting a good fitting model. The standardized residuals did not substantially deviate from the acceptable threshold of +/- 1.96. The last consideration are the MIs, for which the only theoretically justifiable suggestion was to add error covariances between leader support and belonging (32.6 decrease in $\chi^2$) as well as between behavioral engagement and belonging (23.7 decrease in $\chi^2$). As the global fit indices of Model B4 mirrored that of Model B1, Model B4 is not significantly different from Model B1.

The final alternative model tested (Model B5), modified the program engagement latent construct, removing the mediation between the social competence constructs, while maintaining program engagement as an outcome of program implementation. Therefore, the paths from social competence (pre-test) to program engagement as well as program
engagement to social competence (post-test) were removed and a path from social competence (pre-test) to program implementation added. The results indicate statistically significant observed t-values, at the \( \alpha < .05 \) level of significance, with only one structural path significant (\( p < .001 \)) between the social competence constructs. The paths from social competence (pre-test) to program implementation (\( p = .241 \)) as well as from program implementation to program engagement (\( p = .240 \)) were found non-significant. Upon review of the global fit indices (Table 13), five of the six indices indicated a poor fitting model. The largest standardized residuals are between the program engagement observed variables and social competence (post-test) observed variables (ranging from 4.01 to 7.424), contradicting the removal of the path from program engagement to social competence (post-test). The final consideration consists of the theoretically justifiable MIs. The modification suggestions with the largest anticipated decrease in \( \chi^2 \) included the addition of a path from program engagement to social competence (post-test; 304.5 decrease in \( \chi^2 \)) and error covariance between leader support and belonging (276.1 decrease in \( \chi^2 \)). Results of the \( \chi^2 \) difference test found a statistically significant difference between Models B1 and B5, with Model B1 a better fitting model.

When selecting the best model, all five models were compared for goodness of fit, of which Models B1, B2, and B4 demonstrated acceptable fit. Of the three good fitting models, Model B2 was found significantly different from Model B1 based on a \( \chi^2 \) difference test. When comparing the models, Model B1 accounted for a greater amount of the variance in social competence post-test scores (Model B1 \( R^2 = .92 \); Model B2 \( R^2 = .88 \)), demonstrated better fit (Table 13), and was associated with nominal modification
suggestions; therefore Model B1 was superior to Model B2. In respect to Model B4, the global fit indices were equivalent as were the amount of variance in social competence scores (post-test) explained (Models B1 and B4 $R^2 = .92$). Similar to Model B1, Model B4 was associated with minimal standardized residuals and modification suggestions. The primary difference between Models B1 and B4 was the path modification - Model B1 specified a path from social competence (pre-test) to program engagement, while Model B4 specified a path from social competence (pre-test) to program implementation. The results found the path in Model B1 to be significant, while the path in Model B4 was found non-significant. Based upon this evidence, Model B1 is considered the best model, that is both theoretically and empirically sound (Kline, 2005).

**Final model.** To examine the impact of program implementation and engagement on social competence development, 13 structural models were proposed and tested. These models were classified into two primary models – Model A and Model B. Model A specified both program implementation and engagement as mediators of change in social competence, while Model B specified program engagement as a mediator of change in social competence and program implementation as an antecedent of program engagement. Of the 13 models, two models were determined to be superior models, including Model A3 and Model B1. Model A3 is an alternative model in which the program implementation latent construct was removed from the model, therefore only program engagement acts as a mediator of change in social competence. Model A3 was found to demonstrate a good fit (Table 12), with all observed $t$-values and structural paths statistically significant. The structural model was found to account for 91% of the
variance in social competence scores at the post-test. Additionally, the standardized residuals and theoretically justifiable modification suggestions were nominal.

When considering Model B1, program engagement was specified as a mediator of change in social competence and program implementation an antecedent of program engagement. Model B1 was also found to be a good fitting model (Table 13), for which all observed $t$-values were found statistically significant, along with the majority of structural equations. However, the path from program implementation to program engagement was found statistically non-significant ($p = .154$). Despite the non-significant path, the model demonstrated acceptable fit and the standardized residuals as well as modification suggestions were all minimal. Furthermore, the structural model explained 92% of the variance in social competence post-test scores.

Comparing Models A3 and B1, the only difference is the path from program implementation to program engagement found in Model B1. Using a $\chi^2$ difference test, the two models were found to demonstrate a non-significant difference in model fit ($p = .262$), indicating that the removal of the path from program implementation to program engagement did not significantly improve the model. Although the law of parsimony would suggest that the model with the least number of parameters be selected, Model B1 was maintained as the final social competence model for the current study (Figure 13; Kline, 2005).
Figure 12

Social competence Model B1 with standardized solutions

![Diagram showing the relationships between social competence, effort, teamwork, self-control, program engagement, behavioral engagement, sense of belonging, leader support, sports coaching, social skills coaching, and environment with standardized solutions.](image-url)
The purpose of the current study was to examine how program dosage influences PYD, for which program dosage consists of both program implementation and engagement. If the program implementation latent construct was removed from the model merely due to statistical non-significance when the model was found to demonstrate an acceptable fit, a theoretically justifiable and meaningful relationship would be lost. Furthermore, experiential learning (Kolb, 1984; Kolb et al., 2001), cognitive behavioral theory (Beck et al., 1979; Hupp et al., 2012), the Sport Education Model (Siedentop, 1998), and the TPSR model (Hellison, 2003) are theoretical conceptualizations of how learning and development occur. Based upon these theoretical considerations, the program implementation latent construct was maintained to explain this critical mechanism of PYD. Ultimately, in order to fully answer the research questions of the current study and further the exploration of organizational effectiveness within the context of sport-based PYD, Model B1 was selected as the final social competence model.

**Athletic Competence Model**

To investigate the athletic competence structural models, the measurement and structural components previously established were combined into two proposed models (i.e., Model A and B), similar to the social competence models. Model A specifies program implementation and program engagement as mediators of change in athletic competence, while Model B specifies program engagement as a mediator of change in athletic competence and program implementation as an antecedent of program engagement. To determine the best fitting athletic competence model, the proposed
structural models were tested, along with theoretically justifiable alternative models. Of the proposed and alternative models tested, a superior model was selected for both Models A and B. The two superior models were then compared, for which a final athletic competence model was selected based upon empirical support and theoretical justification.

**Model A.** The proposed structural model (i.e., Model A1) specifies program implementation and program engagement as mediators of change in athletic competence. To determine the best fitting model, seven additional theoretically justifiable alternative models were examined. Figure 13 illustrates the proposed and alternative structural athletic competence models tested.
Figure 13

Path diagrams of athletic competence Model A structural models

Note. Ath. Comp. = athletic competence; Program Implement. = program implementation; and Program Engage. = program engagement.
The proposed model (i.e., Model A1) was first tested, for which the parameters and constraints were inputted into LISREL 9.1 software for analysis. To account for the mediating relationships, paths from social competence (pre-test) to program implementation and engagement, as well as paths from program implementation and engagement to social competence (post-test) were specified. An additional path between the athletic competence latent constructs was specified to test whether the pre-test scores act as a predictor of the post-test scores. The reference variables determined when testing the measurement models [i.e., teamwork (pre-test and post-test); adherence to program structure – sports coaching; leader support] were fixed to one to allow for comparison across the factor loadings. As the athletic competence observed variables at the pre-test and post-test correspond, the errors between the corresponding variables (i.e., item 1, item 2, item 3) were permitted to correlate. The final constraint specified was the inclusion of an error covariance between effort and self-control measured at the post-test, based upon the athletic competence (post-test) measurement model previously established.

Once all parameters were specified, SEM was conducted to test the initial proposed (i.e., Model A1). To evaluate the model, the observed $t$-values were compared to the critical $t$-value of a two-tailed test using an $\alpha < .05$ level of significance. All observed variables were found statistically significant ($p < .05$), supporting the measurement models previously established. Upon review of the structural equations, program implementation was found to be a non-significant mediator of change in athletic competence, while results demonstrate that program engagement was a significant
mediator of change in athletic competence at the \( \alpha < .05 \) level of significance. The goodness of fit statistics suggest a good fitting model with five of the six global fit indices indicating an acceptable fit (Table 14). Lastly, when considering any model modifications, the standardized residuals, MIs, and EPC statistics were reviewed. The standardized residuals ranged from -2.68 to 3.01, which are relatively small and not far outside the +/-1.96 standard previously noted. There were no theoretically justifiable MIs suggested to result in a considerable decrease in \( \chi^2 \), therefore the model was found empirically and theoretically sound.

Table 14

Structural equation modeling goodness of fit indices for athletic competence Model A

<table>
<thead>
<tr>
<th>Model Tested</th>
<th>( \chi^2 )</th>
<th>( Df )</th>
<th>( p )</th>
<th>NC</th>
<th>RMSEA</th>
<th>CFI</th>
<th>GFI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model A1</td>
<td>81.64</td>
<td>45</td>
<td>&lt; .001</td>
<td>1.81</td>
<td>0.04</td>
<td>0.99</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>Model A2</td>
<td>81.94</td>
<td>46</td>
<td>&lt; .001</td>
<td>1.78</td>
<td>0.04</td>
<td>0.99</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>(AC_Pre → Implement. removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A3</td>
<td>82.27</td>
<td>47</td>
<td>0.001</td>
<td>1.75</td>
<td>0.04</td>
<td>0.99</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>(AC_Pre → Implement. removed; Implement. → AC_Post removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A4</td>
<td>113.05</td>
<td>46</td>
<td>&lt; .001</td>
<td>2.46</td>
<td>0.06</td>
<td>0.99</td>
<td>0.96</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Table 14 continued

<table>
<thead>
<tr>
<th>Model A5</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(AC_Pre → Engage. removed; Engage. → AC_Post removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AC_Pre → Engage. removed; AC_Pre → Implement. removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Engage. → AC_Post removed; Implement. → AC_Post removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model A8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(AC_Pre → Engage. removed; AC_Pre → Implement. removed; Engage. → AC_Post removed; Implement. → AC_Post removed)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. AC_Pre = athletic competence (pre-test); AC_Post = athletic competence (post-test); Implement. = program implementation; Engage. = program engagement; $\rightarrow =$ path; $\chi^2 =$ chi-square; df = degrees of freedom; NC = normed chi-square; RMSEA = root mean square error of approximation; CFI = comparative fit index; GFI = goodness-of-fit index; SRMR = standardized root-mean square residual.
To obtain a superior theoretical and empirical athletic competence model, seven alternative models were tested, incorporating various modifications to the initial proposed model (see Figure 13). The first alternative model tested (Model A2) did not include a path from athletic competence (pre-test) to program implementation. Therefore, program implementation was only specified as a predictor of athletic competence (post-test). In evaluation of Model A2, all observed t-values were found significant at the $\alpha < .05$ level of significance. In relation to the latent constructs, program implementation did not significantly predict athletic competence post-test scores ($p = .544$), while all other structural paths were found significant. Upon review of model fit (Table 14), five of the six global fit indices suggested a good fitting model. The standardized residuals were fairly small, ranging from -2.71 to 1.94 and there were no major theoretically justifiable MIs to employ. When comparing Models A1 and A2, the $\chi^2$ difference test resulted in a non-significant difference between the models ($p = .583$).

The next alternative model (Model A3) does not include the program implementation latent construct as a mediator. Moreover, the paths from athletic competence (pre-test) to program implementation, and program implementation to athletic competence (post-test) were removed. All observed variables and structural equations were found statistically significant at the $\alpha < .05$ level of significance. When considering the fit of the model (Table 14), five of the six global fit indices indicated a good fitting model. In relation to the standardized residuals, all values were relatively small, ranging from -2.71 to 1.94. Additionally, there were no theoretically justifiable MIs associated with a substantial decrease in $\chi^2$. A $\chi^2$ difference test was conducted to
compare Models A1 and A3, for which the findings demonstrated a non-significant difference between the models ($p = .427$).

Model A4 maintains program implementation as a mediator of change in athletic competence, but does not include the path from athletic competence (pre-test) to program engagement. Therefore, program engagement is only a predictor of athletic competence post-test scores. SEM was run, resulting in an acceptable fitting model (Table 14), with four of the six global fit indices suggesting a good fit. In relation to the significance of the pathways, all observed $t$-values were found significant, while the program implementation latent construct was found to be a non-significant mediator at the $\alpha < .05$ level of significance. The largest standardized residuals were found between the athletic competence (pre-test) observed variables and the program engagement observed variables (ranging from 2.50 to 4.71), contradicting the removal of the path from athletic competence (pre-test) to program engagement. The modification suggestion with the largest anticipated decrease in $\chi^2$ is the addition of a path from athletic competence (pre-test) to program engagement (30.1 decrease in $\chi^2$), also contradicting the removal of this path in Model A4. When comparing Models A1 and A4, the $\chi^2$ difference test found a statistically significant difference between the models, with Model A1 demonstrating a better fitting model.

The next alternative model tested (Model A5) did not include program engagement as a mediator, for which the paths from athletic competence (pre-test) to program engagement and then to athletic competence (post-test) were removed. Results of the SEM indicate significant $t$-values at the $\alpha < .05$ level of significance, for all
observed variables. In relation to the latent variables, there was a significant path between the athletic competence constructs, while program implementation was found to be a non-significant mediator of change in athletic competence (pre-test to implementation: $p = .192$; implementation to post-test: $p = .135$). When considering the fit of the model (Table 14), four of the six global fit indices suggested a good fitting model. The largest standardized residuals were found between the athletic competence latent constructs and program engagement latent construct (ranging from 2.50 to 5.60), contradicting the removal of program engagement as a mediator of change in athletic competence.

Similarly, the theoretically justifiable MIs with the largest anticipated decrease in $\chi^2$ were the addition of paths from athletic competence (pre-test) to program engagement (36.8 decrease in $\chi^2$) and program engagement to athletic competence (post-test; 25.6 decrease in $\chi^2$). A $\chi^2$ difference test was conducted to compare Models A1 and A5, for which a significant difference was found between the models ($p < .001$). When examining the model fit and modification suggestions of the two models, Model A1 was found to be a superior model.

Another alternative model tested (Model A6) specified program implementation and engagement as antecedents of athletic competence post-test scores, but not outcomes of athletic competence pre-test scores. Therefore, the paths from athletic competence (pre-test) to program implementation and engagement were removed. Once again, all observed $t$-values were found significant at the $\alpha < .05$ level of significant. In relation to the latent constructs, all paths were found statistically significant except for the path from program implementation to athletic competence (post-test; $p = .584$). Upon review of the
global fit indices (Table 14), four of the six fit indices suggested a good fitting model. The majority of standardized residuals were within the +/- 1.96 threshold previously established. However, the largest standardized residuals were between athletic competence (pre-test) and program engagement latent constructs, contradicting the removal of this path. The theoretically justifiable modification suggestion with the largest anticipated decrease in $\chi^2$ was to add a path from athletic competence (pre-test) to program engagement (31.0 decrease in $\chi^2$), once again supporting the inclusion of this path. When comparing Model A6 with the initial proposed Model A1, the $\chi^2$ difference test demonstrated a significant difference between the models ($p < .001$), for which Model A1 is a better fitting model.

In contrast to the previous model tested, Model A7 did not include program implementation and engagement as antecedents of athletic competence (post-test). Therefore the paths from program implementation and engagement to athletic competence (post-test) were removed. Results of the analysis indicate statistically significant observed $t$-values, supporting the measurement models previously established. In relation to the structural equations, the only path that was non-significant was from athletic competence (pre-test) to program implementation ($p = .706$). Model A7 was found to be a good fitting model (Table 14), with five of six global fit indices indicating an acceptable fit. The standardized residuals were relatively small, ranging from -2.90 to 3.31, with the largest values between the athletic competence (post-test) observed variables and behavioral engagement. Upon review of the modification suggestions, the theoretically justifiable MIs included the addition of a path (25.3 decrease in $\chi^2$) and error
covariance (21.1 decrease in $\chi^2$) between program engagement and athletic competence (post-out). To determine the best fitting model, Models A1 and A7 were compared using a $\chi^2$ difference test. A statistically significant difference was found between the models ($p < .001$), for which Model A1 demonstrated a better model fit.

The last alternative model tested (Model A8) did not include program implementation nor program engagement as mediators of athletic competence, therefore the only predictor of athletic competence post-test scores were the athletic competence pre-test scores. Review of the results revealed statistically significant observed t-values and structural equation at the $\alpha < .05$ level of significance. The goodness of fit statistics suggest an acceptable model fit as four of the six global fit indices indicate a good fitting model (Table 14). To determine if any modifications are theoretically and empirically justifiable, the standardized residuals, MIs, and EPC statistics were surveyed. In relation to the standardized residuals, the largest residuals were between the athletic competence observed variables and program engagement observed variables (ranging from 2.50 to 5.66), contradicting the removal of program engagement as a mediator of change in athletic competence. Similarly, the MIs with the largest EPC statistic included the addition of paths from athletic competence (pre-test) to program engagement (37.6 decrease in $\chi^2$) and then to athletic competence (post-test; 26.9 decrease in $\chi^2$), as well as the inclusion of an error covariance between program engagement and athletic competence (post-test; 25.8 decrease in $\chi^2$). When comparing Model A8 to Model A1, a $\chi^2$ difference test showed a significant difference between the models. Based upon the
model fit and theoretically justifiable modification suggestions, Model A1 is the superior model.

To determine the best model, the initial proposed model and seven alternative models were compared for goodness of fit, of which all eight models demonstrated acceptable fit. However, of the eight models tested, Models A1, A2, and A3 explained the most amount of variance in the athletic competence post-test scores ($R^2 = .68$) and were associated with a greater number of global fit indices indicating an acceptable model fit. Models A4 through A8 were significantly different from Model A1, while Models A2 and A3 demonstrated a non-significant difference from Model A1. When comparing Models A1, A2, and A3, program implementation was consistently found to be a non-significant latent construct. However, removing program implementation as a mediator of change in athletic competence did not significantly improve the fit of the model (Table 14). Moreover, removing program implementation did not change the amount of variance explained in athletic competence post-test scores. Given that the removal of program implementation as a mediator did not significantly deteriorate the model, nor decrease the athletic competence (post-test) variance explained, Model A3 (which removed program implementation as a mediator) is considered the best model.

**Model B.** The proposed structural model (i.e., Model B1) examined program engagement as a mediator of change in athletic competence and program implementation as an antecedent of program engagement. This model hypothesizes that the staff’s implementation of the program influences the youth’s engagement in the program, for which engagement in the program causes development in athletic competence. To
determine the best model, four alternative models were tested in addition to the initial proposed model (Model B1). Figure 14 illustrates the five structural models tested.
Figure 14

Path diagrams of athletic competence Model B structural models

Note. Ath. Comp. = athletic competence; Program Implement. = program implementation; and Program Engage. = program engagement.
The initial proposed structural model (Model B1) was inputted into LISREL 9.1 software for SEM. A path was specified between the athletic competence latent constructs to show change over time. Program engagement was specified as a mediator by adding paths from athletic competence (pre-test) to program engagement and then to athletic competence (post-test). The final path added was from program implementation to program engagement to test program implementation as a predictor of program engagement. For each measurement model within the structural model, the reference variable was fixed to once, based upon the reference variables previously established [i.e., teamwork (pre-test and post-test); adherence to program structure – sports coaching; leader support]. This allowed for comparison across the factor loadings. The final constraints imposed consisted of error covariances. For each of the corresponding athletic competence observed variables (i.e., items 1, 2, and 3), an error covariance was added as the same items were used at the pre-test and post-test. Additionally, an error covariance was added between the effort and self-control observed variables at the post-test, based upon the measurement model previously established.

SEM was conducted to test the initial proposed model (Model B1). To evaluate the model, the observed \( t \)-values were compared to the critical \( t \)-value of a two-tailed test at the \( \alpha < .05 \) level of significance. All observed \( t \)-values were found statistically significant, as well as all structural equations, at the \( \alpha < .05 \) level of significance, supporting the inclusion of all measurement and structural paths. Upon examination of the goodness of fit statistics (Table 15), Model B1 demonstrated good fit with five of the six global fit indices indicating an acceptable model fit.
Table 15

**Structural equation modeling goodness of fit indices for athletic competence Model B**

<table>
<thead>
<tr>
<th>Model Tested</th>
<th>$\chi^2$</th>
<th>$Df$</th>
<th>$p$</th>
<th>NC</th>
<th>RMSEA</th>
<th>CFI</th>
<th>GFI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model B1</td>
<td>82.06</td>
<td>46</td>
<td>&lt;.001</td>
<td>1.78</td>
<td>0.04</td>
<td>0.99</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>Model B2</td>
<td>114.35</td>
<td>47</td>
<td>&lt;.001</td>
<td>2.43</td>
<td>0.06</td>
<td>0.99</td>
<td>0.96</td>
<td>0.10</td>
</tr>
<tr>
<td>(AC_Pre $\rightarrow$ Engage. removed)</td>
<td>104.58</td>
<td>47</td>
<td>&lt;.001</td>
<td>2.23</td>
<td>0.05</td>
<td>0.99</td>
<td>0.96</td>
<td>0.05</td>
</tr>
<tr>
<td>Model B3</td>
<td>82.07</td>
<td>46</td>
<td>&lt;.001</td>
<td>1.78</td>
<td>0.04</td>
<td>0.99</td>
<td>0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>(AC_Pre $\rightarrow$ Engage. removed; AC_Pre $\rightarrow$ Implement. added)</td>
<td>104.62</td>
<td>47</td>
<td>&lt;.001</td>
<td>2.23</td>
<td>0.05</td>
<td>0.99</td>
<td>0.96</td>
<td>0.05</td>
</tr>
</tbody>
</table>

*Note. AC_Pre = athletic competence (pre-test); AC_Post = athletic competence (post-test); Implement. = program implementation; Engage. = program engagement; $\rightarrow =$ path; $\chi^2$ = chi-square; $Df$ = degrees of freedom; NC = normed chi-square; RMSEA = root mean square error of approximation; CFI = comparative fit index; GFI = goodness-of-fit index; SRMR = standardized root-mean square residual.*
Although Model B1 demonstrated good fit, the standardized residuals, MIs, and EPC statistics were reviewed to determine theoretically justifiable model modifications. The standardized residuals were fairly small, ranging from -2.75 to 2.57, not far outside the +/- 1.96 threshold previously established. In relation to the MIs, there were no modification suggestions expected to result in a large decrease in $\chi^2$. Therefore, the initial proposed model (Model B1) was found theoretically and empirically sound.

To determine the best model, four alternative models were tested, which applied various modifications to the initial proposed model (Model B1). The first alternative model tested (Model B2) did not include program engagement as an outcome of athletic competence pre-test scores. Therefore, the path from athletic competence (pre-test) to program engagement was removed. Upon review of the measurement and structural paths, all observed and latent variables were found statistically significant at the $\alpha < .05$ level of significance. In relation to the model fit (Table 15), four of the six global fit indices indicated a good fitting model. Results revealed relatively large standardized residuals between the athletic competence (pre-test) observed variables and program engagement observed variables, contradicting the removal of athletic competence (pre-test) as a predictor of program engagement. Similarly, the theoretically justifiable MI with the largest EPC suggested the addition of a path from athletic competence (pre-test) to program engagement (31.0 decrease in $\chi^2$). In comparison to Model B1, Model B2 is significantly different ($p < .001$), based on the $\chi^2$ difference test. Of the two models, Model B1 is superior due to the model’s goodness of fit and $R^2$ of the athletic competence post-test scores.
In contrast to Model B2, alternative Model B3 does not include program engagement as a predictor of athletic competence post-test scores, but rather an outcome of athletic competence pre-test scores and program implementation. Therefore, the path from program engagement to athletic competence (post-test) was removed. When evaluating the model, all observed and latent $t$-values were found statistically significant at the $\alpha < .05$ level of significance. In relation to the fit of the model (Table 15), five of the six global fit indices indicated an acceptable model fit. The last consideration was the standardized residuals, MIs, and EPC statistics. The standardized residuals were fairly small (ranging from -2.90 to 3.31), with the largest residuals between behavioral engagement and the athletic competence (post-test) observed variables. In relation to the theoretically justifiable modification suggestions, a path and error covariance from program engagement to athletic competence (post-test) were expected to result in a relatively large decrease in $\chi^2$ (25.3 and 21.1 decrease in $\chi^2$ respectively). These results do not support the removal of program engagement as a predictor of athletic competence post-test scores. When comparing Models B1 and B3, the $\chi^2$ difference test found a significant difference between the models ($p < .001$), for which Model B1 is a better fitting model.

The next alternative model tested (Model B4) did not include program engagement as an outcome of athletic competence pre-test scores, but rather program implementation as an outcome of athletic competence pre-test scores. Therefore, the path from athletic competence (pre-test) to program engagement was removed and a path from athletic competence (pre-test) to program implementation was added. Results of the SEM
demonstrated significant observed $t$-values at the $\alpha < .05$ level of significance. In relation to the structural components, athletic competence pre-test scores and program engagement were found to be significant predictors of athletic competence post-test scores ($p < .001$). However, program implementation was found to be a non-significant predictor of program engagement ($p = .573$) and a non-significant outcome of athletic competence pre-test scores ($p = .565$). Although not all structural paths were found statistically significant, the model demonstrated an acceptable fit (Table 15), with five of the six global fit indices indicating a good fitting model. The standardized residuals were fairly small, ranging from -2.66 to 2.88. Similarly, the modification suggestions were associated with nominal EPC statistics. When comparing Models B1 and B4, the goodness of fit statistics were equivalent, therefore there was no difference in model fit.

The final alternative model tested (Model B5) did not include program engagement as a mediator of change in athletic competence, only an outcome of program implementation. Additionally, Model B5 specified program implementation as an outcome of athletic competence pre-test scores. Moreover, the paths from athletic competence (pre-test) to program engagement then to athletic competence (post-test) were removed and a path from athletic competence (pre-test) to program implementation was added. When evaluating the model, all observed $t$-values were found significant at the $\alpha < .05$ level of significance. In relation to the structural components, the only path found significant was between the athletic competence constructs ($p < .001$). Program implementation was found to be a non-significant outcome of athletic competence pre-test scores ($p = .565$) and a non-significant predictor of program engagement ($p = .572$).
Model B5 demonstrated an acceptable fit (Table 15), with five of the six global fit indices suggesting a good fitting model. To further evaluate the model, the standardized residuals, MIs, and EPC statistics were reviewed. Although the standardized residuals did not considerably fall outside the +/- 1.96 threshold previously established (ranging from -2.54 to 3.32), the largest residuals were between the program engagement observed variables and athletic competence (post-test) observed variables, contradicting the removal of the path from program engagement to athletic competence (post-test). The theoretically justifiable modification suggestions with the largest EPC statistic included the addition of a path from program engagement to athletic competence (post-test; 25.4 decrease in $\chi^2$) and error covariances between program engagement and athletic competence (post-test; 21.1 decrease in $\chi^2$) as well as program implementation and athletic competence (post-test; 20.0 decrease in $\chi^2$). A $\chi^2$ difference test revealed a significant difference between Models B1 and B5 ($p < .001$), with Model B1 demonstrating a better fit and accounting for a greater amount of the variance in athletic competence post-test scores ($R^2 = .68$).

In order to determine the best model, the initial proposed model and four alternative models were compared. All five models demonstrated an acceptable fit, however Models B1 and B4 explained the greatest amount of variance in athletic competence post-test scores ($R^2 = .68$) and demonstrated the best goodness of fit. The primary difference between Models B1 and B4 is the latent construct specified to be the outcome of athletic competence pre-test scores. In Model B1, program engagement is specified to be the outcome of athletic competence (pre-test), while in Model B4,
program implementation is specified to be the outcome of athletic competence (pre-test). When comparing Models B1 and B4, there is a non-significant $\chi^2$ difference between the models. In relation to the significance of the structural paths, Model B1 demonstrated statistical significance for all structural components at the $\alpha < .05$ level of significance. In comparison, Model B4 was associated with two non-significant structural paths, including the paths from athletic competence (pre-test) to program implementation and then to program engagement. Based upon these results, Model B1 is considered the best model, being both theoretically and empirically sound.

**Final model.** A purpose of the current study was to examine how program implementation and engagement impact athletic competence development of youth. As such, two primary models were proposed – Model A and Model B. Model A specifies both program implementation and engagement as mediators of change in athletic competence, while Model B specifies program engagement as a mediator of change in athletic competence and program implementation as a predictor of program engagement. In addition to these proposed models, 11 alternative models, applying various modifications to Models A and B, were tested to determine the best athletic competence model. Of the 13 models tested, two models were considered superior, including Models A3 and B1.

Model A3 is an alternative model which did not include the program implementation latent construct, therefore only program engagement was specified as a mediator of change in athletic competence. Model A3 was found to be a good fitting model (Table 14), with all observed $t$-values and structural paths statistically significant
at the $\alpha < .05$ level of significance. The model accounted for 68% of the variance in athletic competence post-test scores. Additionally, the standardized residuals and MIs did not support any modifications to the model.

Model B1 (proposed Model B) was also found to be a superior athletic competence model, for which program engagement was specified as a mediator of change in athletic competence and program implementation a predictor of program engagement. Model B1 was also a good fitting model (Table 15), with all measurement and structural paths statistically significant at the $\alpha < .05$ level of significance. Furthermore, the model explained 68% of the variance in athletic competence post-test scores, similar to Model A3. Lastly, Model B1 was associated with nominal standardized residuals and MIs.

When comparing Models A3 and B1, both models were found to demonstrate an acceptable fit, with statistically significant parameters, and inconsequential modification suggestions. Moreover, both models explained the same amount of variance in athletic competence post-test scores ($R^2 = .68$). A $\chi^2$ difference test revealed a non-significant difference between Models A3 and B1 ($p = .517$). Although the law of parsimony would support Model A3 based upon the smaller number of parameters in the model, Model B1 was maintained as the final athletic competence model for the current study (Figure 15; Kline, 2005).
Figure 15

*Athletic competence Model B1 with standardized solutions*
As previously mentioned, the purpose of the current study is to examine the impact of both program implementation and program engagement on athletic competence development. Although inclusion of the program implementation latent construct does not result in a greater amount of athletic competence (post-test) variance explained, the construct allows for a greater understanding of program dosage and how program implementation and engagement interact to influence athletic competence development. If the program implementation latent construct was removed from the model due to a statistical guideline that supports parsimonious models, a theoretically justifiable and empirically supported relationship would be lost. Theory proposes experiential learning (Kolb, 1984; Kolb et al., 2001), cognitive behavioral theory (Becky et al., 1979; Hupp et al., 2012), the Sport Education Model (Siedentop, 1998), and the TPSR model (Hellison, 2003) as frameworks for the learning and development process. The theoretical implications of program implementation as an antecedent of program engagement are superior to the guideline of model parsimony. By maintaining both program implementation and program engagement latent constructs, the results provide for a more comprehensive discussion to answer the current research questions. Therefore, Model B1 was selected as the final athletic competence model to further the exploration of organizational effectiveness within the context of sport-based PYD.
Chapter 5: Discussion

The current study was designed to examine the organizational effectiveness of a sport-based PYD program. More specifically, this study investigated how program dosage, measured by program implementation and engagement, mediates PYD. To answer the current research questions, SEM was employed resulting in two structural models demonstrating the relationships between program dosage and PYD. The following chapter will discuss the findings related to the research questions, bringing in theoretical and empirical support to explain the results. Implications for professionals will be provided to translate the findings for sport and youth development practitioners. Lastly, the limitations of the study will be reviewed along with future directions for this line of inquiry.

Organizational Effectiveness of a Sport-Based PYD Program

Within PYD, a program is considered effective if it is evidence-based, which is demonstrated by program relevance, theoretical explanations for the program design, and evaluation of program outcomes (Baldwin et al., 2005). The current study contends that an evidence-based program must not only demonstrate theoretical support for the program design, but also empirical support through conducting both process and outcome evaluation. For a program to claim effectiveness, the program must validate that when
properly implemented, the desired outcomes are achieved (Dusenbury et al., 2003; Law & Shek, 2011). Therefore, to determine the organization effectiveness of the sport-based PYD program under investigation and extend our understanding of the programmatic mechanisms of PYD, two structural models were proposed and tested, integrating program dosage and youth outcomes.

Prior to testing the structural models, the data were first treated for missingness and validity and reliability of the staff and youth measures were established. The data were reduced to mean scores based upon the validity and reliability testing, for which nonnormal factor scores were transformed using a power transformation (Bruce et al., 2008). A two-step SEM approach was adopted, establishing the measurement models before testing the structural models (Kline, 2005). Program dosage was found to best be explained by a two-factor model, consisting of program implementation and engagement. Based upon this result, two different structural models (Model A and B) were proposed and tested to answer the current research questions. Model A hypothesized both program implementation and engagement as mediators of PYD, while Model B hypothesized program implementation as an antecedent of program engagement and program engagement as a mediator of PYD.

Models A and B, along with 11 alternative models, were tested for R1 and R2. Model B1, in which program engagement mediated PYD and program implementation acted as an antecedent of program engagement, was found to be the best model to explain how program dosage mediates both social competence and athletic competence development. However, within the social competence model, program implementation
was found to be a non-significant antecedent of program engagement, but was maintained based upon theoretical and empirical justification. In summary, program engagement was found to significantly mediate PYD and program implementation was found a significant antecedent of program engagement when examining athletic competence outcomes.

**Positive Youth Development**

Sport-based PYD programs intentionally integrate sport and life skill instruction within a series of structured sport activities, while addressing risk and protective factors, to produce PYD (Gould & Carson, 2008; Martinek & Hellison, 1997; Weiss et al., 2012). The current study evaluated PYD through measuring perceived social competence, effort, teamwork, self-control, and athletic competence using a longitudinal design (pre-test and post-test). Results of validity and reliability testing supported the inclusion of all items for each of the scales used as indicators of PYD, including: 5-item PSCS-II measuring general social competence (Anderson-Butcher et al., in press); 5-item modified MSOS commitment subscale measuring effort behaviors and values within the context of sport (Vallerand et al., 1997); 8-item Teamwork Scale for Youth measuring teamwork behaviors (Lower et al., in review); 8-item Social Sports Experiences Scale measuring actions and behaviors in sport which reflect self-control and social responsibility (Anderson-Butcher et al., 2010); and 3-item Perceived Athletic Competence scale measuring general athletic competence (Amorose, 2002).

The longitudinal measures used to evaluate PYD demonstrated acceptable validity and reliability, with all scales showing moderate evidence of invariance across time. These findings support the use of these scales as objective measures of PYD in
longitudinal research, within the context of sport-based PYD. Programs seeking to evaluate program outcomes must ensure the evaluative measures adopted appropriately match the learning goals and desired outcomes associated with the program components (Gosen & Washbusch, 2004). Ultimately, use of these measures can validate sport-based PYD program outcomes.

Sport-based PYD programs are designed to develop athletic and like skills through intentionally designed, structured sport activities (Anderson-Butcher, 2011; Anderson-Butcher et al., 2011; Weiss, 2008; Weiss et al., 2012). Results of the current study found the social and athletic competence pre-test scores to be positive, significant predictors of the corresponding post-test scores. When examining the descriptive statistics of the factor scores, youth reported growth in each of the PYD indicators except for self-control. This finding supports previous research which has found participation in a sport-based PYD program associated with social development (Anderson-Butcher et al., 2011; Donaldson & Ronana, 2006) and athletic competence (Anderson-Butcher et al., 2013; Papacharisis et al., 2007). Overall, the current study demonstrated that four of the five desired outcomes (i.e., development in social competence, effort, teamwork, and athletic competence) were achieved by the sport-based PYD program. These results support the value and utility of sport to holistically develop youth and prepare them for future success.

PYD theorists support the integration of multiple competencies within a program design to contribute to youths’ resiliency (Catalano et al., 2004; Fraser et al., 2004). The program under investigation is designed to enhance the social and athletic competence of
youth participants, targeting nine sports and four primary social skills. The social skills investigated (social competence, effort, teamwork, self-control) were found to load onto a singular latent construct representing overall social competence. These social skills were intentionally integrated into the curriculum to achieve the primary program objective of enhancing the social competence of youth participants. The social competence measurement model supports the use of multiple skills to develop a broader competency. PYD theorists contend that protective factors, such as individual competencies, can have a cumulative, additive effect in that a greater number of protective factors will strengthen positive youth outcomes (The Search Institute, 1997). Sport-based PYD programs should integrate multiple competencies to have a broader and stronger impact on youth development, while keeping in mind the necessary program dosage (number, duration, frequency, and intensity of the intervention) to produce the desired developmental outcomes (Baldwin et al., 2005; The Search Institute, 1997).

Of the PYD indicators, youth reported a decrease in perceived self-control. This finding contradicts research supporting the social developmental outcomes of sport-based PYD programs (Anderson-Butcher et al., 2011; Donaldson & Ronan, 2006). A possible explanation for the digression is content validity of the self-control measure (Andrew, Pedersen, & McEvoy, 2011). The Social Sports Experiences Scale is designed to measure actions and behaviors in sport which reflect both self-control and social responsibility (Anderson-Butcher et al., 2010). This measure may not be the best objective evaluation to assess self-control as the items also reflect social responsibility. It is critical to match a program’s curriculum and activities with the desired outcomes and adopt objective
measures which accurately evaluate these desired outcomes (Bohnert et al., 2010; Gosen & Washbush, 2004; Schoel et al., 1988).

Another explanation for the self-control digression is the possible ceiling effect (Salkind, 2010) as demonstrated by the large self-control pre-test factor score ($M = 4.22$ on a 5-point scale). With youth reporting high responses on the pre-test, there was little room for growth to be captured. To increase the range of potential responses, additional items could be added to the scale or the current items could be revised to make the measure more sensitive to pick up greater variance (Salkind, 2010). When developing curriculum to enhance PYD, programs should carefully consider their target population and create activities that have the optimal degree of challenge to develop the competencies of the participants and motivate engagement while not being too easy nor too challenging to limit or inhibit development (Anderson-Butcher, 2005). Although the PYD findings predominately support the effectiveness of the sport-based PYD program studied, outcome evaluation alone is insufficient evidence to claim organizational effectiveness.

**Program Dosage**

Evidence-based programs are considered effective based upon the assumption that exposure to an intentionally designed intervention, which has theoretical explanations, will produce the desired program outcomes (Nation et al., 2003). Exposure to an intervention requires program implementation and engagement, in that the intervention must occur for exposure to be possible and participants must engage in order to gain the benefits of the intervention. The dimensions of program implementation and engagement
were hypothesized to load onto a singular program dosage latent construct; however, a two-factor model was found to be a better conceptualization of program dosage. The two-factor model differentiated program implementation and engagement into distinct, correlated latent constructs.

The final program dosage measurement model supports the multi-dimensional implementation fidelity framework found in the literature, including elements of program adherence, program delivery, and participant responsiveness (Dusenbury et al., 2003; Fagan et al., 2008; Tucker & Blythe, 2008). The current study focused on program adherence (i.e., program implementation) and participant responsiveness (i.e., program engagement) as critical components of program dosage. The two-factor measurement model, as well as the final structural models, suggest that the dimensions operate differently (e.g., participant engagement mediates PYD while program implementation predicts participant engagement). Therefore when evaluating program dosage, it is imperative to examine the dimensions of implementation fidelity separately in order to tease apart how these dimensions interact and influence PYD. When conducting process evaluation to determine program fidelity, programs should explore the multiple dimensions of fidelity to gain a greater understanding of the participants’ exposure to the intervention as well as the strengths and weaknesses of the program.

**Program implementation.** Program implementation reflects the fidelity dimension of adherence to the program design, which consists of adhering to the specific strategies and approaches specified by the intervention, including the various program components and content (Dusenbury et al., 2003; Fagan et al., 2008; Tucker & Blythe,
The program components that the current study examined included: intervention structure, climate, and curriculum. Overall, the final models found the intervention structure and climate to measure program implementation which then predicted program engagement.

**Intervention structure.** Sport-based PYD programs intentionally integrate sport and life skill instruction within the intervention to enhance holistic development (Gould & Carson, 2008; Martinek & Hellison, 1997; Weiss et al., 2012). The program under investigation integrated sport and social skill instruction within an experiential learning framework (Kolb et al., 2001), adopting the three-step approach of brief, facilitate, debrief (Schoel et al., 1988). The PCA results supported a two-factor conceptualization of intervention structure, for which sport and social skill coaching loaded onto separate factors. The two factors incorporated items representative of the three-step experiential learning framework (Kolb et al., 2001; Schoel et al., 1988), supporting this theoretical learning process.

Sport-based PYD theorists propose that life skills are not developed through sport participation alone (Danish et al., 2005; Danish & Nellen, 1997; Gould & Carson, 2008), rather both sport and life skills must be intentionally taught (Danish & Hale, 1981). The findings of the current study support this assertion of intentionally teaching sport and life skills. The multi-dimensional intervention structure framework suggests that it might be difficult to seamlessly blend sport and life skill instruction. Therefore, practitioners should be intentional about incorporating both sport and life skill instruction, but allow flexibility in relation to how and when life skill instruction is integrated.
Of the 13 items hypothesized to measure implementation of the intervention structure, 2 items did not load onto either of the intervention structure factors (i.e., sports coaching, social skills coaching). The PCA results showed item 5 (i.e., “Staff gave instructions for activity”) to load onto the intervention climate factor. This finding contradicts the three-step experiential learning framework, in that the purpose of an activity and instructions for an activity are presented during the briefing stage of experiential learning to provide focus and clarity (Schoel et al., 1988). The item may have not loaded onto either of the sports coaching or social skills coaching factors based upon its neutrality, for which all loaded items incorporated explicit sport or social skill language (e.g., Item 3 “Staff gave instructions of sport cues”). Additionally, providing instructions for an activity places emphasis on mastering a task which reflects a mastery-oriented climate (Elliot, 2005). Instructions are a necessary component for task mastery, therefore item 5 was allowed to load onto the intervention climate factor. Overall, instructions are a critical component of experiential learning as they provide focus and clarity (Schoel et al., 1988); however, instructions may have a greater effect on the learning environment rather than learning processes.

The second item to not load onto either of the sports coaching or social skills coaching intervention structure factors was item 13 (“Staff engaged youth in debriefing process”), which loaded onto an ambiguous factor. This finding contradicts the three-step experiential learning framework, in that the debriefing phase incorporates time for reflection facilitated through open-ended questions and prompts to engage the participants (Schoel et al., 1988). The debriefing phase requires staff to engage the youth
to facilitate the change process in which reflection translates into decisions and goals. A possible explanation for this finding is in relation to the language of the item. Similar to item 5, item 13 was neutral and not specific to sport nor social skill coaching. Moreover, the other debriefing-related items are specific tasks (e.g., “Staff discussed using S.E.T.S. outside of camp”), while item 13 was more general. This item may be more reflective of the program engagement latent construct as program engagement requires both youth to participate and staff to entice participation (“Engage”, n.d.). From a practical standpoint, engaging youth is necessary for the debriefing phase to be effective in solidifying and transferring the learning process. In relation to scale construction, this finding supports the use of items created at a micro level (e.g., identifying specific behaviors that contribute to the overall goal – debriefing).

When considering the distribution of the data from the current study, the sports coaching factor score was negatively skewed and leptokurtic with a larger mean ($M = 7.73$, $SD = 0.34$), while the social coaching factor score was normally distributed with a smaller mean ($M = 6.79$, $SD = 0.54$). In relation to the type of competencies being developed, sport skills reflect hard, technical skills while social skills reflect soft skills. Sport skills can be easier to articulate, demonstrate, model, evaluate, and reinforce because they are explicit technical skills. Hard, technical skills are teachable skills which can be defined and measured, consisting of the ability and knowledge of specific methods and techniques necessary to perform a task (Lussier & Achua, 2013). Comparatively, soft skills, which include personal qualities, interpersonal skills, as well as social and behavioral skills (Schulz, 2008), have been recognized as more difficult to teach due to
their ambiguous and subjective nature (Terego, 2009). Instructors have claimed that soft skills are difficult to teach as they are dependent on the learner’s personality (Deil-Amen, 2006). It can be difficult to develop or enhance soft skills among youth because it requires self-awareness and a commitment to change (Schulz, 2008).

Education theorists emphasize the need for curriculum which integrates hard and soft skills to develop a well-rounded individual (Schulz, 2008; Terego, 2009). Although it is challenging to screen prospective staff for soft skills as these skills are ambiguous and subjective, it is critical to hire staff with competencies in teaching and modeling both hard and soft skills. In relation to staff training and curriculum development, practitioners should provide staff extra support for soft skill instruction (e.g., incorporating social cues in the curriculum) in order to ensure competence and high fidelity.

The final 5-item sports coaching and six-item social skills coaching scales were found valid and reliable. The results support the use of these scale within the context of sport-based PYD as a measure of adherence to an integrated sport and social skill experiential learning framework. The multi-dimensionality suggests that the instruction of distinct competencies (sport vs. social skill instruction) should be assessed separately, which would allow for a greater understanding of the specific learning processes which contribute to PYD. Additionally, the final scales support the assessment of each phase of experiential learning (i.e., brief, facilitate, debrief), as each phase was represented among the loaded items. As previously mentioned, practitioners must use objective measures which correspond with their program design in order to accurately assess program dosage.
**Intervention climate.** The intervention climate is frequently cited as a critical feature of PYD programs, emphasizing components such as supportive relationships, opportunities to belong, fun, focus on task mastery, and opportunities for skill building (Anderson-Butcher et al., 2011; Eccles & Gootman, 2002). The program under investigation is designed to promote a caring, mastery-oriented climate. The PCA results supported a 4-item intervention climate factor, in which item 5 was added, based upon theoretical and empirical justification previously discussed, and items 18 and 22 were removed to improve the reliability of the scale.

Upon further analysis of the removed items, items 18 (“Staff created an emotionally safe climate”) and 22 (“Staff encouraged participation by all youth”) demonstrated the greatest variance by sport type. The program under investigation facilitates both contact and minimal contact sports. The contact sports include basketball, football, lacrosse, and soccer, while the minimal contact sports consist of dance, softball, swimming, and volleyball. When examining the mean implementation differences of the intervention climate items by sport type, items 18 and 22 demonstrated a 0.27 and 0.29 mean difference respectively, while the other intervention climate items demonstrated a 0 – 0.09 mean difference between sport types, with the results consistently showing lower implementation fidelity in contact sports.

The findings suggest that it is easier to create an emotionally safe climate and encourage participation in sports with minimal contact. Practitioners should address the influence of the activity type on ease of implementation fidelity during staff training and provide additional support in targeted activities (e.g., contact sports) throughout the
program to enhance fidelity. In regards to scale construction when assessing implementation fidelity, items should reflect program design features which can be consistently implemented.

The final 4-item measure of the intervention climate demonstrated acceptable validity and reliability. The items reflect staff practices associated with creating a caring, mastery-oriented climate which can be consistently implemented across curricula and activities. The findings support the use of this scale as a self-assessment of implementation of a caring, mastery-oriented intervention climate within the context of sport-based PYD. As previously mentioned, practitioners must use appropriate implementation fidelity measures which correspond with their program’s design, which may require the modification of existing tools or development of new measures.

**Intervention curriculum.** To enhance PYD, sport-based PYD programs intentionally integrate sport and life skill instruction through use of a dual-focused intervention curriculum (Weiss, 2008). The curriculum outlines the content and skills to be taught. The program under investigation uses a curriculum founded in the Sport Education and TPSR models (Hellison, 2003; Siedentop, 1998), which integrates sport and social skill instruction. The measure evaluating implementation of the intervention climate consists of four items reflecting use of the social skill curriculum.

The 4-item intervention curriculum measure was found valid and reliable, supporting the use of this scale as a tool to measure implementation of social skill curricula, specifically self-control, effort, teamwork, and social responsibility curricula. PYD theorists support the integration of competencies to contribute to PYD (Catalano et
al., 2004; Fraser et al., 2004). In relation to a curriculum’s scope, programs should integrate multiple competences to maximize development and resilience, while taking into consideration Siedentop’s (1998) assertion that fewer competencies taught in greater depth is more effective. Practitioners must determine an appropriate curriculum balance (i.e., number and depth of competencies) based upon their program design and PYD theory and empirical research. One approach that programs can take, similar to the program under investigation, is to focus on a broader skill (e.g., social competence) and break it down into a sequence of smaller skills (e.g., S.E.T.S.) which contribute to the broader skill and build upon each other (ACT Department of Education and Training, 2009; Bisson, 1999).

If a curriculum integrates multiple competencies, an assessment of curriculum implementation should distinguish the multiple competencies, rather than take a broad approach of evaluating overall use of the curricula. Evaluating implementation of the distinct curricula will provide valuable feedback regarding which curricula and/or competencies are easier or more difficult to teach, teasing apart program strengths and weaknesses. Practitioners can also use this evaluation as a means of monitoring a staff’s implementation of the curriculum in order to provide the necessary support for high fidelity.

**Program implementation measurement model.** When testing the construct validity of program dosage, a two-factor measurement model was found to be the best conceptualization of program dosage. The factors represented quantity and quality of program dosage, for which program implementation was considered a measurement of
The four dimensions of program implementation, as well as youth attendance, were hypothesized to measure a single latent construct. CFA found sports coaching and social skills coaching (intervention structure), as well as intervention climate, to significantly measure the latent construct, while youth attendance and intervention curriculum were found to be non-significant observed variables.

Youth attendance has frequently been used as a measure of program dosage in PYD (Anderson-Butcher et al., 2003; Anderson-Butcher et al., 2004; Anderson-Butcher & Cash, 2010; Anderson-Butcher & Fink, 2005; Gottfredson et al., 2010). However, there have been inconsistent findings regarding the outcomes of attendance, with several studies finding attendance unrelated to program outcomes (Anderson-Butcher et al., 2004; Anderson-Butcher & Fink, 2005). In regards to the current study’s descriptive statistics, the attendance data were found negatively skewed and leptokurtic, suggesting that a majority of the youth had high attendance. Many parents cited childcare, affordability, and beneficial programming as reasons for registering their child for the program under investigation, which explains the high attendance for a program that is voluntary. The limited variation of the data may inhibit the model from picking up the influence of attendance on PYD. Alternatively, program attendance does not guarantee quality exposure to the intervention as youth may be distracted, disruptive, or non-participative while in attendance. As attendance is necessary for exposure to an intervention, practitioners should adopt strategies to enhance attendance while prioritizing the quality of the exposure through motivating active youth involvement (Fiester, 2004).
When considering the program implementation dimensions, staff reported the lowest implementation fidelity of the intervention curriculum. The findings that implementation fidelity of the intervention curriculum was not only low, but did not significantly contribute to overall program implementation contradicts PYD and education theorists which argue that a purposefully designed curricula is necessary for learning and development (Danish & Hale, 1981; Weiss, 2008). The Sport Education and TPSR models support an integrated curriculum in which sport and social skills are taught simultaneously (Sidentop, 1998; Hellison, 2003). Moreover, PYD theorists argue that life skills cannot be learned through sport participation alone (Danish et al., 2005; Danish & Nellen, 1997; Gould & Carson, 2008), but must be intentionally taught through instruction, demonstration, modeling, and practice (Danish & Hale, 1981). A curriculum outlines how competencies should be taught, incorporating directive elements of instruction, demonstration, evaluation, and reinforcement.

Social learning theory provides a possible explanation for the removal of intervention curriculum from the program implementation measurement model (Bandura, 1971). Social learning theory proposes that learning can take place through direct experience and observation of others. Reinforcement which is informative and motivational can help regulate learned behaviors, while modeling can help develop new behaviors. Bandura and colleagues identified three models of observational learning, including the live model (demonstration of behavior), verbal instructional model (descriptions and explanations of behavior), and symbolic model (characters displaying behaviors; Bandura, Ross, & Ross, 1961). Youth may have learned social skills through
observation of the staff and peer group modeling the social skills. Modeling of social skills may consist of direct demonstration or indirect behaviors that are not intentionally performed due to directions from a curriculum but rather are innate social behaviors and personal qualities. The implications regarding this finding is limited as the research design for the current study only accounted for implementation of the social skill curriculum, neglecting implementation of the sport skill curriculum.

Overall, an intervention curriculum is intentionally designed to produce desired participant outcomes. Practitioners must ensure use of established curriculum, as high program fidelity increases the likelihood that the desired program outcomes will be achieved (Fagan et al., 2008). This can be accomplished through informal staff interviews, monitoring of staff practices, and review of the curriculum throughout the program. If a curriculum is comprehensive, overwhelming, and/or difficult to implement (e.g., soft skills are hard to teach; Deil-Amen, 2006), practitioners should consider highlighting the critical components that must be delivered, focusing the staff’s attention on what is most pertinent. Programs that are overwhelming have been found to be a barrier to high implementation fidelity (Dusenbury et al., 2003). Practitioners can enhance implementation fidelity by providing a comprehensive training prior to the program and support throughout the program. Keeping the detrimental effects of an overwhelming program in mind (Dusenbury et al., 2003), practitioners should integrate the multiple forms of learning (i.e., verbal instruction, live demonstration, use of symbolism) within a curriculum to enhance participant learning and development (Bandura et al., 1961).
Of the program implementation dimensions, staff reported highest implementation fidelity of the intervention climate. Upon examination of the data distribution, the intervention climate data were negatively skewed and leptokurtic, suggesting that a majority of the staff reported high fidelity of the intervention climate. A possible explanation for this finding is the two-way influential process of the leader-subordinate dyad (Herold, 1977). Organizational behavior theorists propose that the behavior of the leader will affect the behaviors and attitudes of the subordinate (Herold, 1977). A leader’s power will determine the strength of their influence, while their behaviors will affect the direction of that influence. Leaders must keep in mind that subordinates will react to their behaviors differently.

A primary objective of the program under investigation is to create a sense of belonging among the youth with the staff and program. This emphasis is reflected in the leader’s hiring, training, and supportive practices. For example, with the program’s target population vulnerable youth who live in poor conditions, staff training emphasizes strategies for working with youth (e.g., youth development principles, protecting youth, positive behavior management), highlighting the interactive and environmental features of the program. Based upon the leader-subordinate dyad, implementation fidelity will be influenced by what is emphasized by the leaders and program design. Leaders can use this power of influence to positively affect the behaviors of their subordinates. Through modeling and reinforcing expected behaviors, leaders can support program adherence.

For the current study, program implementation was measured through use of self-assessment. When considering the descriptive statistics, the majority of program
implementation dimensions (i.e., sports coaching, intervention climate, intervention
curriculum) demonstrated negatively skewed, leptokurtic data distributions, indicating
that most staff reported high fidelity. The literature suggests that use of self-assessments
can end in flawed results due to unrealistic optimism (Dunning et al., 2004). People often
have overinflated perceptions of their skill and performance, in that when compared to
objective measures, people’s claims exceed actual performance (Dunning et al., 2004).

To control for the potential inflation of staff self-assessment of implementation
fidelity, practitioners should ensure that the staff receive an accurate picture of what is
expected of them and what high implementation fidelity looks like. For example, the
program under investigation used two curriculum demonstrations during training, for
which the staff were trained on use of the implementation fidelity tool and underwent
validity checks. Practitioners should consider evaluating the staff’s understanding of the
program design prior to the beginning of the program to enhance implementation fidelity.
If using self-assessments to measure implementation fidelity, practitioners should
consider facilitating the evaluation through use of a third party so staff do not feel their
performance evaluation will have any direct or indirect effect on their job security.

Program implementation as an antecedent of program engagement. Following
the two-step approach, the program implementation measurement model was inputted
into the structural model, for which Model A hypothesized program implementation as a
mediator of PYD, while Model B hypothesized program implementation as an antecedent
of program engagement. Model B1 was found to be the best model, supporting program
implementation as an antecedent of program engagement. Program implementation
reflects the program adherence dimension of implementation fidelity. Additional
dimensions of implementation fidelity include dosage, quality of program delivery, and
participant responsiveness (Dusenbury et al., 2003; Fagan et al., 2008; Tucker & Blythe,
2008). The current study incorporated dosage into the model by calculating program
implementation based upon youth attendance. While quality of program delivery was not
addressed in the model, participant responsiveness is reflective of program engagement.

Model B1 suggests that the impact of program implementation is contingent upon
program engagement. If youth are not engaged, they will not receive full exposure to the
program implemented which will consequently influence their developmental outcomes.
Comprehensively, program dosage consist of responsibilities for both the staff and youth,
in that staff must adhere to the program design when delivering the intervention to
increase the likeliness of desired outcomes, and youth must engage and respond to the
intervention to gain any benefit from the intervention. As the impact of a program is
contingent upon program engagement, program implementation cannot be a separate
mediator of PYD because it requires program engagement to produce any effect.

While program implementation is not a separate mediator of PYD, it is still a
necessary component of PYD. Without the delivery of an intentional, evidence-based
intervention, the desired youth outcomes are unlikely to be achieved. As the effectiveness
of program implementation is contingent upon program engagement, it is imperative for
staff to establish relationships with youth at the very beginning of the program to entice
engagement and thus increase exposure to the intervention being implemented.
Ultimately, in order for a program to have the greatest chance for achieving the desired
outcomes, the program design must have theoretical explanations and be implemented as intended (Baldwin et al., 2005; Fagan et al., 2008).

Model B1, hypothesizing program implementation as an antecedent of program engagement, was found to be the best model for R1 and R2 based upon theoretical justification and empirical evidence. However, program implementation was found to be a non-significant predictor of program engagement when examining social skill development (R1), in comparison to athletic skill development (R2). This finding contradicts the TPSR model which proposes the integration of sport and social skill instruction to develop personal and social responsibility within sport (Hellison, 2003). Pascual and colleagues (2011) found that when implementation fidelity of the TPSR model is high, greater student outcomes are achieved. Additionally, Wright and Burton (2008) found TPSR to positively influence student behaviors. These studies adopted a qualitative approach, therefore their results conceptually associated TPSR with participant outcomes, while the current study attempted to empirically associate TPSR with youth outcomes and found a non-significant relationship. Additionally, the findings contradict empirical research which supports the use of experiential learning as a learning framework (Gosen & Washbush, 2004).

Social learning theory and the difficulty of teaching soft skills are possible explanations for the non-significant relationship of program implementation and program engagement within the social competence model. As previously discussed social learning theory postulates that learning takes place through experience and observation (Bandura, 1971). Gosen and Washbush (2004) highlight the difficulty of evaluating learning as it is
an internal, complex process and is unique to the individual. The current study captured verbal instruction (e.g., Session Log item 4 “Staff taught how to use S.E.T.S. in activity”) and reinforcement (e.g., Session Log item 8 “Staff praised campers when they demonstrated good S.E.T.S.”; Session Log item 9 “Staff pointed out use of S.E.T.S. in activity”). Verbal instruction and reinforcement of social skills may be difficult to implement as soft skills are ambiguous and subjective (Terego, 2009). The literature suggests that soft skills are more difficult to teach than hard skills (Deil-Amen, 2006), which may inhibit the quality and effectiveness of social skill instruction.

When considering the program implementation dimensions, staff reported the lowest implementation fidelity for the intervention curriculum (i.e., use of social skill curricula) and social skills coaching, supporting the assertion that social skills are difficult to teach. To enhance the implementation fidelity of social skill instruction, practitioners should provide in-depth training and support. Implementation of the curriculum can be taught, demonstrated, and reviewed prior to the program and throughout the program. Practitioners should consider including helpful cues and tips within the curriculum to ease implementation of the curriculum. Moreover, practitioners should conduct validity checks throughout the program (e.g., monitor staff and provide feedback) to ensure fidelity.

In addition to verbal instruction, social learning theory identifies two other models of observational learning which are likely to occur in an experiential-based program – modeling and use of symbolism. The findings of the current study suggest that verbal instruction alone may be an insufficient learning process for social skill development, but
significantly contributes to athletic skill development. While modeling and use of symbolism were not captured in the model, these forms of observational learning may be greater contributors to the learning process. Through engagement in the program, youth may be exposed to the staff and peer group modeling the social skills, resulting in enhanced social competence. To maximize learning and development, practitioners should seek to incorporate multiple forms of learning (e.g., verbal instruction, live demonstration, use of symbolism). Based upon social learning theory and the power of observational learning, practitioners should consider recruiting staff who possess strong social skills and are more likely to model those skills to the youth participants. Additionally, programs can encourage peer modeling of social skills through positive reinforcement (e.g., rewarding youth who demonstrate positive social skills).

Model B1, examining social skill development, also contradicts PYD literature which encourages the creation of a caring, mastery-oriented climate. Empirical research has demonstrated the positive outcomes of caring-based programming (Newton et al., 2007), supportive relationships (Allodi, 2010; Anderson-Butcher, 2010; Eccles et al., 2003; Larson & Walker, 2005; Ullrich-French & Smith, 2009), and a mastery-oriented climate (Barkoukis et al., 2008; Boyce et al., 2009; Miller et al., 2004; Newton et al., 2006). Eccles and Gootman (2002) contend that perceptions of supportive relationships is more important than specific supportive qualities, suggesting that the perceived environment may be more critical than the implemented environment. This assertion is supported by Riley’s (2013) study which found staff practices reflecting care and support to not significantly predict youth outcomes. As the current study examined perceived
belonging and leader support, as measures of program engagement, which reflect a caring environment and were found to mediate PYD, perceptions of the environment may be most critical. To enhance perceptions of the intervention climate, staff should build relationships with the participants in order to learn their distinct preferences and be able to gauge their affective response to staff practices and the overall environment.

The current study sought to extend the theoretical framework of an evidence-based program by incorporating both a theoretical and empirical connection between program components and outcomes (Baldwin et al., 2005). When examining athletic competence outcomes, Model B1 was found to be the best model, for which program implementation was a significant predictor of program engagement, which in turn mediated PYD. This finding provides empirical evidence that the program under investigation is evidence-based and supports Durlack and Weissburg’s (2007) study which found that evidence-based programs were more successful at achieving positive youth outcomes. Overall, empirical research has consistently found program implementation fidelity associated with program effectiveness (Catalano et al., 2004), in that adherence to program components (e.g., intervention structure, climate, curriculum) should result in stronger outcomes (Fagan et al., 2008). Although the current study did not have a control group to test the effectiveness of the program design, program implementation was found to indirectly contribute to PYD.

As program implementation was found a significant predictor of program engagement, and engagement a significant mediator of athletic competence, the program design which was implemented is supported as a significant programmatic mechanism of
PYD. More specifically, the findings support the three-step experiential learning framework (i.e., brief, facilitate, debrief) as an effective learning process within sport (Kolb et al., 2001; Schoel et al., 1988), contributing to Gosen and Washbush’s (2004) findings. In regards to the learning process, experiential learning theory highlights an experience as an agent for learning while social learning theory focuses on learning through observation (Bandura, 1971; Kolb et al., 2001). The current study’s results suggest that use of an experience (e.g., Session Log item 6 “Staff provided opportunity for youth to practice sport skill”), verbal instruction (e.g., Session Log item 3 “Staff gave instructions of sport cues”), and reinforcement (e.g., Session Log item 7 “Staff praised campers when they demonstrated good sport skill”) are effective learning processes for athletic competence. Practitioners should integrate multiple forms of learning (e.g., instruction, modeling, use of symbolism) within an intentional experience to maximize development.

The athletic competence model, in which program implementation is a significant predictor of program engagement, supports the critical features of a PYD setting proposed by PYD theorists (Anderson-Butcher et al., 2011; Eccles & Gootman, 2002). The findings highlight an environment which is inclusive, characterized by care and supportive relationships, with an emphasis on task mastery. These results support past empirical research which has demonstrated the positive outcomes of a caring, mastery-oriented program climate (Barkoukis et al., 2008; Boyce et al., 2009; Miller et al., 2004; Newton et al., 2006; Newton et al., 2007). To intentionally create this type of environment, practitioners should consider the program design as well as hiring and
training practices. Activities that primarily focus on skill development that lead to a culminating competition will promote task mastery while still adhering to key features of the Sport Education model (Siedentop, 1998). Additionally, programs should hire staff who have experience working with children and train staff on child management strategies which support a caring, relational climate. Overall, the findings of the current study support program implementation as a critical mechanism for PYD. Therefore, programs should be intentional not only in the development of an evidence-based intervention, but also in the delivery of the intervention. If an intervention is not implemented with high fidelity, the desired youth outcomes will be limited if at all realized.

To promote high fidelity of program implementation, practitioners should consider four key elements, including: teacher training, program characteristics, teacher characteristics, and organizational characteristics (Dusenbury et al., 2003). Training and staff development can increase knowledge and change staff attitudes toward the intervention. Researchers have found that live teaching (versus video-trainings), facilitated by an administrator with experience and understanding of the program, who uses detailed, concrete instruction, is most effective. In regards to program characteristics, programs should simplify implementation and reduce complexity by designing the intervention so that it is easy to administer, highlighting the essential elements of the program, and providing clear instructions. Research has demonstrated that teachers that are trained, confident, enthusiastic, and possess the skills to be taught are more likely to adhere to the program design and implement with high quality. Lastly,
high fidelity is dependent on the organization and is influenced by the leadership, support, communication, problem solving, and overall culture of the organization. Comparatively, organizational characteristics which act as barriers to effective implementation primarily consist of a lack of resources (e.g., money, time, staff) and turbulence or an organization that is overwhelmed.

**Program engagement.** Program engagement reflects an interaction between the intervention, staff, and participants. More specifically, program staff deliver the intervention and invite the participants to engage in the intervention, for which the participants respond through active involvement in the intervention. As program engagement incorporates elements of both staff practices and participant behaviors, the current study measured program engagement through perceived behavioral engagement, belonging, and leader support. Overall, the final models found program engagement to be a significant, positive mediator of PYD.

**Behavioral engagement.** To be fully exposed to an intervention, a participant must not only attend the program but also be actively involved (Feister, 2004). Participants who are aloof, distracted, disruptive, or resistant to participate will receive less exposure to an intervention, which is why program engagement is a critical component of program dosage. The current study examined behavioral engagement as a measure of participant engagement quality. The 5-item modified Engagement versus Disaffections with Learning behavioral engagement subscale was found valid and reliable, supporting the use of this scale as an objective measure of participant engagement behaviors within the context of sport-based PYD (Skinner et al., 2009).
The findings of the current study support research which has consistently found engagement in PYD programs associated with positive outcomes reflective of the programs’ objectives (Anderson-Butcher & Case, 2010; Bohnert et al., 2010; Fredricks & Simpkins, 2012; Mueller et al., 2011; Roth et al., 2010; Walker & Arbreton, 2004; Weiss et al., 2005). However, the quality of engagement (e.g., behavioral, emotional, cognitive engagement) has rarely been empirically investigated. The current study’s findings support the review conducted by Roth & colleagues (2010) which found engagement quality associated with positive youth outcomes. Overall, attendance is insufficient for youth development to occur, suggesting a deeper level of engagement necessary for program outcomes (Mahoney et al., 2007; Weiss et al., 2005). Practitioners should train staff on how to actively involve youth participants and facilitate activities which promote participation by all youth. Moreover, concepts such as time on task should be integrated into the program design in order to increase engaged time and accelerate development (Florida Education Association, n.d.).

**Belonging.** PYD theorists suggest several program design features which promote participant engagement, including creating a sense of belonging and developing positive, supportive relationships between adults and peers (Anderson-Butcher & Fink, 2005; Ullrich-French & McDonough, 2013). A primary objective of the program under investigation is to develop belonging and connectedness between the participants, program, and staff. To measure perceived feelings of belonging, the 6-item modified sense of belonging scale was utilized (Anderson-Butcher & Conroy, 2002). The tool was
found valid and reliable, supporting the use of this scale as an objective measure of perceived belonging within the context of sport-based PYD.

The current study found belonging to be a significant measure of program engagement, which in turn mediated PYD. These findings support previous empirical research which has found perceived belonging associated with several positive youth outcomes (Anderson-Butcher et al., 2014; Anderson-Butcher & Fink, 2005; McDonough et al., 2013; Newman et al., 2007). Additionally, the findings support the team affiliation feature of the Sport Education model, which encourages team cooperation and membership (Siedentop, 1998). Siedentop (1998) argued that much of the social meaning and development attributed to sport experiences is related to affiliation with a team. The program under investigation randomly assigns youth participants into a group, which persists throughout the entire program, in order to create a sense of belonging. Program staff can enhance a sense of belonging by providing opportunities for youth to develop positive relationships with their peers. Moreover, practitioners should integrate a team structure into the program design to strengthen the participants’ connection to a persisting group and consequently the program.

Leader support. Another PYD program feature suggested to entice participant engagement is leader support. Through care, warmth, support, and acceptance program staff can not only develop relationships with youth participants, but through those relationships influence their behaviors (Herold, 1977). An 8-item modified PSSM teacher support subscale was used to measure perceived feelings of leader support. Results of the pilot test suggested the removal of the two negatively worded items (items 2 and 3).
Therefore, to determine the most valid and reliable scale for the current study, two corresponding, positively worded items were added to retain the measurement content while examining the effectiveness of negatively versus positively worded items within a youth population.

Results of the study support the removal of the two negatively worded items and maintenance of the two corresponding positively worded items, suggesting that negatively worded items might not be effective for use within the youth population. Negatively worded items are often included during scale construction to detect invalid response sets (Eys, Loughead, Bray, & Carron, 2009); however, mixed items can confuse participants resulting in misinterpreted responses and reduced scale reliability (Eys, Carron, Bray, & Brawley, 2007; Spector, 1992). Eys and colleagues (2009) suggested the use of negatively worded items to detect invalid response sets, while not including the negatively worded items in the final measurement model. Overall, the reduced 6-item scale was found valid and reliable, supporting the use of this tool as an objective measure of perceived feelings of leader support within the context of sport-based PYD.

Findings from the current study found leader support to be a significant contributor to PYD, supporting past research which has found perceived leader support associated with many positive youth outcomes (Anderson-Butcher et al., 2004; DuBois & Silverthorn, 2005; McDonough et al., 2013; Ullrich-French & McDonough, 2013). Additionally, the TPSR teacher-student relationship theme is supported by the results of the current study (Hellison, 2003). Moreover, the leader-subordinate dyad may explain the influence perceived leader support has on participants’ behavioral engagement.
(Herold, 1977). To enhance perceived leader support, program staff should develop relationships with the participants, providing both emotional and instrumental support, while keeping in mind youths’ varying preferences and responses to relational qualities (Eccles & Gootman, 2002).

Program engagement measurement model. To investigate program engagement as a mediator of PYD, a program engagement measurement model was proposed and tested. Behavioral engagement, belonging, and leader support were hypothesized to measure a single latent construct reflecting program engagement. CFA results found all three observed variables to be significant indicators of the latent construct, supporting the assertion that program engagement incorporates elements of both participant and staff behaviors. Broadly defined, the term “engage” reflects the actions of inducing participation as well as active participation (“Engage”, n.d.). Practitioners only have control over the act of inducing participation, through intentional staff practices. Therefore, as program engagement was found to be a significant mediator of PYD and a necessary component of program dosage, it is critical to entice participation through adhering to theoretically and empirically supported PYD program design features. PYD theorists have identified several evidence-based design features, including: prosocial opportunities, ideal challenge, supportive relationships, physical and psychological safety, promotion of a mastery motivational climate, and fun (Anderson-Butcher, 2005; Anderson-Butcher et al., 2011; Eccles & Gootman, 2002; Weiss et al., 2005).

When examining the indicators of program engagement, behavioral engagement had a greater factor score than perceived belonging and leader support, which were
relatively similar. Belonging and leader support are suggested to entice participant engagement, but actual engagement is necessary for youth to be exposed to the intervention and gain the resulting benefits of that exposure. As behavioral engagement accounts for the greatest variance in program engagement, it is imperative that program staff monitor participant engagement and seek to motivate disengaged youth while positively reinforcing active involvement. For example, the program under investigation provides a button to youth who demonstrate one of the four social skills, which has been found to be an effective positive behavior management strategy (Anderson-Butcher, Wade-Mdivanian, Paluta, Lower, Amorose, & Davis, 2014).

The current study adopted self-report tools completed by the youth participants to measure behavioral engagement, belonging, and leader support. Danielson and Phelps (2003) support the use of self-report as a means to understand youth perceptions and cognitions. However, a multiple informant approach is preferred to gain a complete picture of the child. Practitioners can adopt a multiple informant approach by integrating youth self-report with parent and teacher ratings. However, combining multiple sources within an evaluation or research design may not be feasible based upon time and financial limitations.

*Program engagement as a mediator of positive youth development.* Adhering to the two-step approach, the program engagement measurement model was inputted into the structural models (Models A and B) as a mediator of PYD. Model B1 was found to be the best model, explaining how program dosage acts as a programmatic mechanism of PYD, for which program engagement was found to be a significant mediator of PYD.
Program engagement is reflective of the participant responsiveness dimension of implementation fidelity (Dusenbury et al., 2003; Fagan et al., 2008). To gain the benefits of an intervention, participants must first be exposed and respond to the intervention (Anderson-Butcher et al., 2003). Through program engagement, the implemented intervention acts as the catalyst for PYD.

Model B1 supports PYD theorists who contend that engagement in socialization experiences allow for the development of skills (Bohnert et al., 2010). The current study’s findings extend previous empirical research on program engagement and developmental outcomes by examining the quality of engagement (Roth et al., 2010; Weiss et al., 2005), which was found to be a significant mediator of PYD. This finding supports the review conducted by Roth and colleagues (2010) which found engagement quality associated with positive youth outcomes. As the impact of program engagement is contingent on program implementation, the programmatic mechanisms are interdependent, reflecting program dosage. Therefore, when considering the necessary dosage for an intervention to produce the desired effect, practitioners must consider both program implementation and engagement. This calls attention to the program staff’s dual responsibility of both delivering the intervention with high fidelity and inducing active involvement from participants simultaneously.

When comparing the social and athletic competence models, program engagement was found to be a stronger mediator of social skill development than athletic skill development. Social learning theory provides a possible explanation (Bandura, 1971). As previously discussed, social learning theory proposed three models of observational
learning, including the live model, verbal instructional model, and symbolic model. As program implementation was found to be a significant antecedent of program engagement within the athletic competence model, but not the social competence model, verbal instruction and reinforcement (measured by the Session Log tool) appear to be more effective for teaching hard skills, which are defined and measurable (Bandura et al., 1961; Lussier & Achua, 2013). However live modeling may be more effective when teaching social skills, which are ambiguous and subjective, as program engagement was found to be a stronger mediator of social skill development than athletic skill development (Bandura et al., 1961; Terego, 2009).

Through program engagement, participants interact with the staff and peer group who model various behaviors. During this interaction, participants are able to observe the behaviors of others and through that observation learn new behaviors (Bandura, 1971). Through engagement, youth can observe modeled social skills which enhances self-awareness of their own social skills and results in new learned behaviors (Bandura, 1971; Schulz, 2008). These learned behaviors can then be regulated through reinforcement and teachable moments (Bandura, 1971). Therefore, to enhance social learning and development, program staff should model the targeted social skills and reinforce modeled social skills through encouragement, teachable moments, as well as correction. Based upon the power of influence demonstrated by the leader-subordinate dyad, program staff should highlight and encourage participants who are demonstrating positive social skills and engagement while correcting participants who are distracted, resistant to
participation, or disruptive to not only reinforce social competence, but also enhance program engagement (Herold, 1977).

Upon examination of the athletic competence Model B1, the pre-test athletic competence scores, in comparison to program engagement, accounted for a greater amount of variance in the post-test athletic competence scores. This finding suggests that the intervention may not have been strong enough to produce the desired outcomes.

When considering the necessary program dosage to produce a desired effect, practitioners must consider the number, duration, frequency, and intensity of the intervention (Nation et al., 2003). The Sport Education model suggests sport seasons that have a minimum of 12 lessons (Siedentop, 1998). The program under investigation consists of a 19 day intervention, in which youth engage in each of the nine sports for five, one-hour lessons.

To strengthen an intervention, practitioners may consider increasing the number, duration, frequency, and/or intensity of the intervention (Nation et al., 2003), pending available resources. Alternatively, practitioners may consider decreasing the number of sports taught and increasing the depth per sport (i.e., scope). For a long lasting effect, practitioners should incorporate follow-up sessions after the initial intervention (Nation et al., 2003). For example, the program under investigation facilitates five sport clinics throughout the year to reinforce the targeted sport and social skills. Practitioners should keep in mind that the greater the deficits of the participants, the stronger the intervention must be to produce the desired outcomes (Carnahan, 1994).

There are many challenges and barriers which may inhibit program engagement, such as alternative engagements and responsibilities of the participant, safety and
transportation issues, disinterest of the participant, and lack of good rapport with staff (Roth et al., 2010; Weiss et al., 2005). Practitioners should seek to proactively address these barriers in order to enhance participant engagement. Although alternative engagements and responsibilities of the participant cannot be mitigated, practitioners can enhance the evidence-based program design features which entice participation to increase retention and engagement. To address safety and transportation issues, programs should be intentionally placed where the target population is located and has access to it (Lawson & Anderson-Butcher, 2000). If possible, programs should provide free transportation to reduce this barrier.

To recruit and retain youth, programs should provide new opportunities which fulfill a participant’s need for autonomy, competence, and relatedness (Anderson-Butcher, 2005). The activities should be designed to provide an ideal amount of challenge so that youth do not get bored or disinterested but can still experience success. Lastly, the adult-child relationship is a critical component of PYD and program engagement. PYD theorists consistently emphasize the importance of positive adult-child relationships (Allodi, 2010; Anderson-Butcher, 2010; Anderson-Butcher et al., 2004; Eccles & Gootman, 2002; Hellison, 2003; Larson & Walker, 2005; Ullrich-French & McDonough, 2013). Program staff should recognize that each participant has unique strengths and personality (Hellison, 2003). By extending care, warmth, acceptance, and support, program staff can develop positive relationships with the participants, consequently enhancing program engagement and ultimately PYD (Bowers et al., 2012; Eccles & Gootman, 2002).
Impact of the Study

Economically disadvantaged youth, living in impoverished areas of the United States, represent a vulnerable population exposed to risk in their environment which can negatively influence healthy development (Federal Interagency Forum on Child and Family Statistics, 2013; Fraser, 2004). Poverty can affect a youth’s access to quality food and housing, health services, and transportation to prosocial opportunities (Federal Interagency Forum on Child and Family Statistics, 2013). Furthermore, many youth living in impoverished homes engage in risky behaviors due to lack of supervision and social support (Forum on Child and Family Statistics, 2013; Fraser, 2004).

PYD emerged in response to the needs of vulnerable youth, for which sport-based PYD programs utilize intentional interventions designed to address risk factors and contribute to healthy development (Anderson-Butcher, 2011; Anthony et al., 2009; Lerner et al., 2005; Vierimaa et al., 2012). The effectiveness of sport-based PYD programs is imperative due to the positive impact the interventions can have on vulnerable populations. Research exploring the impact of these programs on both youth social and athletic competence continues to be limited (Newton, Watson, Kim, & Beacham, 2006). Only a few studies identified have focused specifically on the impact of summer-based programming (Newton et al., 2006; Schilling, Martinek, & Carsons, 2007; Schilling, 2001). Overall, research in relation to the unique challenges and outcomes of sport-based PYD programming in underserved communities is lacking (Anderson-Butcher et al., 2014; Fraser-Thomas et al., 2005).
To advance youth sport and assist practitioners with the effective design, implementation, and evaluation of sport-based PYD programs, research must consider “characteristics of interventions that are effective or ineffective” in promoting healthy youth outcomes (Chalip, 2006). The current study found program implementation and engagement to significantly influence positive youth development among an underserved youth population. More specifically, the study confirmed the positive outcomes of a sport-based PYD summer program through use of a longitudinal research design. Additionally, the study supported implementation of an experiential-based, mastery-focused intervention as a predictor of program engagement which was found to significantly mediate change in social and athletic competence. The findings of this study can be applied in practice to maximize the impact of sport-based PYD interventions, contributing to healthy development and successful transition to adulthood.

Implications for Practice

The results of the current study have implications for both practitioners and researchers in youth serving organizations. The following section will provide a summary of the implications for practice, with the caveat that many suggestions are dependent on a program’s available resources (e.g., time, personnel, funding; Dusenbury et al., 2003). More specifically, implications related to program design, staff recruitment and hiring practices, staff training, staff support, and program evaluation will be discussed.

Program Design

An evidence-based PYD program adopts theoretical frameworks and approaches when designing a purposeful intervention intended to produce desired outcomes (Baldwin
et al., 2005). The current study supports an experiential learning framework as an
effective learning process (Kolb et al., 2001). As sport-based PYD programs consist of a
sequence of activities (i.e., experiences) designed to enhance sport competence as well as
life skills, practitioners should consider adopting the three-step framework of brief,
facilitate, debrief to teach the targeted skills, provide an opportunity for the youth to
practice the skills, and transfer the learning to alternative contexts (Schoel et al., 1998).
These activities should integrate sport and life skill instruction to prepare youth for
success in diverse environments (Danish et al., 2005; Danish & Hale, 1981; Danish &

Use of a structured intervention curriculum is encouraged to enhance strategic,
intentional instruction of a program’s targeted skills (Catalano et al., 2004; Weiss, 2008).
When developing an intervention curriculum, practitioners should consider instruction
and activities that are age appropriate, to enhance participant engagement and
development (Fraser et al., 2004; Nation et al., 2003). Anderson-Butcher (2005)
contended that an optimal amount of challenge and opportunities to experience new and
exciting activities are critical factors influencing youth retention. If curriculum is
designed to be developmentally appropriate, as youth return to a program in subsequent
years, they will be exposed to new and challenging curriculum which will retain their
interest and allow for the continued development of competence.

The results of the current study support PYD theorists’ argument for the
integration of multiple competencies to strengthen the impact of an intervention
(Catalano et al., 2004; Fraser et al., 2004). Although the integration of multiple
competencies can have a cumulative, additive effect (Fraser et al., 2004), program dosage must be strong enough to produce the desired outcomes (Nation et al., 2003). Therefore, practitioners should carefully consider the scope of the curriculum, in relation to number and depth of skills taught, for which Siedentop (1998) claimed fewer activities with greater depth is most effective. Once the scope of the curriculum is determined, those skills should be broken down into a progressive sequence for effective learning and development (Clumpner, 2003; Schoel et al., 1988). Practitioners can provide cues and helpful tips, highlighting the most important elements within a curriculum, to enhance implementation fidelity, especially for social skill instruction which was found to be difficult to adhere to within the current study.

The program environment is another critical design feature that must be considered. The current study found a caring, mastery-oriented climate, in which the youth perceived a sense of belonging to the program and staff as well as leader support, to be effective (Anderson-Butcher et al., 2011; Eccles & Gootman, 2002). Practitioners can create a mastery-oriented climate through integrating goal setting, providing opportunities for skill development, and limiting competition. A sense of belonging can be enhanced through interactive games and exercises in which youth cooperate with their peers and the staff through engaging in the program.

Staff Recruitment and Hiring

When recruiting staff to implement a sport-based PYD program, it is critical to hire staff with competencies in both sport and life skill instruction as the curriculum reflects this dual-focus (Danish et al., 2005; Danish & Hale, 1981; Danish & Nellen,
1997; Gould & Carson, 2008; Hellison, 2003). The results of the current study suggest that implementing social skill curriculum is more difficult than teaching sport skills (Deil-Amen, 2006); as such, social skills may best be learned through observation (Bandura, 1971). Program staff who possess strong social skills can model the targeted social skills and enhance learning. Additionally, practitioners should recruit individuals who have previous experience working with youth as they will be more equipped to effectively handle behavior management and create a positive environment while implementing the intervention. Research has found that program staff who are enthusiastic, confident, and possess the necessary teaching skills are more likely to implement with fidelity (Dusenbury et al., 2003).

**Staff Training**

To enhance implementation fidelity of the program, staff must be trained on the program’s design. This training should incorporate the primary program components (e.g., structure, curriculum, environment) to increase knowledge and understanding of the program design and influence staff attitudes towards the intervention (Dusenbury et al., 2003). For example, the program under investigation provides multiple curriculum demonstrations during staff training to increase understanding of the program design and implementation fidelity. Practitioners may want to provide additional training on social skill instruction, as soft skills have been found more difficult to teach (Deil-Amen, 2006). Overall, the literature suggests that live training, facilitated by experienced program administrators, with detailed, explicit instruction is most effective (Dusenbury et al., 2003).
The current study found program engagement to act as a programmatic mechanism of PYD. To enhance the quality of exposure to an intervention, effective management of youth is critical. Therefore, staff training should incorporate positive behavior management strategies as well as safety concerns and procedures. Lastly, if program staff are expected to complete self-assessments of their implementation of the program as a means of process evaluation or monitoring, staff should first be trained on how to evaluate. More specifically, to enhance the validity of the evaluation and value of the self-reflection, program staff should be trained on the instrument and what constitutes high fidelity, as people often have a tendency to be overly optimistic when self-evaluating (Dunning et al., 2004).

Staff Support

During the implementation of a program, practitioners can positively influence the effectiveness of a program intervention through providing support. Research supports the power of the leader-subordinate dyad in that the behaviors of a program leader can influence staff practices (Herold, 1977). Practitioners can enhance implementation fidelity of program staff by reinforcing the critical components of the program, reviewing the curriculum, and providing additional trainings if necessary. Moreover, practitioners should monitor staff practices and provide feedback, including encouragement and corrective instruction, to reinforce effective staff practices. Another consideration is providing support through assistance and shared responsibility so that the program staff do not feel overwhelmed. Dusenbury and colleagues (2003) suggest that an overwhelming program inhibits implementation fidelity. Therefore, through support,
practitioners can not only meet the needs of program staff but also enhance their program adherence.

**Program Evaluation**

An evidence-based PYD program not only has theoretical justifications for the program design, but demonstrates the outcomes of the program through evaluation (Baldwin et al., 2005). To accurately measure program outcomes, practitioners and researchers must match objective measures with the desired outcomes (Gosen & Washbush, 2004). As youth development is a primary outcome of PYD programs, use of a longitudinal research design is the most effective means of capturing growth over time (Anderson-Butcher et al., 2011; Petitpas et al., 2005; Riley & Anderson-Butcher, 2012; Vierimaa et al., 2012; Weiss et al., 2012). Use of self-reports as a means of measuring youth perceptions and cognitions is supported; however, a multiple informant approach, integrating parent and teacher evaluations of the child, can provide a more comprehensive picture (Danielson & Phelps, 2003). When using self-assessment to measure implementation fidelity, practitioners and researchers should conduct validity checks to ensure responses reflect actual implementation (Dunning et al., 2004).

To measure program dosage, researchers and practitioners should consider integrating multiple dimensions of implementation fidelity (Dusenbury et al., 2003; Fagan et al., 2008; Tucker & Blythe, 2008). From a pragmatic standpoint, a comprehensive assessment will provide greater understanding of the strengths and weaknesses of a program which will inform program planning. In relation to research, process evaluation answers how and why PYD programs are effective, which can extend
our current understanding of organizational effectiveness in sport-based PYD (Petitpas et al., 2005). Only through understanding how a program acts as a mechanism of PYD, can practitioners effectively meet program goals and objectives.

The examination of organizational effectiveness within the context of sport-based PYD is critical due to the significant impact programs can have on vulnerable youth populations. Participation in sport-based PYD programs has consistently been found to enhance PYD (Anderson-Butcher, 2011; Anderson-Butcher et al., 2011; Weiss, 2008; Weiss et al., 2012), promoting success both inside and outside of sport (Gould & Carson, 2008; Papacharisis et al., 2007; Theokas et al., 2008). Sport can be used as a hook to capture youth engagement and reach the underserved population in need of PYD programming (Halpern, 2000; Hartmann, 2003; Hellison, 2000; Li et al., 2007; Pedersen & Seidman, 2005). Research has found that evidence-based programs are more likely to produce positive outcomes (Durlack & Weissberg, 2007), for which the current study contends that an evidence-based program must demonstrate both theoretical and empirical connections between the program components and outcomes to claim effectiveness. Ultimately, a sport-based PYD program that is evidence-based and effective has the power to make a significant positive impact on vulnerable populations through the enhancement of resiliency and PYD, which contributes to lifelong success.

Limitations and Future Directions

While the current study contributes to the examination of organizational effectiveness and PYD, the study design and results have several limitations and therefore should be interpreted with caution. First, to study organizational effectiveness
within the context of sport-based PYD, a single sport-based PYD summer program was
investigated. Therefore, the intervention as well as staff and youth populations were
confined to one context, limiting the external validity of the results. Future research
should test the current study’s models in alternative sport-based PYD contexts to lend
credence to the findings and implications presented in this study.

In consideration of the current study’s research design, the outcomes of a sport-
based PYD intervention were examined without use of a control group. To fully
understand the effectiveness and impact of an intervention, an experimental design using
a control and treatment group should be considered, as a true experiment controls for the
multiple possible threats to internal validity (Bryman, 2012). Similar to longitudinal
research, an experimental design is not heavily used in social research due to the rigor,
time, and cost involved (Bryman, 2012). Future research should consider use of a control
group when examining the learning structure and curriculum of a sport-based PYD
intervention to increase the robustness of the results.

In relation to the longitudinal research design employed in the current study,
program and study attrition was a major limitation (Bryman, 2012). The program under
investigation consisted of a 19 day intervention, for which youth participants completed a
survey at multiple times throughout the intervention (i.e., pre-test and post-test). Of the
539 youth with parental consent to participate in the current study, only 417 completed
both measures (i.e., pre-test and post-test). The loss of participants was due to
participants absent on the day of data collection, dropping out of the program, or being
removed from the program. As the program was voluntary and free of charge, future research may consider ways to incentivize study participation.

To measure program implementation fidelity, the current study developed a self-assessment tool appropriate for the intervention being studied, for which program staff reported on their implementation of the program. The use of self-assessment may limit the accuracy of the results, as this methodological approach has been associated with inflated responses due to people’s unrealistic optimism (Dunning et al., 2004). To address this potential testing threat, the program staff completed multiple trainings on the Session Log tool, including a curriculum demonstration in which the staff practiced evaluating program implementation fidelity (Baumgartner & Hensley, 2006). Additionally, the researchers assigned a research identification number (ID) to each program staff member and identified completed surveys by research ID to protect confidentiality and encourage honest responses. Although self-assessment may influence the accuracy of the data, this methodological approach is more feasible to adopt than observation or ratings from a third party.

Another research design limitation, in relation to the Session Log tool, was the collection of implementation fidelity data confined to structured time. While implementation of the curriculum and lesson plans occurs during structured sessions, program staff continue to contribute to the learning process and environment during unstructured time (e.g., lunch, traveling). Therefore, the implementation fidelity data may not represent a comprehensive picture of the quantity of program dosage. This may account for the small relationships between program implementation and engagement.
Future research should seek to capture program adherence during structured and unstructured time.

The current study had several measurement-related limitations to be considered. The program under investigation is designed to enhance athletic and social competence through the integration of sport and social skill instruction. While the intervention structure subscales measured sport and social skill briefing, reinforcement, and debriefing separately, the curriculum items only focused on use of the social skill curriculum (self-control, effort, teamwork, social responsibility), neglecting the sport skill curriculum as it was only relevant for the recreation leaders. Future research should examine use of a dual-focused intervention curriculum (i.e., sport and life skill instruction) by measuring implementation of sport and social curriculum separately. Although the intervention structure items differentiated sport and social skill instruction, the measure does not tease apart instruction of the specific social skills (i.e., self-control, effort, teamwork, social responsibility). To gain a better understanding of program implementation fidelity as well as the strengths and weaknesses of a program’s design, future research should adopt measures which incorporate and differentiate the multiple competencies taught.

Another measurement-related issue was the appropriate matching of measures with the program’s outcomes (Bohnert et al., 2010; Gosen & Washbush, 2004; Schoel et al., 1988). The Social Sports Experiences Scale was used to measure perceived self-control (Anderson-Butcher et al., 2010). However, the scale is designed to measure both self-control and social responsibility, limiting the content validity of the tool as a measure of self-control alone (Anderson et al., 2011). Future research examining the self-control
social skill should consider use of a different measure which appropriately matches the self-control construct.

Lastly, the use of perceived constructs collected from a single population constituted a measurement limitation. More specifically, to measure youth development only youth provided information on their perceived competencies and to measure program implementation only staff members provided information on their perceived implementation. Future studies should consider examination of these constructs from a multiple informant approach (Danielson & Phelps, 2003). For youth development constructs, parent/guardian and staff perceptions of youth competencies could provide a comprehensive picture of development. Additionally, use of a third party observer to evaluate program implementation could contribute valuable information to staff self-assessment data.

In addition to future research seeking to address the limitations presented in the current study, there are several avenues through which to continue this line of inquiry. The current study did not integrate demographic considerations into the models tested. PYD theorists suggest that youth react differently to interventions as they mature due to changes in individual and environmental conditions (Bohnert et al., 2010; Fraser et al., 2004; Nation et al., 2003). Maturation may not only influence program engagement but also youth development based upon the skills the youth possess entering the program. As a one-size fits all approach does not effectively address group differences, future research should incorporate youth demographics such as age, gender, and new versus returning participants to understand the influence of individual characteristics on program
engagement and youth development. Knowledge of how group differences influence program engagement and youth development will inform suggested staff practices and curriculum development.

Within the current study, program implementation was examined as an antecedent of program engagement, and consisted of implementation of the program structure, curriculum, and environment. As program implementation was found to be a significant antecedent in the athletic competence model and is a necessary component of program dosage and PYD, future research should explore the antecedents of high fidelity within the context of sport-based PYD, such as staff training, program characteristics, staff characteristics, and organizational characteristics (Dusenbury et al., 2003). Understanding antecedents of high fidelity will inform program design, staff recruitment and hiring, training, and support.

Another consideration in relation to the program components implemented is the model of learning adopted. Social learning theory identifies live, verbal instruction, and symbolic modeling as forms of observational learning (Bandura et al., 1961). While the program structure component used the verbal instruction framework, live and symbolic modeling were not captured in the current study. Future research should explore an integration of verbal instruction, live, and symbolic modeling to capture which learning processes are effective for different competencies.

To measure program engagement, the current study examined youth participation and PYD features which entice participation. Quality of youth engagement is suggested to include behavioral, emotional, and cognitive dimensions (Bohnert et al., 2010;
Fredricks et al., 2004; Mahoney et al., 2007; Roth et al., 2010; Weiss et al., 2005), for which the current study focused on behavioral engagement. Future research should incorporate the emotional and cognitive dimensions for a comprehensive picture of quality of engagement. Additionally, researchers should explore the minimum engagement necessary for developmental gains in order to inform program dosage. In relation to the factors which entice youth participation, peer relationships should be considered, as research has found that positive relationships with peers’ influences continued sport engagement and the adoption of pro-social norms (Eccles et al., 2003; Ullrich-French & Smith, 2009). An understanding of the peer influence on youth engagement can inform staff practices and program strategies, such as activity facilitation, group assignment, and behavioral management.

Another dimension of implementation fidelity to consider is that of quality of program delivery. While program implementation and engagement were examined in the current study, reflecting the dimensions of program adherence and participant responsiveness, quality of program delivery was not explored (Dusenbury et al., 2003; Fagan et al., 2008; Tucker & Blythe, 2008). Instruction, facilitation, reinforcement, and engagement may occur during a program intervention; however there is a continuum of how those elements are delivered. A program staff member may reinforce the targeted skills by recognizing peer modeling once or several times throughout the entire activity. The quality of delivery may be a critical factor to explaining the strength of the program intervention, and inform strategies for staff recruitment, training, and support.
Lastly, the current study focused on social and athletic competence development through sport-based PYD due to the primary objectives of the program under investigation. While it is appropriate to examine the developmental outcomes emphasized within the context being studied, there is opportunity to explore how program implementation and engagement influences alternative PYD outcomes. The models of the current study provide valuable insight into how program implementation and engagement operate differently depending on the competency. Therefore, future research should test this framework with alternative competencies to expand the implications and provide valuable insight for diverse sport-based PYD programs.

**Conclusion**

Sport-based PYD emerged as a means of meeting the needs of vulnerable youth through sport (Gould & Carson, 2008; Martinek & Hellison, 1997; Weiss et al., 2012). Organizational effectiveness of sport-based PYD programs is of vital importance due to the influence programs can have on youths’ successful transition to adulthood. Research examining the effectiveness of sport-based PYD programs can inform program planning, guide intervention strategies, improve program practices, and ultimately contribute to PYD. The purpose of the current study was to examine program dosage as a programmatic mechanism of PYD, within the context of sport-based PYD. A sport-based PYD program is considered effective if, when implemented with fidelity, the program achieves its desired outcomes (Dusenbury et al., 2003; Law & Shek, 2011). Therefore, an investigation of the theoretical and empirical connections between program components and outcomes was warranted.
To evaluate program dosage as a programmatic mechanism of PYD, data were collected from staff and youth participants engaged in a sport-based PYD summer program. Program staff completed self-assessments of their implementation of the program components, including program structure, curriculum, and environment. While youth completed self-reports of their perceived competencies and behavioral engagement as well as perceptions of belonging and leader support. Additionally, secondary data of youth attendance was used as a measure of quantity of program dosage. A two-factor framework of program dosage was found to be the best measurement model, reflecting program implementation and engagement. Program implementation was measured by implementation of the program structure and environment, while program engagement was measured by behavioral engagement, belonging, and leader support. Implementation of program curriculum and attendance were found to be non-significant indicators of quantity of program dosage. The developmental outcomes examined included social (general social competence, effort, teamwork, self-control) and athletic competence.

The final two models, differentiated by developmental outcome, specified program engagement as a mediator of PYD and program implementation as an antecedent of program engagement. When comparing the models, program implementation was found to be a significant antecedent of program engagement in the athletic competence model only, while program engagement was found to be a significant mediator of PYD in both models, and the pre-test PYD indicators were found to significantly predict the post-test PYD indicators. Overall, the current study found that four of the five developmental outcomes were achieved (general social competence,
effort, teamwork, athletic competence), for which engagement in the implemented sport-based PYD intervention mediated the developmental growth.

PYD theorists contend that an evidence-based program demonstrates effectiveness through evaluation of program outcomes (Baldwin et al., 2005). The current study extends this notion by integrating process and outcome evaluation to establish empirical connections between exposure to a program intervention (i.e., program dosage) and resulting youth outcomes. The results support the interdependency of program implementation and engagement, for which the effectiveness of program implementation is contingent upon program engagement and engagement is dependent on an implemented intervention. As program implementation was a non-significant predictor and program engagement a stronger mediator in the social competence model, the findings suggest that social skills may be learned best through live modeling from the program staff and peer group (Bandura et al., 1961). Although program implementation was found to be a significant antecedent of program engagement in the athletic competence model, supporting the use of verbal instruction and experiential learning in sport (Bandura et al., 1961; Kolb et al., 2001), program engagement had a moderate mediating influence on PYD, suggesting that the intervention may not have been strong enough to produce the desired effects (Siedentop, 1998).

Overall, the program implementation and engagement dimensions of program dosage were found to be significant programmatic mechanisms of PYD. The findings have implications for program design and evaluation, as well as staff recruitment, training, and support. The current study provides a foundational framework for
examining the interaction of program implementation, engagement, and PYD within the context of sport-based PYD. Future research should seek to build off the findings presented to further explore how programs can effectively develop youth through sport.
References


Amorose, A.J. (2002). The influence of reflected appraisals on middle school and high school athletes’ self-perceptions of sport competence. *Pediatric Exercise Science,


of youth development: The significance of supportive staff-youth relationships. 

*The Journal of Human Behavior in the Social Environment, 9*(1/2), 83-99. (co-published as a book chapter simultaneously in *How Institutions are Shaping the Future of our Children*).


Psychology, 31(1), 39-55.


models: From theory to outcomes. In P.A. Witt & L. L. Caldwell (Eds.),


Diego, CA: Naval Personnel Research Center.


Community and Youth Collaborative Institute (CAYCI). (2010). *The teamwork scale*. 313
Columbus, OH: College of Social Work, Ohio State University.


    Ringel & J. Brandell (Eds.), *Trauma: Contemporary directions in theory,

    adolescents: Scale development and educational correlates. *Psychology in the
    Schools, 30*, 79-90.


    learning effectiveness. *Simulation & Gaming, 35*(2), 270-293.

    experimental evaluation of the All Stars prevention curriculum in a community

Gottfredson, D.C., & Gottfredson, G.D. (2002). Quality of school-based prevention
    programs: Results from a national survey. *Journal of Research in Crime and
    Delinquency, 39*, 3-35.

    future directions. *International Review of Sport and Exercise Psychology, 1*(1),
    58-78.

    we’ve been, where we are, and where we should go. *Education & Treatment of


Hellison, D. (2003). Teaching responsibility through physical activity (2nd ed.).
Champaign, IL: Human Kinetics.


Education and Sport, Waterville Valley, NH.


National Collaboration for Youth Members (1998). What is youth development?


Mahwah: Lawrence Erlbaum Associates.


Riley, A.B. (2013). *Staff practices and youth outcomes in a summer sport-based positive youth development program* (Unpublished doctoral dissertation). The Ohio State University, Columbus, OH.


physical activity. *Medicine & Science in Sports & Exercise, 39* (9), 1493-1500.


Appendix A: Staff Instrument
### Session Log

**Session Type:** Recreation / Chalk Talk

<table>
<thead>
<tr>
<th>Component</th>
<th>Degree of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Framing (Before Activity)</td>
<td></td>
</tr>
<tr>
<td>Staff overviewed today's S.E.T.S.</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff overviewed today's sport skill</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff gave instructions of sport cues</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff taught how to use S.E.T.S. in activity</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff gave instructions for activity</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Facilitating (During Activity)</td>
<td></td>
</tr>
<tr>
<td>Staff provided opportunity for youth to practice sport skill</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff praised campers when they demonstrated good sport skill</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff praised campers when they demonstrated good S.E.T.S.</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff pointed out use of S.E.T.S. in activity</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Debriefing (After Activity)</td>
<td></td>
</tr>
<tr>
<td>Staff reviewed today's S.E.T.S.</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff reviewed today's sport skill</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff discussed using S.E.T.S. outside of camp</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff engaged youth in debriefing process</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
</tbody>
</table>

**Group #:**

- Organization: Groups combined / Groups separated / Groups rotated

<table>
<thead>
<tr>
<th>Component</th>
<th>Degree of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Curriculum</td>
<td>None  Total</td>
</tr>
<tr>
<td>Staff emphasized self-control</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff emphasized effort</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff emphasized teamwork</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff emphasized social responsibility</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Interaction with Youth</td>
<td></td>
</tr>
<tr>
<td>Staff created an emotionally safe climate</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff greeted youth as they arrived</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff used warm tone of voice</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff encouraged youth to try new skills</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff encouraged participation by all youth</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff provided opportunity for youth to make at least one choice</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
<tr>
<td>Staff kept youth on task</td>
<td>1 2 3 4 5 6 7 8 9 10</td>
</tr>
</tbody>
</table>
Appendix B: Youth Pre-Test Instrument

2014 Life Sports Social Competence Education Survey
Instructions: Fill in each circle completely using the following example:
Like this:  ●  Not like this: ✓  ◯  ⊘

**What You Do.** Please read each statement and decide if the statement is “not at all true” for you, “a little true” for you, “somewhat true” for you, “pretty true” for you, or “really true” for you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all true</th>
<th>A little true</th>
<th>Somewhat true</th>
<th>Pretty true</th>
<th>Really true</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I help other people.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>2. I ask others if I can be of help.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>3. I show concern for others.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>4. I show care for others.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>5. I give support to others.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
**Teamwork in Sports.** Please read each statement and decide if the statement is “not at all true” for you, “a little true” for you, “somewhat true” for you, “pretty true” for you, or “really true” for you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all true</th>
<th>A little true</th>
<th>Somewhat true</th>
<th>Pretty true</th>
<th>Really true</th>
</tr>
</thead>
<tbody>
<tr>
<td>When playing sports...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. I feel confident in my ability to work in a team.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>2. I know how to give my group members feedback that will not hurt their feelings.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>3. I ask others for feedback.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>4. I make an effort to include other members of my group.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>5. I value the contributions of my group members.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>6. I treat my group members as equal members of the team.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>7. I am good at communicating with my group members.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>8. I feel confident in my ability to be a leader.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

**Your Actions and Behavior in Sports.** Please read each statement and decide if the statement is “not at all true” for you, “a little true” for you, “somewhat true” for you, “pretty true” for you, or “really true” for you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all true</th>
<th>A little true</th>
<th>Somewhat true</th>
<th>Pretty true</th>
<th>Really true</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I respect others when playing sports.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>2. I work well with my group members when playing sports.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>3. I control my temper when playing sports.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>4. I play sports fairly even when an adult is not around.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
5. I am good at working together with my group members when playing sports.  
6. I take responsibility for my actions when playing sports.  
7. I act responsibly when playing sports.  
8. I cooperate with others when playing sports.

**Effort in Sport.** Please read each statement and decide how well the statement “corresponds,” or describes, you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Doesn’t correspond to me at all</th>
<th>Corresponds to me a little</th>
<th>Corresponds to me partly</th>
<th>Corresponds to me a lot</th>
<th>Corresponds to me exactly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. In competition, I go all out even if I’m almost sure to lose.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>2. I don’t give up even after making many mistakes.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>3. I think about ways to improve my weaknesses.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>4. It is important to me to be present at all practices.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>5. During practices, I go all out.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

**Your Ability in Sport.** Some kids think they are pretty good at sport but other kids think they are not so good. Please fill in the circle that best reflects how you feel about your sports ability.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not good at all</th>
<th>Not good</th>
<th>So-so</th>
<th>Good</th>
<th>Very good</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How good do you think you are at sports?</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td>Not skilled at all</td>
<td>Not skilled</td>
<td>Sort of skilled</td>
<td>Skilled</td>
<td>Very skilled</td>
</tr>
<tr>
<td>---</td>
<td>-------------------</td>
<td>-------------</td>
<td>-----------------</td>
<td>---------</td>
<td>--------------</td>
</tr>
<tr>
<td>2.</td>
<td>How <strong>skilled</strong> do you think you are at sports?</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Not much ability at all</th>
<th>Not much ability</th>
<th>Some ability</th>
<th>A lot of ability</th>
<th>A whole lot of ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.</td>
<td>How much <strong>ability</strong> do you think you have when it comes to sports?</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Appendix C: Youth Post-Test Instrument

2014 LiFE Sports Social Competence Education Survey

Instructions: Fill in each circle completely using the following example:

Like this: ☐  Not like this: ☑  ❌  ☐

Belonging. Please read each statement and decide if the statement is “not at all true” for you, “a little true” for you, “somewhat true” for you, “pretty true” for you, or “really true” for you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all true</th>
<th>A little true</th>
<th>Somewhat true</th>
<th>Pretty true</th>
<th>Really true</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I feel comfortable with people at LiFE Sports.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>2. I am part of LiFE Sports.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>3. I am committed to LiFE Sports.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>4. I am supported at LiFE Sports.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>5. I am accepted at LiFE Sports.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>6. I feel safe at LiFE Sports.</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
Relatedness and Leader Support. Think about your LiFE Sports Group Coach, the leader of your group. Please read each statement and decide if the statement is “not at all true” for you, “a little true” for you, “somewhat true” for you, “pretty true” for you, or “really true” for you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all true</th>
<th>A little true</th>
<th>Somewhat true</th>
<th>Pretty true</th>
<th>Really true</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I do not wish I were in a different LiFE Sports group.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
<td>O</td>
</tr>
<tr>
<td>2. My Group Coach is interested in people like me.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>3. My Group Coach respects me.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>4. My Group Coach thinks I am interesting.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>5. I can talk to my Group Coach if I have a problem.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>6. I wish I were in a different LiFE Sports group.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>7. My Group Coach is not interested in people like me.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>8. I am treated with as much respect as the other kids in my LiFE Sports group.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

What You Do. Please read each statement and decide if the statement is “not at all true” for you, “a little true” for you, “somewhat true” for you, “pretty true” for you, or “really true” for you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all true</th>
<th>A little true</th>
<th>Somewhat true</th>
<th>Pretty true</th>
<th>Really true</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. I help other people.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>7. I ask others if I can be of help.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>8. I show concern for others.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>9. I show care for others.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>10. I give support to others.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
**Teamwork in Sports**. Please read each statement and decide if the statement is “not at all true” for you, “a little true” for you, “somewhat true” for you, “pretty true” for you, or “really true” for you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all true</th>
<th>A little true</th>
<th>Somewhat true</th>
<th>Pretty true</th>
<th>Really true</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. I feel confident in my ability to work in a team.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>10. I know how to give my group members feedback that will not hurt their feelings.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>11. I ask others for feedback.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>12. I make an effort to include other members of my group.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>13. I value the contributions of my group members.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>14. I treat my group members as equal members of the team.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>15. I am good at communicating with my group members.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>16. I feel confident in my ability to be a leader.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

**Your Actions and Behavior in Sports**. Please read each statement and decide if the statement is “not at all true” for you, “a little true” for you, “somewhat true” for you, “pretty true” for you, or “really true” for you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all true</th>
<th>A little true</th>
<th>Somewhat true</th>
<th>Pretty true</th>
<th>Really true</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. I respect others when playing sports.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>10. I work well with my group members when playing sports.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>11. I control my temper when playing sports.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>12. I play sports fairly even when an adult is not around.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Engagement. Please read each statement and decide if the statement is “not at all true” for you, “a little true” for you, “somewhat true” for you, “pretty true” for you, or “really true” for you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not at all true</th>
<th>A little true</th>
<th>Somewhat true</th>
<th>Pretty true</th>
<th>Really true</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. I am good at working together with my group members when playing sports.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>14. I take responsibility for my actions when playing sports.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>15. I act responsibly when playing sports.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>16. I cooperate with others when playing sports.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Effort in Sports. Please read each statement and decide how well the statement “corresponds,” or describes, you. Fill in the circle that best shows your answer.

<table>
<thead>
<tr>
<th>Question</th>
<th>Doesn’t correspond to me at all</th>
<th>Corresponds to me a little</th>
<th>Corresponds to me partly</th>
<th>Corresponds to me a lot</th>
<th>Corresponds to me exactly</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. In competition, I go all out even if I’m almost sure to lose.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>7. I don’t give up even after making many mistakes.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Question</td>
<td>Doesn’t correspond to me at all</td>
<td>Corresponds to me a little</td>
<td>Corresponds to me partly</td>
<td>Corresponds to me a lot</td>
<td>Corresponds to me exactly</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>---------------------------------</td>
<td>---------------------------</td>
<td>--------------------------</td>
<td>------------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>8. I think about ways to improve my weaknesses.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>9. It is important to me to be present at all practices.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>10. During practices, I go all out.</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

**Your Ability in Sport.** Some kids think they are pretty good at sport but other kids think they are not so good. Please fill in the circle that best reflects how you feel about your sports ability.

<table>
<thead>
<tr>
<th>Question</th>
<th>Not good at all</th>
<th>Not good</th>
<th>So-so</th>
<th>Good</th>
<th>Very good</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. How good do you think you are at sports?</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question</th>
<th>Not skilled at all</th>
<th>Not skilled</th>
<th>Sort of skilled</th>
<th>Skilled</th>
<th>Very skilled</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. How skilled do you think you are at sports?</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question</th>
<th>Not much ability at all</th>
<th>Not much ability</th>
<th>Some ability</th>
<th>A lot of ability</th>
<th>A whole lot of ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. How much ability do you think you have when it comes to sports?</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Appendix D: Session Log Training Handout

Session Log Tips

• Pick-up / Drop-off
  o Pick up your session logs from Lee and Mickey before leaving the ice rink every Morning
  o At the end of the day drop off your completed session logs to Lee and Mickey before going to your bus or leaving

• Timing
  o Complete your logs immediately after each session (before rotating)

• Preliminary information
  o Circle the type of activity session the log corresponds with (Recreation or Chalk Talk)
  o Write down the (your) group number present at the session you are evaluating
  o Circle the group organization of the session (groups combined, groups separated, groups rotated)
    ▪ If groups are combined or rotated, include both group numbers

• Evaluation
  o Evaluate the implementation of the entire session (not just your own implementation)
  o Respond to every question
  o If you would like to provide immediate feedback, please write in the comments section on the back of the session log

• Honesty
  o Critically consider each question and make sure your responses accurately reflect what occurred during the session
  o We want to receive your honest feedback (assists with future program planning)

• Clarity

347
Circle your responses clearly so that our research staff can identify the correct numbers
Write the %ages clearly so that our research staff can identify the correct numbers

• Terminology
  - Overview: Introduce the day’s sport skill and/or social competence
  - Sport cues: Key terms associated with sport skills (e.g., dribbling with your fingertips)
  - S.E.T.S.
    - Within framing and debriefing - S.E.T.S. refers to the day’s social competency
    - Within facilitating - S.E.T.S. refers to any of the four social competencies
  - Self-control: The ability to appropriately manage and exhibit control of your own actions and behaviors
  - Effort: Self-directed behavior about how hard you are willing to work and the extent of your exertion
  - Teamwork: The ability to collaborate and work with others in order to achieve a common goal
  - Social responsibility: Your ability to think about how your actions and behaviors will ultimately affect others
  - Framing: Overviewing the day’s activities and goals
  - Facilitating: Guiding youth to develop sport skill and social competence (includes demonstration, reinforcement, direct feedback)
  - Debriefing: Reviewing what happened in the activity and discussing how S.E.T.S. can be used in other contexts
  - Emphasizing: Identifying, highlighting, and reinforcing the social competencies that are occurring
  - Emotionally safe climate: Youth feel safe and secure in the environment created; youth are able to ask questions and not afraid to try new things

348
- **Warm tone of voice**: Speaking voice is pleasant, calm, and polite
- **On task**: Youth are participating in the activity and paying attention during instruction
- **Fully engaged**: Youth are concentrated on the activity
Appendix E: Parental Consent Form

The Ohio State University Parental Permission
For Child’s Participation in Research

Study Title: Examining the Impacts of a Summer Youth Development Program
Researcher: Dawn Anderson-Butcher, Ph.D.
Sponsor:

This is a parental permission form for research participation. It contains important information about this study and what to expect if you permit your child to participate.

Your child’s participation is voluntary.
Please consider the information carefully. Feel free to discuss the study with your friends and family and to ask questions before making your decision whether or not to permit your child to participate. If you permit your child to participate, you will be asked to sign this form and will receive a copy of the form.

Purpose:
We suspect that children involved in “LiFE Sports” enhance their sport skills and have better social and personal skills. We want to understand whether your child has these types of benefits from his/her coming to the program. To do this, we want to ask your child to do a survey at the beginning and the end of the program. The survey will have questions on it that looks at their experiences, feelings, and thoughts related to themselves and the program. We will also get attendance, registration, social competence education, and health and fitness information about your child from the LiFE Sports camp records.

Procedures/Tasks/Duration:
We need your permission to let your child do this survey and to let us get information from the LiFE Sports camp. If you agree, your child will answer some questions at the
beginning of the program, which will take about 30-40 minutes. In addition, your child also will answer some questions at the end of the program, which also will take about 30-40 minutes. The total time required throughout the program for your child to answer all of these questions will be approximately 60-80 minutes. Your child will be permitted to answer the questions at both the beginning and end of the program over the course of several days in order to accommodate each child’s ability to understand and answer the questions.

**Participant Rights:**
You get to choose if you want your child to do this. That means you don’t have to have your child participate if you don’t want to. Nothing will happen if you decide not to have your child participate. He/she will still have the same opportunity to be involved in “LiFE Sports.”

**Confidentiality:**
Anything your child says will not be tracked back to him/her. What that means is that we won’t say “Jamal said this or that.” The information will be put together so that it says something like “the kids said this or that.” Efforts will be made to keep your child’s study-related information confidential. However, there may be circumstances where this information must be released. For example, personal information regarding your child’s participation in this study may be disclosed if required by state law. Also, your child’s records may be reviewed by the following groups (as applicable to the research):

- Office for Human Research Protections or other federal, state, or international regulatory agencies;
- The Ohio State University Institutional Review Board or Office of Responsible Research Practices;
- The sponsor, if any, or agency (including the Food and Drug Administration for FDA-regulated research) supporting the study.

**Risks and Benefits:**
We don’t think there are any bad things that could happen because your child answers these questions. One benefit, however, is that LiFE Sports leaders can use the information you give them to make the program better.

If you have any questions about this research or if you feel your child has been harmed by study participation, please contact Dawn Anderson-Butcher at (614) 292-6934.

For questions about your child’s rights as a participant in this study or to discuss other study-related concerns or complaints with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6251.
Signing the parental permission form

I have read (or someone has read to me) this form and I am aware that I am being asked to provide permission for my child to participate in a research study. I have had the opportunity to ask questions and have had them answered to my satisfaction. I voluntarily agree to permit my child to participate in this study.

I am not giving up any legal rights by signing this form. I will be given a copy of this form.

Printed name of subject

Printed name of person authorized to provide permission for subject      Signature of person authorized to provide permission for subject

Relationship to the subject

Date and time

AM/PM

Investigator/Research Staff

I have explained the research to the participant or his/her representative before requesting the signature(s) above. There are no blanks in this document. A copy of this form has been given to the participant or his/her representative.

Printed name of person obtaining consent      Signature of person obtaining consent

Date and time

AM/PM
The Ohio State University Assent to Participate in Research

Study Title: Examining the Impacts of a Summer Youth Development Program
Researcher: Dawn Anderson-Butcher, Ph.D.
Sponsor:

1. What is this study about?

We are interested in looking at the benefits of your coming to LiFE Sports. We want your help and want you to find out what you like, don’t like, and what happens in general at the program.

2. What will I need to do if I am in this study?

If you agree to help us, you will answer some questions on a survey now and at the end of the program. We will also get attendance, registration, social competence education, and health and fitness information about you from the LiFE Sports camp records.

3. How long will I be in the study?

It will take about 30-40 minutes to answer these questions now and at the end of the session. So, in total, you are going to be answering questions for about 60-80 minutes during the program. Because some children might need more time reading and understanding questions, which is ok, you might be answering all of these questions over several days.
4. Can I stop being in the study?

You may stop being in the study at any time. In other words, you get to choose if you want to do this. That means you don’t have to participate if you don’t want to. Also, you don’t have to answer any of the questions that you don’t want to. Or if you decide you don’t want to participate anymore, you can just hand the survey back in with no answers. Nothing bad will happen if you decide not to participate. You will just wait until the others have finished and then go back to participating in the program.

Also, anything you answer will not be tracked back to you specifically. What that means is that I can’t say “Jamal said this or that.” The information will be put together so that when reported it will say something like “the kids said this or that.”

5. What bad things might happen to me if I am in the study?

We don’t think there are any bad things that could happen because you answer these questions.

6. What good things might happen to me if I am in the study?

One good thing that might happen if you participate is that leaders at this program can use the information you give them to make the program better for you.

7. Will I be given anything for being in this study?

You will not be given anything for being in this study.

8. Who can I talk to about the study?

For questions about the study you may contact Dawn Anderson-Butcher in the College of Social Work at 614-292-6934.

To discuss other study-related questions with someone who is not part of the research team, you may contact Ms. Sandra Meadows in the Office of Responsible Research Practices at 1-800-678-6251.
Signing the assent form

I have read (or someone has read to me) this form. I have had a chance to ask questions before making up my mind. I want to be in this research study.

______________________________  ______________________________ AM/PM
Signature or printed name of subject               Date and time

Investigator/Research Staff

I have explained the research to the participant before requesting the signature above. There are no blanks in this document. A copy of this form has been given to the participant or his/her representative.

______________________________  ______________________________ AM/PM
Printed name of person obtaining assent               Signature of person obtaining assent

______________________________  ______________________________ AM/PM
Date and time

This form must be accompanied by an IRB approved parental permission form signed by a parent/guardian.