

# **Research and Practice: An Applied Examination of the Parental Perceptions of Developmental Outcomes Scale**

**Ryan J. Gagnon**  
**Barry A. Garst**  
Clemson University

## **Abstract**

To support practitioners' need for reliable and valid measures of the youth camp experience, this study compares the results of two analytic approaches—(1) composite-based paired samples *t* tests and (2) a latent structural equation model approach—using the Parental Perceptions of Developmental Outcomes (PPDO) scale in a sample of 930 parents of youth attending residential summer camps. Through a retrospective pretest (RPT) design, parents reported statistically significant growth across the five dimensions of the PPDO (responsibility, exploration, self-regulation, attitude, and communication) in both analytic approaches. The study results indicated similar effect sizes and conclusions were reached using both analytical approaches, suggesting a more practitioner-friendly approach (composite) may tell the same story as a more complex and research-oriented (latent) approach. This finding, as well as the revalidation of the PPDO in this sample, indicates the PPDO's usefulness in applied and research contexts.

**KEYWORDS:** camp; youth development; applied multivariate research; parents; Parental Perceptions of Developmental Outcomes scale

Many out-of-school-time (OST), outdoor, and recreation-centered organizations aim to provide the best available tools, resources, and context necessary to ensure the youth they serve progress in their socioemotional, physical, and cognitive development. These organizations must examine both *how* and *if* their desired outcomes are achieved and identify areas for improvement. In this pursuit toward evidence-based practice, organizations such as summer camps experience escalating pressure from both internal (e.g., program staff and participants) and external stakeholders (e.g., accreditors and funders) to demonstrate program efficacy using the best available methods (Bialeschki & Sibthorp, 2011; Preskill & Boyle, 2008). Consequently, it is no longer sufficient for youth-serving organizations to say they are achieving their desired goals—organizations now must articulate if and how success was achieved. Such an explanation regularly includes assessment processes used for measuring progression toward targeted goals, the rationale for specific evaluation approaches, how conclusions of success were reached, and (arguably more important) where the organization fell short in the pursuit of its desired goals (Powers, Maley, Purington, Schantz, & Dotterweich, 2015).

In response to the increasing expectations for documentation of program impact and opportunities for program and organizational improvement, recreation practitioners, researchers, and evaluators have noted the need for reliable and valid measures of assessing both parent and child perceptions of program outcomes (Sibthorp, Bialeschki, Morgan, & Browne, 2013). Correspondingly, an increase in the methodological rigor of research and evaluation in OST, outdoor, and recreation-centered programs is occurring, where increasingly sophisticated methodological techniques are becoming normative (Roth & Brooks-Gunn, 2016). For instance, studies using structural equation modeling (Garst & Gagnon, 2016), instrumental case studies (Povilaitis & Tamminen, 2017), and photovoice elicitation (Bultas, Steurer, Balakas, Brooks, & Fields, 2015) have emerged within the camp research literature. These studies have advanced understanding of programmatic and contextual factors influencing developmental outcomes in youth campers. Paradoxically, elevation in the sophistication of camp research has potentially furthered the divide between researchers and practitioners, as the need for ongoing training in evaluation methods, data analysis, and data visualization techniques can make certain evaluation and assessment techniques unobtainable for practitioners with limited time and resources.

These trends, the need for psychometrically valid instruments appropriate for use in OST, outdoor, and recreation-centered settings, and the related gap between the analytic capabilities of researchers and practitioners, provide the context for this study. Specifically, this study compares two methodological approaches to assess parent observations of youth socioemotional skill development resulting from their child's camp participation. In the first, more practitioner-centered approach, this study used a paired sample *t*-test technique to determine if parents perceived meaningful growth before and after camp across five dimensions of the Parental Perceptions of Developmental Outcomes (PPDO) scale (Garst & Gagnon, 2016). In the second more technically sophisticated approach, the study utilized a structural equation model (SEM) to determine if parents perceived meaningful growth in the same five dimensions of the PPDO before and after their child's camp experience. Recognizing that if similar results are found despite the difference in approach, then this study could provide evidence of the usefulness of the PPDO for less computationally demanding analyses by practitioners (e.g., paired sample *t* tests), and therefore provide a more useful and practitioner-friendly measure of parent perceptions of their child's socioemotional development in a camp setting.

## **Camp as a Developmental Context: Outcomes and Measurement**

As a noted developmental setting, camp has been demonstrated as impactful to children and adolescents due to the duration, intensity, and novelty of the experience (Henderson, Bialeschki, et al., 2007). A broad base of evidence highlights potential outcomes for youth associated with camp including *resilience* (Merryman, Mezei, Bush, & Weinstein, 2012), *leadership develop-*

ment (Kendellen, Camiré, Bean, & Forneris, 2016), *career readiness* (Bhattacharyya, Mead, & Nathaniel, 2011), *physical activity* (Hickerson & Henderson, 2013), and *environmental awareness* (Collado, Staats, & Corraliza, 2013). Further investigations have supported camp as an environment for nurturing socioemotional skills in youth (Thurber, Scanlin, Scheuler, & Henderson, 2007). This socioemotional skill-centered research indicates camp can positively impact a youth's *personal responsibility* (Garst & Gagnon, 2016), *motivation to explore* (Thurber et al., 2007), *attitude* (Gillard & Watts, 2013), and *communication skills* (Clary & Ferrari, 2015).

In addition to studies of the impact of camp experiences from the youth perspective, research has also examined parents' perceptions and observations of their children following camp (Henderson, Whitaker, Bialeschki, Scanlin, & Thurber, 2007), and these studies have yielded comparable results to youth-focused studies. For example, Henderson, Whitaker, et al. (2007) found parents noted the greatest precamp/postincreases in their child's development (with modest effect sizes) related to positive identity development, independence, peer relationships, and exploration/adventure. Similarly, Michalski, Mishna, Worthington, and Cummings (2003) noted that parents associated camp participation with positive changes in child cooperation, responsibility, and self-control levels. In another study, Baughman, Garst, and Fuhrman (2009) found that parents viewed their children as more likely to take care of personal belongings and to share work/chore responsibilities due to their camp experience.

As illustrated, literature exploring outcomes associated with the camp experience is variable yet comprehensive, reflecting a broad array of outcomes and/or skills. This contrast reflects the differentiation in camp vision and populations served. Indeed, in an extensive study of outcomes associated with camp experiences, the American Camp Association (ACA, 2005) demonstrated wide variation in camp ownership/sponsorship (e.g., for-profit vs. nonprofit, day vs. residential, single-gender vs. coeducational) and camp outcomes, with at least 10 outcomes directly resulting from the camp experience (e.g., self-esteem, leadership, spirituality). This variability is reflected in the numerous measures currently used for the assessment of camp outcomes. For instance, the scale development investigation underpinning the Youth Outcomes Battery<sup>1</sup> (YOB) conducted by Sibthorp et al. (2013) illustrated six unique outcomes associated with the camp experience: family citizenship behavior, competence, responsibility, independence, problem solving, and teamwork. While Sibthorp et al. (2013) designed the YOB specifically for camp outcome assessment and research, E. Hill et al. (2015) modified existing measures reflecting self-determination theory (SDT) to understand outcomes associated with medical specialty camp attendance. Beyond measures developed specifically for measuring holistic outcomes of camp experiences (e.g., Camper Growth Index by Henderson, Thurber, Whitaker, Bialeschki, & Scanlin, 2006; YOB by Sibthorp et al., 2013) or adapted for the study of specific camp populations (e.g., E. Hill et al., 2015), numerous other measures have been used for examining camp-specific outcomes (Bobzien & Judge, 2014; Williams, Ma, Prejean, Ford, & Li, 2007). As a result, no widely used common measures have emerged for assessing youth and/or parent perceptions of outcomes of the camp experiences.

This differentiation presents a challenge for those seeking to measure global outcomes associated with the camp experience (Lewis et al., 2018; Sibthorp et al., 2013) in a way that also considers the local conditions necessary for utilization-focused program evaluation (Patton, 2008). While global measures of outcomes of the camp experience may have the advantage of allowing for macro-level comparison (e.g., how Camp A compares to Camp B), they may fail to capture the unique focus of a specific camp with a specific mission addressing a few targeted outcomes.

The wide array of outcome measures available in combination with differentiation across programmatic mission and focus represents a challenge to those evaluating and researching camp experiences. For example, a foundational element underpinning outcomes measurement

---

<sup>1</sup>As of the production of this manuscript, peer-reviewed reports of the validity and reliability of the ACA Staff and Parent Perceptions YOB were unavailable (ACA, 2018).

in camp relates to the psychometric properties of the measures implemented where camp evaluators must have evidence of the validity of how selected measures reflect desired outcomes (i.e., convergent, discriminant, construct, and criterion validity; Brown, 2015) and how the measures perform in applied settings. However, the support for many current measures and the techniques underpinning their “validity” are somewhat suspect. Specifically, many measures of outcomes associated with the broader OST field reflect outdated techniques for justifying their validity (Gagnon, Stone, & Garst, 2017) and can result in a misalignment of what the scales purport to measure when compared with what they actually measure. For instance, a common approach is the utilization of composite scores for assessing camp-related outcomes; indeed, the author team has utilized similar techniques within OST and camp settings (Gagnon & Bumpus, 2016; Garst, Baughman, Franz, & Seidel, 2013). Further, escalating expectations for methodological rigor within the broader social sciences have highlighted limitations of composite-score-based validation approaches (Freire & Caldwell, 2013; Gagnon et al., 2017). The fundamental challenge embedded within composite-score approaches (i.e., summing the items composing a subscale and dividing this sum by the number of items to create a composite score) is the assumption that items composing the composite are measured error-free (Gagnon et al., 2017), whereas more contemporary approaches such as confirmatory factor analyses (CFA) do not assume perfect measurement (Brown, 2015). Further, within the few camp-related studies using CFA, utilization of the corresponding findings is difficult if not impossible in the practical sense (e.g., for site-specific program evaluation) due to the level of training and resources necessary for conducting these analyses. For instance, a default approach recommended by the American Psychological Association (Appelbaum et al., 2018) is to consider the convergent and discriminant validity of measures prior to relationship testing (e.g., precamp to postcamp score change; Appelbaum et al., 2018). More basic techniques such as exploratory factor analysis (EFA) do not readily facilitate these statistics without the utilization of composite-based approaches for tests beyond those examining factor structure. As a relatively exorbitant level of training and resources is necessary for conducting these more comprehensive analyses (e.g., CFA; Brown, 2015) than is typically available to camp program evaluators, one mitigation strategy is for camp program evaluators to determine the performance of identical camp outcome measures, utilizing composite methods as compared to latent methods (i.e., CFA). In effect, this comparative approach represents a logical progression of the latent techniques presented in prior studies of camp research (Garst & Gagnon, 2016; Sibthorp et al., 2013). Specifically, if similar conclusions are reached, despite the differentiation in approach (e.g., latent analyses vs. composite analyses), both regarding the psychometric properties of the scale (where comparison is statistically available), as well as the effect sizes of the relationships being explored (e.g., change from precamp to postcamp; Cohen's  $d$ ), then the evaluated scale could serve as a useful tool for those interested in applying the scale with the use of practitioner-friendly composite techniques.

## Study Purpose and Contribution

The literature associated with camp experiences provides compelling evidence—from the perspective of youth and parents—for camp as a positive developmental experience for youth across a range of desirable outcomes. However, the training and resources necessary for camp practitioners to evaluate the impact of the camp experience on youth outcomes are becoming increasingly complex. To support the need for reliable and valid measures that require less training for practitioners to analyze and interpret, this study compares the results of two analytic approaches investigating the same research question: *Do parents of campers report socioemotional change across the five dimensions of the PPDO precamp to postcamp?* Prior research utilizing the PPDO indicates that it is a psychometrically reliable and valid instrument for programs targeting the outcomes presented in Table 1 (Garst & Gagnon, 2016). Further, the work of Garst and Gagnon (2016) also suggests across a large sample ( $N = 2,952$ ), parents generally reported

statistically significant ( $p \leq .05$ ) growth in factors composing the PPDO following the camp experience. As such, this study first wants to confirm the findings of Garst and Gagnon (2016) by answering the first hypothesis ( $H_1$ ), *parents will observe and report statistically significant ( $p \leq .05$ ) growth across the five dimensions of the PPDO*.

Beyond the examination of youth change resulting from the camp experience, this study has the parallel intent to determine how the PPDO performs in the use of two distinct analytic techniques. Specifically, this study compares findings resulting from a composite-based paired samples  $t$ -test approach to findings using a latent structural equation model (i.e., latent paired samples  $t$  tests). Research by Coman et al. (2013) indicates that latent analytic approaches for determining growth across two time points should be similar to composite-based techniques. Thus, given the likely parallel findings despite the differing analytic approach, the second hypothesis ( $H_2$ ) of the current study is, *the two analytic approaches will yield the same findings (i.e., significance) and effect sizes (i.e., Cohen's  $d$ )*. If parallel conclusions are reached despite the differing analytical complexity, then these findings may support the PPDO as a promising measure for practitioners to assess changes in youth outcomes resulting from the camp experience.

## Method

### Study Setting

As part of a larger multiorganization effort exploring the developmental outcomes of the camp experience, this study was conducted with two not-for-profit organizations operating 18 residential summer camps within two states, one in the northeastern United States (Organization Northeast; 12 camps) and the other in the southeastern United States (Organization Southeast; 6 camps). These organizations were selected for this study due to their program assessment experience and well-established rapport with camper parents. Both organizations offered residential coeducational camp sessions lasting 5 to 7 days. Organization Northeast and Southeast targeted the same developmental outcomes (i.e., responsibility, exploration, self-regulation, attitude, and communication) through the provision of youth activities such as shooting sports, robotics, outdoor skills, and environmental sciences.

### Participants and Procedures

Data were collected in the summer of 2015 from parents of campers attending one of the 18 camps, through the e-mail address parents used to register their child for camp. Specifically, 7 days after their child's camp experience, parents were sent an e-mail from their corresponding camp with a study description, instructions on how to participate in the study, and a link to the online questionnaire, which was administered through Qualtrics software. Nonresponsive parents were sent a reminder e-mail 7 days later (14 days after camp completion). As an incentive to participate, upon completion of the questionnaire parents could enter a raffle to receive one of six \$100.00 gift cards.

The combination of two e-mails requesting participation and the gift card raffles resulted in an overall response rate of 25.71% to the questionnaire (930 parent respondents  $\div$  3,617 parents who received the link = 25.71%). A power analysis for determining the minimum sample size for the exploration of predictive relationships precamp to postcamp indicated a sample of 167 respondents ( $R^2 = .10$ ,  $p \leq .01$ ,  $\alpha = .99$ ,  $\lambda = 17.87$ , 5 independent variables) would be sufficient for the analyses of interest (Cohen, Cohen, West, & Aiken, 2003). Parent respondents tended to be highly educated (65% with a bachelor's degree or above), female (89%), White (82.8%), and married/never divorced (79.3%). Annual respondent income averaged \$104,637.45 ( $SD = \$63,079.48$ ) with 63.2% of respondents reporting a yearly income at or above \$87,500.00. Campers tended to be White (84.6%) and female (50.8%) and were on average 11.96 years old

( $SD = 2.30$  years). Outcomes targeted by both organizations were matched to the factors composing the PPDO. This study replicated the measurement approach presented in Garst and Gagnon (2016), who used a retrospective pretest design “where parents were asked to reflect on their child’s behavior prior to attending camp and following the camp experience” (p. 57). In this approach, parents were presented with the item “my child . . . participates in a discussion” and then asked to share their observation of their child’s ability level on a 1 (*strongly disagree*) to 5 (*strongly agree*) Likert-style scale, rating their child both before and after camp. Retrospective pretest (RPT) designs have been identified as helpful for reducing concerns associated with incomplete data (Raidl et al., 2004), financial and implementation challenges associated with administering multiple measures (Chang & Little, 2018), and response shift bias (i.e., when respondents’ personal understanding of a construct may change over time; Chang & Little, 2018; Sibthorp, Paisley, Gookin, & Ward, 2007), which has been alternately described as a situation in which respondents “don’t know what [they] don’t know” (Browne, 2018, para. 2). Such a response shift occurs in a traditional pretest–posttest design because participants have limited knowledge or experience with which to assess their preprogram perceptions or behaviors, and this bias can have a negative influence on outcome measures by making it appear that the program made participants worse. The use of an RPT design was further supported by the camp collaborators, who expressed parents would respond more favorably to completing one rather than two measures (Marshall, Higginbotham, Harris, & Lee, 2007).

## Measurement

The PPDO presented within Garst and Gagnon (2016) comprised five factors: (1) communication, (2) responsibility, (3) self-regulation, (4) attitude, and (5) exploration. Table 1 provides the construct definitions of these factors and supporting research. While the study conducted by Garst and Gagnon (2016) was focused on concepts beyond these five factors, the methodological approach of assessing the reliability and validity of the PPDO within their study is reflective of the more contemporary approach described earlier (Brown, 2015). Specifically, to examine the psychometric properties of the PPDO, Garst and Gagnon used a CFA and supporting statistics, which indicated acceptable levels of reliability and emerging evidence of convergent and discriminant validity of the PPDO in a sample of 2,952 parents of residential campers. The CFA results suggested the PPDO demonstrated good convergent validity in terms of factor loadings ( $\lambda = .545-.874$ ), average variance extracted (AVE) scores (AVE =  $.548-.672$ ), and reliability ( $\alpha = .825-.894$ ). Additionally, evidence of discriminant validity was indicated by the between-factor correlations composing the PPDO ( $r = .446-.689$ ). Given these promising psychometric properties, the PPDO was unmodified within the current study.

**Table 1**

*Construct Definitions of the Parental Perceptions of Developmental Outcomes (PPDO) Scale Dimensions (Garst & Gagnon, 2016)*

<b>Dimension</b>	<b>Definition</b>
Responsibility	The ability to start and accomplish tasks without external motivation
Exploration	The ability to engage in new and/or challenging experiences
Self-Regulation	The ability to adapt through stress and adversity
Attitude	The ability to monitor and regulate behaviors
Communication	The ability to verbally express oneself

## Data Setup and Analyses

Two distinct analytic approaches were selected for this study, a composite-based approach (i.e., paired sample  $t$  tests) and a latent approach (i.e., structural equation modeling; SEM), to determine how the PPDO performs with the more practitioner-friendly composite approach. Prior investigations exploring these comparative analyses have indicated if the same sample is used, then this redundancy can introduce type 1 error to a study (Kline, 2011), where researchers may unintentionally use findings from one approach to inform the other (e.g., introducing confirmation bias). To mitigate this challenge, the researchers randomly assigned the sample ( $N = 930$ ) to one of two analysis conditions<sup>2</sup>: (1) composite (*Condition A*;  $n = 464$ ) or (2) latent (*Condition B*;  $n = 466$ ). The additional data setup and analysis of each subsample is described below.

**Condition A: Paired sample  $t$  test (composite approach).** Prior to the analyses, the data were examined for multivariate normality through the Mahalanobis distance and the chi-square distribution ( $p \leq .001$ ) function in SPSS 24 software. This examination indicated 38 respondents were negatively influencing multivariate normality within the data set and thus were removed from further analyses. Next, the reliability of the 10 proposed factors was examined through Cronbach's alphas ( $\alpha$ ), which determined the strength of relationships and internal consistency of the items composing each factor. The results indicated acceptable  $\alpha$  levels across all 10 factors (see Table 2). The researchers then transformed the items into their respective composite variables by dividing the total score of items composing a factor by the number of items within said factor. Next,  $H_1$  was examined through a combined set of paired sample  $t$  tests.

**Condition B: Structural equation model (latent approach).** The data for the latent analysis (SEM) were also examined for multivariate normality through Mahalanobis distance and the chi-square distribution ( $p \leq .001$ ) function in SPSS 24 software. This examination indicated 37 respondents significantly contributed to multivariate non-normality within the data set, which were removed from later analyses. Next, the data were screened for systematic causes of missingness using Little's test of MCAR (missing completely at random;  $p \geq .001$ ) in EQS 6.3 software (Little, 1988). The nonsignificant results of Little's test of MCAR,  $\chi^2(1008) = 1103.265$ ,  $p = .0191$ , indicated the data were MCAR. Thus, the researchers selected an expectation maximization (EM) technique to generate missing values using all available information (e.g., covariance matrices and standard errors) and predicted population level parameters (Enders, 2010).

Prior to testing  $H_1$ , the researchers examined the data for model fit and convergent and discriminant validity through CFA and supporting statistics. Specifically, they examined the model fit of the PPDO by using a combination of the comparative fit indices (CFI), non-normed fit indices (N-NFI), and the root mean square error approximation (RMSEA). Both the CFI and N-NFI provide evidence of a model fit, where levels closer 1.00 indicate better model fit (e.g.,  $\geq .90$ ; Fornell & Larcker, 1981). Finally, the RMSEA compares hypothesized model parameters to the data, where levels lower than .06 indicate better model fit (Marsh, Hau, & Wen, 2004).

The relationship strength of items proposed (i.e., convergent validity) to make up the PPDO factors was assessed through an examination of the item covariance matrices,<sup>3</sup> factor loadings ( $\lambda$ ), and Cronbach's alpha ( $\alpha$ ). Factor loadings represent the strength of relationship between an item and factor, where higher levels (e.g.,  $\lambda \geq .5$ ) indicate better predictive quality (Brown, 2015). Similarly,  $\alpha$  provides information on the relationships between items composing a factor, where higher levels (e.g.,  $\alpha \geq .80$ ) provide evidence of the internal consistency of a factor. The discriminant validity of the proposed measurement model was determined through an examination of between-factor Pearson correlation ( $r$ ) levels and square roots of AVE scores ( $\sqrt{\text{AVE}}$ ). Specifically,

<sup>2</sup>The sample sizes presented reflect pre-outlier screening levels described in the relevant sections below; the final sample sizes for the two conditions were 426 for Condition A and 421 for Condition B.

<sup>3</sup>A copy of the final covariance matrices available upon request from first author.

**Table 2***PPDO Scale Descriptive Statistics by Analytic Approach*

Construct/Item	Composite analyses (Condition A)				Latent analysis (Condition B)							
	Pre M (SD)	$\alpha$	Post M (SD)	$\alpha$	<sup>a</sup> Pre M (SD)	$\lambda$	$\alpha$	AVE	<sup>a</sup> Post M (SD)	$\lambda$	$\alpha$	AVE
<i>Responsibility</i>	3.96 (.715)	.906	4.08 (.618)	.879	-	-	.880	.601	-	-	.884	.604
takes responsibility for his/her own actions.	4.05 (.803)		4.13 (.736)		4.09 (.716)	.729			4.15 (.683)	.750		
takes care of his/her own things.	3.89 (.879)		4.00 (.769)		3.87 (.856)	.671			3.99 (.784)	.687		
shares work responsibilities.	4.00 (.812)		4.13 (.712)		4.05 (.746)	.805			4.16 (.719)	.773		
follows through when asked to do something.	3.86 (.895)		4.00 (.804)		3.91 (.852)	.848			4.01 (.782)	.851		
follows directions.	4.02 (.807)		4.13 (.742)		4.11 (.746)	.811			4.14 (.722)	.814		
<i>Exploration</i>	4.04 (.676)	.868	4.23 (.587)	.839	-	-	.864	.616	-	-	.855	.596
participates in new learning experiences.	4.18 (.726)		4.36 (.626)		4.17 (.739)	.855			4.37 (.627)	.836		
is curious about new topics and subjects.	4.16 (.730)		4.26 (.687)		4.20 (.687)	.684			4.32 (.646)	.671		
seeks challenges beyond his/her comfort zone.	3.73 (.954)		3.99 (.883)		3.85 (.888)	.783			4.09 (.813)	.761		
is willing to try new experiences.	4.10 (.764)		4.31 (.650)		4.18 (.741)	.808			4.36 (.639)	.809		
<i>Self-Regulation</i>	3.68 (.753)	.871	3.81 (.679)	.861	-	-	.880	.654	-	-	.876	.647
handles success and failure.	3.88 (.808)		4.02 (.697)		3.94 (.769)	.844			4.02 (.714)	.844		
manages disappointment well.	3.62 (.914)		3.76 (.834)		3.69 (.885)	.813			3.76 (.865)	.812		
deals effectively with conflict.	3.68 (.852)		3.80 (.796)		3.81 (.818)	.847			3.87 (.812)	.828		
doesn't get frustrated easily.	3.56 (.965)		3.68 (.914)		3.64 (.933)	.726			3.72 (.898)	.729		
<i>Attitude</i>	4.21 (.683)	.943	4.29 (.618)	.917	-	-	.938	.794	-	-	.936	.786
has a good mental attitude.	4.26 (.717)		4.34 (.663)		4.32 (.625)	.910			4.36 (.618)	.914		
has a generally "positive" view on life.	4.24 (.759)		4.32 (.699)		4.32 (.643)	.910			4.35 (.643)	.899		
is more hopeful about the future.	4.12 (.757)		4.21 (.719)		4.22 (.690)	.833			4.26 (.686)	.828		
shows a positive attitude when around others.	4.24 (.726)		4.31 (.693)		4.31 (.632)	.909			4.35 (.622)	.903		
<i>Communication</i>	4.23 (.672)	.930	4.31 (.615)	.917	-	-	.921	.750	-	-	.901	.739
participates in a discussion.	4.20 (.747)		4.28 (.704)		4.29 (.670)	.935			4.36 (.639)	.928		
communicates well with others.	4.21 (.742)		4.29 (.691)		4.24 (.701)	.922			4.31 (.660)	.921		
shares thoughts and ideas verbally.	4.25 (.739)		4.33 (.694)		4.28 (.719)	.912			4.34 (.683)	.901		
asks questions.	4.26 (.730)		4.34 (.666)		4.26 (.700)	.667			4.30 (.680)	.660		

Note. *M* indicates mean; *SD* indicates standard deviation;  $\alpha$  = Cronbach's alpha;  $\lambda$  indicates factor loading.

<sup>a</sup>Latent means are representative of EM missing data imputation; raw scores available upon request from first author.



correlation levels were examined for excessively high relationships (e.g.,  $r \geq .9$ ), as this suggests factors account for the same variance (Kline, 2011). Additional evidence of discriminant validity was established through a comparison of  $\sqrt{\text{AVE}}$  scores to between-factor correlations, where  $\sqrt{\text{AVE}}$  scores should be higher than correlation levels to establish the variance accounted for by individual factors was due to their unique variance, rather than variance shared with other factors (Brown, 2015). Finally, after evidence of the psychometric reliability of the PPDO was established,  $H_1$  was tested through latent paired sample  $t$  tests<sup>4</sup>. While well-established within nonlatency research, paired sample  $t$  tests are only just emerging as a viable process of examining differences across latent means (Coman et al., 2013). As such, the analyses used in this study reflect this emerging approach, where groups can be statistically compared and significant latent mean differences determined. The results of this approach can present similar results as those in composite-based approaches (e.g., significance values and Cohen's  $d$ ).

As part of the CFA, data were screened for nonnormality, which indicated eight additional cases were contributing to multivariate kurtosis and were removed from further analyses. After removal of these additional outliers, the final 10-factor PPDO CFA model indicated acceptable levels of fit,  $S/B\chi^2(753) = 1127.1691$ ,  $p \leq .001$ , N-NFI = .966, CFI = .970, RMSEA = .034 (90% CI .030–.038). Evidence of the PPDO's convergent validity (see Table 2) was established by the relatively high and consistent factor loadings across all items ( $\lambda = .660$ –.935) and acceptable  $\alpha$  levels ( $\alpha = .864$ –.938). Evidence of discriminant validity between factors was partly indicated by the general lack of excessively high correlations (see Table 3). However, all five factors of interest indicated high correlation levels (e.g.,  $r \geq .9$ ) from precorrelation to postcorrelation. This high correlation likely reflects construct stability across time for the five PPDO factors. More specifically, the constructs are “likely to have a stable empirical correlate if the relationships between indicators are invariant across contexts” (Lillibacka, 2006, p. 208), suggesting invariance from pretest to posttest, providing emerging support for discriminant and criterion validity of the PPDO factors. Beyond the five high pretest to posttest correlations, the  $\sqrt{\text{AVE}}$  values (see bolded diagonal line in Table 3) were all above the between-factor correlation levels, further illustrating the discriminant validity of the measures. In aggregate, these findings provide further evidence of the psychometric validity and reliability of the PPDO and the appropriateness of testing  $H_1$  through latent paired sample  $t$  tests.

**Table 3**  
*Latent Factor Correlations*

Factor	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
F1. Responsibility-Pre	<b>.775</b>									
F2. Responsibility-Post	.964	<b>.777</b>								
F3. Exploration-Pre	.651	.646	<b>.784</b>							
F4. Exploration-Post	.604	.688	.841	<b>.772</b>						
F5. Self-Regulation-Pre	.820	.788	.642	.601	<b>.808</b>					
F6. Self-Regulation-Post	.780	.808	.622	.686	.957	<b>.804</b>				
F7. Attitude-Pre	.725	.739	.706	.730	.710	.698	<b>.891</b>			
F8. Attitude-Post	.697	.737	.686	.767	.693	.713	.979	<b>.886</b>		
F9. Communication-Pre	.603	.602	.651	.624	.559	.536	.786	.760	<b>.866</b>	
F10. Communication-Post	.573	.622	.622	.686	.549	.583	.787	.806	.958	<b>.859</b>

Note. All factors correlated at  $p \leq .05$ . Bold indicates  $\sqrt{\text{AVE}}$ .

<sup>4</sup>In condition B,  $t$  tests were conducted as a fully latent model.

## Results

### Hypothesis 1 Results

The first hypothesis ( $H_1$ ) was that parents would observe and report statistically significant ( $p \leq .05$ ) growth across the five dimensions of the PPDO. The paired sample  $t$  tests (i.e., Condition A) indicated statistically significant increases in all five dimensions of the PPDO from precamp to postcamp, thus failing to reject  $H_1$ , indicating parents observed significant positive change in their child (see Table 4). However, upon examination of effect sizes across the five factors, only the exploration factor indicated a medium effect size ( $d = .302$ ), with self-regulation ( $d = .164$ ), responsibility ( $d = .157$ ), communication ( $d = .112$ ), and attitude ( $d = .105$ ) indicating small effect sizes (Cohen et al., 2003). More simply, there were significant increases in all five dimensions of the PPDO, but the meaningfulness of these increases was low in four of the five dimensions.

The SEM in Condition B, examining the latent  $t$  tests, indicated acceptable model fit:  $S/B\chi^2(787) = 1200.4250$ ,  $p \leq .001$ , N-NFI = .965, CFI = .969, RMSEA = .035 (90% CI .031-.039). The latent paired sample  $t$  tests (i.e., Condition B) indicated positive growth across all dimensions with parents reporting significant changes from precamp to postcamp in responsibility ( $M_{diff} = +.898$ ,  $p \leq .001$ ), exploration ( $M_{diff} = +.777$ ,  $p \leq .001$ ), self-regulation ( $M_{diff} = +.924$ ,  $p \leq .001$ ), attitude ( $M_{diff} = +.949$ ,  $p \leq .001$ ), and communication ( $M_{diff} = +.898$ ,  $p \leq .001$ ). The statistically significant results across all five dimensions of the PPDO within Condition B indicate a failure to reject  $H_1$ , as with the composite-based results in Condition A, albeit tempered by the relatively small effect sizes across four of the five dimensions. Specifically, responsibility ( $d = .117$ ), self-regulation ( $d = .104$ ), attitude ( $d = .060$ ), and communication ( $d = .099$ ) indicated small effect sizes, and, as in Condition A, only exploration ( $d = .322$ ) indicated a medium effect size.

### Hypothesis 2 Results

The second hypothesis ( $H_2$ ) was that the two analytic approaches would yield the same findings and effect sizes. With a focus on effect sizes (where a Cohen's  $d$  value less than .20 equals a small effect, a value between .20 and .50 equals a medium effect, and a value greater than .50 equals a large effect; Cohen et al., 2003), this study found similar effects for all outcomes across the two approaches. Responsibility, self-regulation, attitude, and communication indicated similar increases across the two analyses approaches, with exploration indicating a medium increase in both the composite (Condition A) and latent approach (Condition B). More simply, the effect sizes across the two approaches suggested near identical results; thus we also failed to reject  $H_2$ .

**Table 4**  
*Condition A (Composite): Paired t-Test Results*

Composite factor	Precamp <i>M (SD)</i>	Postcamp <i>M (SD)</i>	<i>t (df)</i>	Paired correlation
Responsibility	3.97 (.678)	4.07 (.622)	6.874* (425)	.883*
Exploration	4.05 (.630)	4.24 (.585)	10.526* (425)	.813*
Self-Regulation	3.71 (.723)	3.83 (.681)	6.638* (425)	.861*
Attitude	4.24 (.625)	4.30 (.617)	4.643* (425)	.890*
Communication	4.25 (.622)	4.32 (.611)	5.019* (425)	.890*

\* $p \leq .001$ .

## Discussion

The results illustrate the utility of the PPDO as a tool which performs equally well with a practitioner-centered or more computationally demanding analytical approach. As the results suggested, the PPDO is a promising metric for OST practitioners assessing developmental outcomes associated with the camp experience from the perspective of parents. The significance of engaging parents through measuring their perceptions of programs and services is also supported in the broader social science literature (Ball, Bates, Amorose, & Anderson-Butcher, 2017; Schueler, Capotosto, Bahena, McIntyre, & Gehlbach, 2014). Furthermore, access to measures of parent perceptions are particularly important at a time when camp program providers have identified the intensification of parental expectations for their child's camp experience (Garst, Gagnon, & Bennett, 2016), and being able to better understand parent perceptions of their child's camp experience can help program providers better meet the needs of parents—their primary stakeholder.

The support for statistically significant growth across all dimensions of the PPDO continues to illustrate residential summer camp as a potential context for socioemotional skill development (Schelbe, Hansen, France, Rony, & Twichell, 2018; Woods, Mayes, Bartley, Fedele, & Ryan, 2013). While the first hypothesis guiding this study (i.e.,  $H_1$ , *Parents will observe and report statistically significant ( $p \leq .05$ ) growth across the five dimensions of the PPDO*) has been seemingly well investigated within the camp literature (e.g., Baughman et al., 2009; Henderson, Whitaker, et al., 2007; Michalski et al., 2003), the finding of growth also addresses the “replication crisis” facing the broader social and behavioral sciences (Baker, 2016). Specifically, an effort by an international research team (Camerer et al., 2018) has indicated that when many well-regarded and implemented studies are replicated, they fail to duplicate the original results, calling into question the veracity and strength of findings. While the factors contributing to the lack of reproducibility of studies are often multidimensional, the similarity of findings demonstrated in the current study to those illustrated in Garst and Gagnon (2016) provide a foundation for the strength of the PPDO.

The results of the CFA provided in the second (latent) analyses offer additional support beyond Garst and Gagnon (2016) for the psychometric validity of the PPDO as a tool for camp researchers to explore parent observations of development using an RPT format. However, the process of scale validation and revalidation is, and should be, ongoing (Furr, 2018). Specifically, the current study findings provide evidence of the construct stability and convergent validity across the five measured outcomes as indicated in Table 2 (Bashkov & Finney, 2013). Further, the high correlations between the premeasures and postmeasures across the five measured factors illustrate additional support for their construct and/or predictive validity, suggesting that despite the significant mean differences from preprogram to postprogram, items were measuring the same construct.

This developing evidence of construct validity is tempered by the high between-factor correlations outside of the expected relationships within Table 3 (e.g., precamp responsibility should correlate strongly with postcamp responsibility scores). While  $\sqrt{\text{AVE}}$  levels were higher than nearly all between-factor correlations, some correlations were above .700, which indicates factors shared more than 50% of variance. Similar high levels of between-factor correlations have been demonstrated in other measures of camp outcomes. For instance, in their study of the YOB “A,” Sibthorp et al. (2013) indicated between-factor correlations ranging from .71 to .88. These limitations suggest an opportunity for future scale development within the camp environment, particularly with the development of measures demonstrating stronger levels of discriminant validity. One technique for advancing the validity of measures of camp outcomes is offered by Chen et al.'s (2015) construction and implementation of the Basic Psychological Need Satisfaction and Frustration (BPNFSF) scale. Specifically, the use of both positively valenced factors (e.g., autonomy satisfaction; *I feel a sense of choice and freedom in the things I undertake*)

and negatively valenced factors (e.g., autonomy frustration; *Most of the things I do feel like "I have to"*) provides a mechanism of demonstrating not only discriminant validity of measures, but also criterion/predictive validity (i.e., frustration and satisfaction should be negatively correlated). However, much of current camp scale research, including the PPDO and YOB, reflect only a "satisfaction" approach. Thus, future scale development and validation research in the camp environment should incorporate the negatively valenced measures to provide evidence of the accuracy of measures.

Future scale development targeting camp parents should also consider scale format (i.e., the number of choices provided by the scale). For example, although common, 5-point scales such as the one used with the PPDO produce higher mean scores because respondents tend to provide positive responses (Dawes, 2008), particularly when given only two options for a positive response (i.e., either a 4 or a 5). When a large proportion of respondents select the maximum score on a scale, a "ceiling effect" may occur. In the current study, a ceiling effect is potentially reflected in the item means and relative variances for attitude and communication presented in Table 2. A solution in future studies is to design scales with 7 or more points. Contemporary scale development within the recreation and leisure literature reflects this approach. For example, Gould, Moore, McGuire, and Stebbins (2008) used a 9-point scale in a measure of serious leisure.

Camp-related parent perception measurement could also be advanced through a comparison of the PPDO with similar measures of parent perceptions of these experiences. For instance, by collecting responses concurrently from the same population of parents using the PPDO and the ACA Staff and Parent Perceptions YOB, and then examining the factor covariances and supporting model fit statistics, we can determine where there may be convergence across conceptually similar factors across the different measures. Beyond providing comparative evidence regarding the performance of the PPDO and the ACA Staff and Parent Perceptions YOB, such a study would provide psychometrics for the ACA Staff and Parent Perceptions YOB, which have not yet been published in a peer-reviewed journal as of the production of this manuscript. As noted, the measures across camps are broad, while in many cases this is due to mission, it is also likely due to "siloeing." As suggested by Lewis et al. (2018), collaboration and replication of measures represents the next important "direction" for the assessment of camp-related youth outcomes. While establishing a "common" measure may be a difficult endeavor, most good science is (Popper, 1968). Collaboration across research teams represents a necessary condition for better understanding the unique outcomes frequently associated with camp experiences.

Although camp measures are often selected based on a camp's targeted outcomes, another way of further developing the camp-related parent perception measurement is to use the PPDO with camps that intentionally target youth outcomes *not reflected* in the PPDO. In such a scenario, if parent reports of child development is an accurate approach, then there should be lower scores on the PPDO outcomes and higher scores for the alternate outcomes specifically targeted by the camp. If scores on PPDO outcomes are as high or higher than scores for the alternate outcomes, then it may reflect response or recall bias on the part of parents, thereby calling into question the effectiveness of parent-reported measures of camp-related youth outcomes.

This study found that an RPT design was an appropriate technique for measuring parent perceptions following their child's camp experience. The effectiveness of RPT for studying program-related growth based on subjective perceptions of experiences supports the recommendations of L. Hill and Betz (2005) and offers additional confirmation of the usefulness of the RPT format for practitioners. Practitioners looking for a cost-effective and parent-friendly approach for measuring parent perceptions can benefit from the PPDO's strength as a validated measure in an RPT format. Specifically, there is no cost to utilize the PPDO, and the retrospective format only requires a single data entry per parent, also saving on costs associated with multiple survey administrations. Further, the emergence of easy-to-use spreadsheet-based *t*-test

calculators widely available at no to little cost to practitioners mitigates an additional burden associated with the training and resources necessary for more complex software packages (e.g., SPSS, EQS, Mplus, R).

One potential limitation of this study is embedded in the design of the PPDO. Although the rationale for and strength of RPT designs have been established (Allen & Nimon, 2007; Raidl et al., 2004; Sibthorp et al., 2007), there are also weaknesses inherent in such designs, including the inaccuracy of recall (Pratt, McGuigan, & Katzev, 2000), the tendency for study participants to respond in socially desirable ways (e.g., learning effects; Marshall et al., 2007), the propensity of participants to inflate program ratings to make the program look good (Pratt et al., 2000), and the fact that RPT feedback is collected only from participants who completed a program, which fails to capture the perspective of those who started but did not finish a program (Chang & Little, 2018). While some suggest a limitation of the RPT design is the desire to demonstrate a learning effect, the relatively low effect sizes (Table 4), evidence of discriminant validity (Table 3), and construct stability across the five dimensions of the PPDO illustrates the effectiveness of RPT as a method in a residential camp context, when using the PPDO.

When considering camp assessment through the researcher lens, the study findings also illustrate the benefits of SEM as compared to composite approaches. SEM presents a more nuanced story of the analyses of interest, provides additional tools to common challenges facing practitioners (e.g., missing data management), and provides detailed information on the validity and reliability of the measures of interest (Kline, 2011). SEM's strength in managing missing data is particularly relevant in this study. As recreation and leisure researchers, we must respect the time our participants provide in their responses to surveys and use techniques to address, rather than delete, missing data and incomplete responses (Freire & Caldwell, 2013). In this study, addressing missing data through expectation maximization (Condition B) instead of listwise deletion (Condition A) allowed for the retention of more of the overall sample. Using missing data management approaches facilitates research efficiency and permits researchers to be good stewards of respondents' time and effort (Enders, 2010).

The finding of support for the PPDO's psychometrics is encouraging, particularly as an applied measure. To further substantiate the PPDO as a practitioner-friendly measure, comparative analyses such as those presented in the current study should continue to identify the usefulness and/or limitations of the PPDO as measure of observed camper socioemotional growth. As noted, data in this study were collected entirely online. While emerging evidence suggests equivalent performance of Web-based and paper-based measures in terms of response variance, validity, and reliability (Campbell, Ali, Finlay, & Salek, 2015), the performance of the PPDO has only been assessed using an online format (e.g., Garst & Gagnon, 2016). It is possible the PPDO may perform differently if implemented in a paper format. Additionally, while beyond the scope of the current study, there is evidence that groups that are historically underrepresented within camp research (e.g., non-White and/or socioeconomically disadvantaged) tend to respond at lower rates to online questionnaires and more broadly to questionnaire-based studies regardless of their format (e.g., paper, telephone; Heiervang & Goodman, 2011). As such, researchers interested in assessing the efficacy of the PPDO should consider exploring its utility in alternate formats such as paper and, more broadly, how to engage groups underrepresented within camp research using measures such as the PPDO.

Measures like the PPDO can have broad applications within camp settings. Accurately capturing parent perceptions of a child's camp-related outcomes allows camps to develop parent-focused messaging prior to camp, which may be useful for encouraging new parents to sign up their child for camp, as well as existing parents to continue their child's engagement with camp. For example, evidence that a child became more responsible, was better able to regulate his or her emotions, or improved his or her attitude because of camp, and that these results were based on feedback received from parents, can create a compelling narrative that a particular

camp is an important summertime experience worthy of ongoing parental investment. When paired with other data sources (e.g., staff data), the PPDO may be useful for identifying which staff perform best in producing positive outcomes in youth. Measures such as the PPDO can also be helpful for practitioners interested in establishing an organizational culture more attuned to and trained in program evaluation. The PPDO is a user-friendly tool well suited for practitioners who want to expand their parent evaluation strategies while building capacity among camp employees to be engaged in a camp program evaluation process.

## Conclusion

Using evidence to inform programmatic and operational decision making within the context of camp programs is an important factor impacting program quality. However, when it comes to identifying and serving the needs of campers and their parents, many camp program providers merely “scratch the surface.” Camp administrators often collect limited data around parents’ experiences with select point-of-service dimensions (e.g., satisfaction with the camper check-in process), without considering how parents can appraise the camp of how youths’ knowledge, skills, behaviors, and attitudes have been influenced in the days and weeks following camp. The tendency to address what happens within the geographic or operational bounds (i.e., “four walls”) of a program rather than the longer term and broader impacts of programs is a common theme among camp program providers. An important factor in reversing this trend is the development of measures that are both psychometrically valid and practically useful. Although the PPDO is not the only measure with these characteristics, it is among a short list of measures designed and validated with camp populations. This study supported the PPDO as a promising measure for research *and* practice. Broadening the tools available to camp program providers as well as camp researchers will strengthen the body of knowledge supporting our understanding of the camp experiences and better position camp providers to advance their capacity to provide high-quality programs based on sound evidence.

## References

- Allen, J. M., & Nimon, K. (2007). Retrospective pretest: A practical technique for professional development evaluation. *Journal of Industrial Teacher Education, 44*(3), 1–7.
- American Camp Association. (2005). *Directions: Youth development outcomes of the camp experience*. Retrieved from [https://www.acacamps.org/sites/default/files/resource\\_library/report-directions-youth-development-outcomes.pdf](https://www.acacamps.org/sites/default/files/resource_library/report-directions-youth-development-outcomes.pdf)
- American Camp Association. (2018). *ACA youth outcomes battery*. Retrieved from <https://www.acacamps.org/resource-library/research/aca-youth-outcomes-battery>
- Appelbaum, M., Cooper, H., Kline, R. B., May-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist, 73*(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature, 533*, 452–454. <https://doi.org/10.1038/533452a>
- Ball, A., Bates, S., Amorose, A., & Anderson-Butcher, D. (2017). The parent perceptions of overall school experiences scale: Initial development and validation. *Journal of Psychoeducational Assessment*. Advance online publication. <https://doi.org/10.1177/0734282917742310>
- Bashkov, B. M., & Finney, S. J. (2013). Applying longitudinal mean and covariance structures (LMACS) analysis to assess construct stability over two time points. *Measurement and Evaluation in Counseling and Development, 46*(4), 289–314. <https://doi.org/10.1177/0748175613497038>

- Baughman, S., Garst, B., & Fuhrman, N. E. (2009). Consistency of developmental outcomes of 4-H camp experiences over time and across sites. *Journal of Youth Development, 4*(2), 17–28. <https://doi.org/10.5195/JYD.2009.261>
- Bhattacharyya, S., Mead, T. P., & Nathaniel, R. (2011). The influence of science summer camp on African-American high school students' career choices. *School Science and Mathematics, 111*(7), 345–353. <https://doi.org/10.1111/j.1949-8594.2011.00097.x>
- Bialeschki, M. D., & Sibthorp, J. (2011). Celebrating the camp experience through eighty years of camp research. *Taproot, 20*(2), 13–24.
- Bobzien, J., & Judge, S. (2014). Characteristics of peer models at a summer camp for children with autism. *Journal for Multicultural Education, 8*(4), 237–248. <https://doi.org/10.1108/JME-04-2014-0017>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York, NY: Guilford Press.
- Browne, L. (2018, August 13). Let's get retrospective. Retrieved from <https://www.acacamps.org/news-publications/blogs/research-360/lets-get-retrospective>
- Bultas, M. W., Steurer, L. M., Balakas, K., Brooks, C., & Fields, H. (2015). Psychosocial outcomes of a summer overnight recreational experience for children with heart disease. *Journal of Child Health Care, 19*(4), 542–549. <https://doi.org/10.1177/1367493514540350>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T., Huber, J., Johannesson, M., . . . Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature: Human Behavior, 2*, 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- Campbell, N., Ali, F., Finlay, A. Y., & Salek, S. S. (2015). Equivalence of electronic and paper-based patient-reported outcome measures. *Quality of Life Research, 24*(8), 1949–1961. <https://doi.org/10.1007/s11136-015-0937-3>
- Chang, R., & Little, T. D. (2018). Innovations for evaluation research: Multiform protocols, visual analog scaling, and the retrospective pretest–posttest design. *Evaluation and the Health Professions, 41*(2), 246–269. <https://doi.org/10.1177/0163278718759396>
- Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., . . . Ryan, R. M. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion, 39*(2), 216–236. <https://doi.org/10.1007/s11031-014-9450-1>
- Clary, C. D., & Ferrari, T. M. (2015). Communication, coping, and connections: Campers' and parents' perspectives of self-efficacy and benefits of participation in deployment support camps. *Journal of Youth Development, 10*(2), 31–54. <https://doi.org/10.5195/JYD.2015.407>
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New York, NY: Routledge.
- Collado, S., Staats, H., & Corraliza, J. A. (2013). Experiencing nature in children's summer camps: Affective, cognitive, and behavioural consequences. *Journal of Environmental Psychology, 33*, 37–44. <https://doi.org/10.1016/j.jenvp.2012.08.002>
- Coman, E. N., Picho, K., McArdle, J. J., Villagra, V., Dierker, L., & Lordache, E. (2013). The paired t-test as a simple latent change score model. *Frontiers in Psychology, 4*(738), 1–3. <https://doi.org/10.3389/fpsyg.2013.00738>
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point, and 10-point scales. *International Journal of Market Research, 50*(1), 61–104. <https://doi.org/10.1177/147078530805000106>
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

- Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18, 39–50. <https://doi.org/10.2307/3151312>
- Freire, T., & Caldwell, L. L. (2013). Afterthoughts on leisure and future research directions. In T. Freire (Ed.), *Positive leisure science: From subjective experience to social contexts* (pp. 213–223). New York, NY: Springer. [https://doi.org/10.1007/978-94-007-5058-6\\_12](https://doi.org/10.1007/978-94-007-5058-6_12)
- Furr, R. M. (2018). *Psychometrics: An introduction* (3rd ed.). Thousand Oaks, CA: Sage.
- Gagnon, R. J., & Bumpus, M. F. (2016). Fidelity and its importance to experiential and outdoor education. *Journal of Outdoor Recreation, Education, and Leadership*, 8(1), 10–25. <https://doi.org/10.18666/JOREL-2016-V8-I1-7279>
- Gagnon, R. J., Stone, G., & Garst, B. (2017). Beyond composite scores and Cronbach's alpha: Advancing methodological rigor in recreation research. *Journal of Outdoor Recreation, Education, and Leadership*, 9(4), 383–402. <https://doi.org/10.18666/JOREL-2017-V9-I4-7959>
- Garst, B., Baughman, S., Franz, N., & Seidel, R. (2013). Strengthening families: Exploring the impacts of family camp experiences on family functioning and parenting. *Journal of Experiential Education*, 36(1), 65–77. <https://doi.org/10.1177/1053825913481582>
- Garst, B. A., & Gagnon, R. J. (2016). A structural model of camp director practices and outcomes: Does intention toward program outcomes matter? *Journal of Park and Recreation Administration*, 34(4), 51–65. <https://doi.org/10.18666/JPARA-2016-V34-I4-7293>
- Garst, B., Gagnon, R., & Bennett, T. (2016). Parent anxiety causes and consequences: Perspectives from camp program providers. *LARNet: The Cyber Journal of Applied Leisure and Recreation Research*, 18(1), 21–39.
- Gillard, A., & Watts, C. E. (2013). Program features and developmental experiences at a camp for youth with cancer. *Children and Youth Services Review*, 35(5), 890–898. <https://doi.org/10.1016/j.childyouth.2013.02.017>
- Gould, J., Moore, D., McGuire, F., & Stebbins, R. (2008). Development of the serious leisure inventory and measure. *Journal of Leisure Research*, 40(1), 47–68. <https://doi.org/10.1080/0022216.2008.11950132>
- Heiervang, E., & Goodman, R. (2011). Advantages and limitations of web-based surveys: Evidence from a child mental health survey. *Social Psychiatry and Psychiatric Epidemiology*, 46(1), 69–76. <https://doi.org/10.1007/s00127-009-0171-9>
- Henderson, K. A., Bialeschki, M. D., Scanlin, M. M., Thurber, C., Whitaker, L. S., & Marsh, P. E. (2007). Components of camp experiences for positive youth development. *Journal of Youth Development*, 1(3), 1–12. <https://doi.org/10.5195/JYD.2007.371>
- Henderson, K. A., Thurber, C. A., Whitaker, L. S., Bialeschki, M. D., & Scanlin, M. M. (2006). Development and application of a camper growth index for youth. *Journal of Experiential Education*, 29(1), 1–17. <https://doi.org/10.1177/105382590602900103>
- Henderson, K. A., Whitaker, L. S., Bialeschki, M. D., Scanlin, M. M., & Thurber, C. (2007). Summer camp experiences: Parental perceptions of youth development outcomes. *Journal of Family Issues*, 28(8), 987–1007. <https://doi.org/10.1177/0192513X07301428>
- Hickerson, B., & Henderson, K. A. (2013). Opportunities for promoting youth physical activity: An examination of youth summer camps. *Journal of Physical Activity and Health*, 11(1), 199–205. <https://doi.org/10.1123/jpah.2011-0263>
- Hill, E., Gagnon, R., Ramsing, R., Goff, J., Kennedy, B., & Hooker, T. (2015). Measuring the impact of a medical specialty camp. *Therapeutic Recreation Journal*, 49(4), 310–325. <https://doi.org/10.18666/TRJ-2015-V49-I4-6308>
- Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26(4), 501–517. <https://doi.org/10.1177/1098214005281356>



- Kendellen, K., Camiré, M., Bean, C. N., & Forneris, T. (2016). Facilitators and barriers to leadership development at a Canadian residential summer camp. *Journal of Park and Recreation Administration, 34*(4), 36–50. <https://doi.org/10.18666/JPra-2016-V34-14-6514>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Lewis, K. M., Bird, M., Wilkins, T., Borba, J., Nathaniel, K., & Schoenfelder, E. (2018). Developing a common evaluation tool for camps. *Journal of Youth Development, 13*(1–2), 306–315. <https://doi.org/10.5195/JYD.2018.539>
- Lillibacka, R. (2006). Measuring social capital: Assessing construct stability of various operationalizations of social capital in a Finnish sample. *ACTA Sociologica, 49*(2), 201–220. <https://doi.org/10.1177/0001699306064774>
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>
- Marsh, H., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320–341. [https://doi.org/10.1207/s15328007sem1103\\_2](https://doi.org/10.1207/s15328007sem1103_2)
- Marshall, J. P., Higginbotham, B. J., Harris, V. W., & Lee, T. R. (2007). Assessing program outcomes: Rationale and benefits of posttest-then-retrospective-pretest designs. *Journal of Youth Development, 2*(1), 118–123. <https://doi.org/10.5195/JYD.2007.366>
- Merryman, M., Mezei, A., Bush, J. A., & Weinstein, M. (2012). The effects of a summer camp experience on factors of resilience in at-risk youth. *Open Journal of Occupational Therapy, 1*(1). <https://doi.org/10.15453/2168-6408.1016>
- Michalski, J. M., Mishna, F., Worthington, C., & Cummings, R. (2003). A multi-method impact evaluation of a therapeutic summer camp program. *Child and Adolescent Social Work Journal, 20*(1), 53–76. <https://doi.org/10.1023/A:1021467503588>
- Patton, M. Q. (2008). *Utilization focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Popper, K. (1968). *The logic of scientific discovery*. New York, NY: Routledge.
- Povilaitis, V., & Tamminen, K. A. (2017). Delivering positive youth development at a residential summer sport camp. *Journal of Adolescent Research, 33*(4), 470–495. <https://doi.org/10.1177/0743558417702478>
- Powers, J., Maley, M., Purington, A., Schantz, K., & Dotterweich, J. (2015). Implementing evidence-based programs: Lessons learned from the field. *Applied Developmental Science, 19*(2), 108–116. <https://doi.org/10.1080/10888691.2015.1020155>
- Pratt, C. C., McGuigan, W. M., & Katzev, A. R. (2000). Measuring program outcomes: Using retrospective pretest methodology. *American Journal of Evaluation, 21*(3), 341–349. <https://doi.org/10.1177/109821400002100305>
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation, 29*(4), 443–459. <https://doi.org/10.1177/1098214008324182>
- Raidl, M., Johnson, S., Gardiner, K., Denham, M., Spain, K., Lanting, R., . . . Barron, K. (2004). Use retrospective surveys to obtain complete data sets and measure impact in extension programs. *Journal of Extension, 42*(2). Retrieved from <https://www.joe.org/joe/2004april/rb2.php>
- Roth, J. L., & Brooks-Gunn, J. (2016). Evaluating youth development programs: Progress and promise. *Applied Developmental Science, 20*(3), 188–202. <https://doi.org/10.1080/10888691.2015.1113879>
- Schelbe, L., Hansen, M. E. D., France, V. L., Rony, M., & Twichell, K. E. (2018). Does camp make a difference? Camp counselors' perceptions of how camp impacted youth. *Children and Youth Services Review, 93*, 441–450. <https://doi.org/10.1016/j.chilyouth.2018.08.022>

- Schueler, B. E., Capotosto, L., Bahena, S., McIntyre, J., & Gehlbach, H. (2014). Measuring parent perceptions of school climate. *Psychological Assessment, 26*(1), 314–320. <https://doi.org/10.1037/a0034830>
- Sibthorp, J., Bialeschki, M. D., Morgan, C., & Browne, L. (2013). Validating, norming, and utility of a youth outcomes battery for recreation programs and camps. *Journal of Leisure Research, 45*(4), 514–536. <https://doi.org/10.18666/jlr-2013-v45-i4-3897>
- Sibthorp, J., Paisley, K., Gookin, J., & Ward, P. (2007). Addressing response-shift bias: Retrospective pretests in recreation research and evaluation. *Journal of Leisure Research, 39*(2), 295–315. <https://doi.org/10.1080/00222216.2007.11950109>
- Thurber, C. A., Scanlin, M. M., Scheuler, L., & Henderson, K. A. (2007). Youth development outcomes of the camp experience: Evidence for multidimensional growth. *Journal of Youth and Adolescence, 36*(3), 241–254. <https://doi.org/10.1007/s10964-006-9142-6>
- Williams, D. C., Ma, Y., Prejean, L., Ford, M. J., & Lai, G. (2007). Acquisition of physics content knowledge and scientific inquiry skills in a robotics summer camp. *Journal of Research on Technology in Education, 40*(2), 201–216. <https://doi.org/10.1080/15391523.2007.10782505>
- Woods, K., Mayes, S., Bartley, E., Fedele, D., & Ryan, J. (2013). An evaluation of psychosocial outcomes for children and adolescents attending a summer camp for youth with chronic illness. *Children's Health Care, 42*(1), 85–98. <https://doi.org/10.1080/02739615.2013.753822>