16

Leadership in Sport

Calum A. Arthur[1] *and Nicolas Bastardoz*[2]

[1] *University of Stirling, Stirling, Scotland*
[2] *University of Lausanne, Lausanne, Switzerland*

> *"The key to success for a coach lies in the relationships he holds with his players."*
> — Carlo Ancelotti, winner of the UEFA Champions League multiple times as a coach and as a player.

Coaches, via the behaviors they adopt, have a profound and lasting impact on their athletes and teams. Whether this is encouraging young people to pursue a lifetime of recreational sport and physical activity, impacting development or psychological variables such as self-esteem, or inspiring young athletes to pursue successful careers in professional sport, coaches are key influencers in this regard. Thus, it is of utmost importance that rigorous and robust scientific method be applied in sport coaching research. In this chapter, we review the various leadership theories and models, as well as empirical investigations within the sport leadership context.

The interest in sport coaches has been longstanding in the sport psychology discipline. Indeed, Gould and Wright (2012) citing Griffith (1925) highlighted that the sports coach is an important person to be researched, suggesting that observing what great coaches do and transmitting this information on to less-experienced coaches was a viable way to enhance best coaching practice. Since the birth of sport psychology, examining the behaviors coaches use, why the coach behaves in a certain way, and the effectiveness of coach behaviors have received considerable attention from sport psychology researchers. In this chapter, we trace the development and application of leadership theories in a sport context and review this substantial amount of work. Within the leadership and sport coaching literature there have been four broad strands of research: (1) Researchers have developed models and theories of leadership within sport (e.g., multidimensional model of leadership and mediational model of leadership); (2) researchers have applied leadership theories developed in other disciplines to the sport context (e.g., transformational leadership in Organizational Behavior and Industrial Organization; OB/IO); (3) researchers have observed "great coaches" (e.g., John Wooden, Patt Summit) and drawn learnings from them; and (4) researchers have developed standalone measures of coach behaviors (e.g., Côté's coaching behavior scale for sport) without underpinning theory. Within this review, we summarize the extant knowledge that has been developed in regards to the major leadership approaches within sport. We also make relevant comparisons with OB/IO fields, where the leadership phenomenon has been broadly studied, and which can provide new avenues for research.

In a second part, we argue that applying rigorous scientific methods is of paramount importance if we are to advance our understanding of the social environment of sport. In conducting our evaluation of the state of knowledge of leadership research in sport, we draw on methodological developments from other disciplines (e.g., economics) and provide a commentary on the sport leadership literature via this lens. We frame our discussion around the notions of causality and endogeneity and explain why the latter impedes the development of sport research. We offer solutions in the forms of research designs (e.g., randomized experiments, quasi-experiments), statistical methods (e.g., instrumental variable estimation, structural equation modeling) and coding of behavior. We conclude with a discussion about tautological definitions and theories that would in our opinion benefit from fresh theorizing.

Handbook of Sport Psychology, Fourth Edition, Volume I. Edited by Gershon Tenenbaum and Robert C. Eklund.

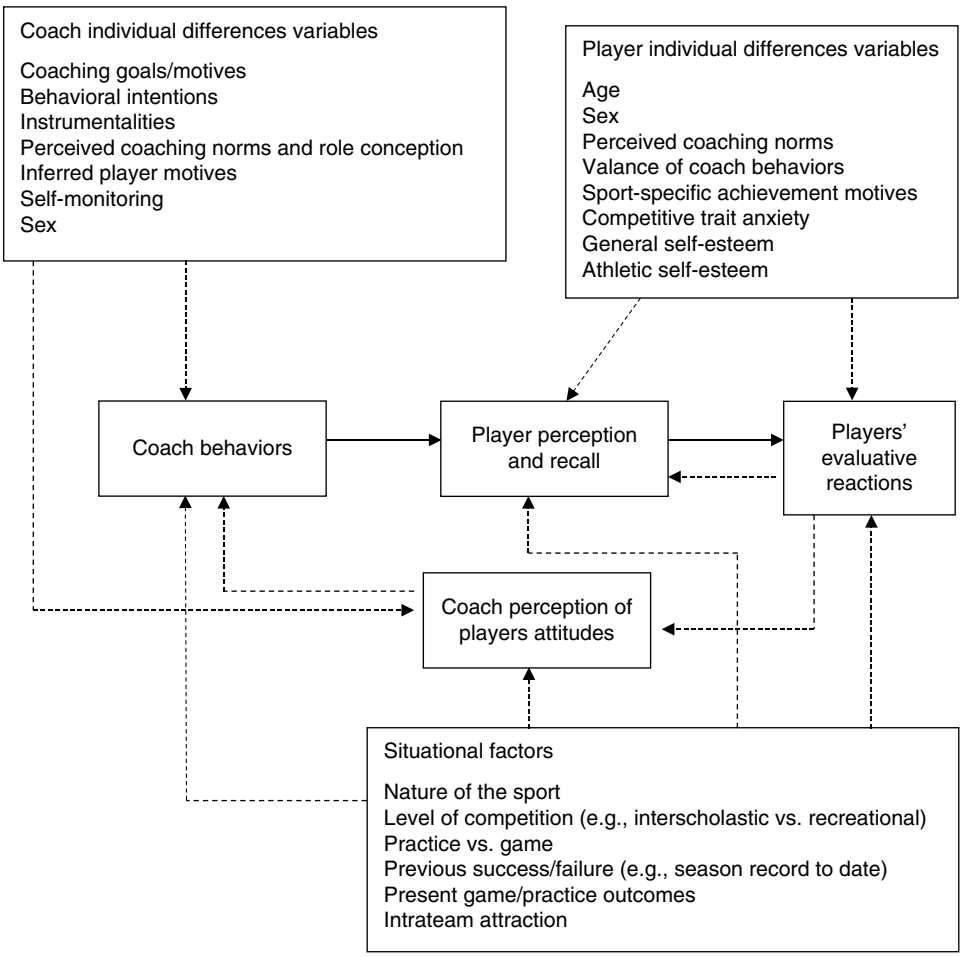


Figure 16.1 The mediational model of leadership. P. Chelladurai, 2007, Leadership in sports. In G. Tenenbaum and R.C. Eklund (Eds.) *Handbook of Sport Psychology* (3rd ed., pp. 113–135). Hoboken, NJ: Wiley. Reproduced with permission of John Wiley & Sons.

Mediational Model of Leadership

Originally developed by Smoll, Smith, Curtis, and Hunt (1978), the mediational model of leadership describes a leadership process that involves three central components: coach behaviors, players' perception and recall of the coach's behaviors, and players' evaluative reactions to the coach's behaviors (see Figure 16.1). Embedding their model within the social-cognitive perspective (e.g., Bandura, 1986), Smith and colleagues argue that coach behaviors affect athletes' evaluative reactions to these behaviors via athletes' perceptions and recall of the coach behaviors. Their model depicts a complex reciprocal interaction between contextual, personal, and behavioral factors, placing the influence process of coach behaviors within the environment in which it occurs. In line with the social-cognitive paradigm, the mediational model contains a complex array of individual differences and situational variables intertwined in a reciprocal causal network. The resultant model is inevitably rather complex with 12 basic relationships being hypothesized. Furthermore, when each of the higher order factors (i.e., coach individual differences, athlete individual differences, and situational factors) is broken down into their constituent parts, the model contains over 50 different parameters. The model depicts three antecedent factors of coach behaviors, namely, coach's individual differences, situational factors, and coach's perceptions of their athlete's attitude. In the athlete section of the model (i.e., player perception and recall, and player evaluative reactions), three antecedents are also specified, namely, athlete individual differences, situational factors, and the coach behaviors. Coach's perceptions of their player's attitude is identified as being influenced by coach's individual differences, situational factors, and player's evaluative reactions. This is in turn proposed to influence the coach's subsequent behaviors. A very complex picture of interacting and mediating

pathways emerges from this model. It is important to note that complexity is not a criticism of the model, rather the complexity of the model reflects the complex reality of the process it depicts. Indeed, a key strength of the mediational model of leadership explicitly embeds the leader influence process within the broader context in which it occurs.

Besides their multidimensional model, Smith and Colleagues developed a measure of coach behaviors entitled the coach behavior assessment system (CBAS) (Smith, Smoll, & Hunt, 1977). The CBAS is an observational behavioral assessment tool that researchers can use to record coach behaviors in a naturalistic setting. At a time when the interest in the youth sport environment was growing, Smith and colleagues observed very limited empirical investigation into the sports coach and specifically the behaviors they exhibited, as well as the consequences of these behaviors. In particular, Smith, Zane, Smoll, and Coppel (1983) noted that while there were lots of opinions about how the typical sport coach behaved, these opinions were unfortunately formed "…almost entirely on non-systematic observation, hearsay, and extreme examples of 'good' and 'bad' coaching." (p. 208). Thus, in developing the CBAS, Smith and colleagues hoped to provide a tool that enabled for systematic observations of coach behaviors thus providing a valuable insight into what coaches do. The development of the CBAS took several years: it was initially developed with soccer coaches but was subsequently broadened to include basketball, baseball, and football (see Smith et al., 1977, for a detailed description). Observers carried portable tape recorders and did a "play-by-play" account of the coach behaviors, and these recordings were subsequently transcribed and content analyzed (Smith et al., 1977). This process resulted in 12 behavioral categories being identified. The 12 CBAS categories are divided into two major classes of behaviors—reactive and spontaneous. Reactive behaviors are those that are in response to a discernible preceding event and spontaneous behaviors are those that are not preceded by a definable event. The reactive behaviors are further subdivided into three main categories: responses to desirable performance or effort, responses to mistakes, and response to misbehavior. The spontaneous category is divided into two subsections: game related and game irrelevant. In a recent review, Smith and Smoll (2007) noted the extensive use of the CBAS, with more than 85,000 behaviors coded for 80 male coaches in youth sport, and nearly 1,000 children were interviewed and administered questionnaires.

The CBAS offers a tool in which individual differences across behavioral patterns of coaches could be ascertained. Of particular note is the extensive training given to the coders. In a typical example, the coding of behaviors included training 22 observers over a 4-week program (Smith et al., 1983). The training comprised an extended study of a training manual, group instruction, written tests, and scoring of videotaped sequences (40 randomly ordered coaching behaviors). The observers were given practice and reliability checking was in place, and only 17 of the 22 observers that were trained were retained based on their performance scores (over 90% of correct coding). Clearly, the intensive training program and large number of trained observers is evidence of a diligent and comprehensive approach to measuring coach behaviors. A disadvantage is the time and resources needed to collect this type of data, which undoubtedly limit a wider application of this measurement paradigm. We suggest research to be conducted that can establish the minimum required training and observers that enable for accurate and reliable observations. Beyond some cutoff, there are certainly diminishing returns on additional training and observers that would enable a more efficient allocation of resources. Yet, we believe well-performed behavioral coding (compared to athlete ratings of coaches' behaviors) can provide data unaffected by relational, motivational, or attributional biases, an issue we revisit later.

The CBAS has also been modified to measure the athlete perspective of their coach behaviors and the coach perspective of their own behaviors. In this form of the CBAS, respondents are asked to rate how frequently their coach engaged in the behaviors from each of the 12 categories from CBAS (1 for "almost never" to 7 for "to almost always"). It is important to note that athlete report and coach self-report versions of the CBAS are measured by single item scales for each of the behavioral categories. Therefore, these versions of the CBAS have the standard problems associated with single item scales (for a critique of this approach, see Chelladurai & Riemer, 1998). The three measurement perspectives of the CBAS enabled descriptive research to be conducted that explored the differences between these three measurement perspectives. A notable finding is the typically non-significant or small correlations between coach self-reported behaviors and the observational measures of the same behavior. Interestingly, the punitive behavioral category produces the highest correlations between the various perspectives (around .5) compared to the other behavioral categories of the CBAS (Smith & Smoll, 2007). Athlete measurement typically correlates more strongly with observers' reports than with coach self-report measures. This led Smith and Smoll (2007) to conclude that "coaches were, for the most part, blissfully unaware of how they behaved and that athletes were more accurate perceivers of actual coach behaviors" (p. 79).

While the comparison of the different measurement perspectives has made a significant contribution to the literature and raises important questions about how

the different measurement perspectives can be used, the above conclusion may be somewhat premature. First, there is an implicit assumption within the conclusion that the observational measure is the gold standard method for measuring coach behaviors and is the only measure of actual coach behaviors. We do not believe the method is flawless and should necessarily predominate over other measurement perspectives. Each of the measurement approaches has potential utility, different strengths and weakness across contexts, and the chosen method should match the constraints of study design and the research question. Second, this conclusion was based on studies using predominately coaches that work with young athletes. This is problematic because youth athletes may rate coaches' behaviors differently from their adult counterparts. Also, coaches involved in youth sport may differ from adult sport coaches on important factors such as age or experience, and this may affect the correlations between the different measurement perspectives. A third aspect is that individual differences may moderate the strength of the correlation between the various perspectives. For example, individuals high on trait anxiety selectively attend to negative and threat-related information in the environment (Spielberger, 1966). High trait anxious athletes might thus report more negative perceptions of the coaches' behaviors compared to low trait anxious athletes, biasing the correlation between coach and athlete measures. The coach's personality can also affect measurement correlations, such that more self-aware coaches will be more aware of the behaviors they engage in. We encourage future research around measurement perspectives of coach's leader behavior to discuss how these contingencies may affect current knowledge.

Smith and Smoll (2007) summarized that the key findings of the research underpinned by the mediational model included:

- Coaches play a crucial role in determining the experiences of their athletes.
- The most positive athlete outcomes are observed when coaches engage with high levels of reinforcement for both desirable performance and effort, and when coaches respond to mistakes with encouragement and technical instruction.
- Coaches who display more reinforcement are liked more by their athletes and the athletes report having more fun and liking their teammates more.
- Team's win-loss record was unrelated to the extent to which athletes liked their coach and wanted to play for them; the win-loss record did influence athlete's perception of how much they thought their parents liked their coach.

One of the original motivations for developing the CBAS was to enable a systematic observation of coach behaviors and thus bring more scientific rigor to the debate on how coaches behave toward their athletes. In the 1970s, coaches were criticized for being highly punitive and creating excessively stressful environments. Systematic observational tools thus provided a mechanism by which researchers could examine and test such preconceptions. In fact, findings revealed that only about 3% of the coaches' behaviors were actually punitive and critical. Indeed, Curtis et al. (1979) summarized their results by concluding that "It is also refreshing to find that contrary to some stereotypes, observers and players perceived punitiveness relatively infrequently in the overall stream of coaching behavior" (p. 399). In examining the predictive properties of the different perspectives (observer and athlete report), Smith et al. (1978) found that both perspectives individually accounted for around 20% of variance in attitudinal outcomes. When entered simultaneously into the regression equation, the percentage of explained variance increased to around 40%, suggesting that both perspectives make unique contributions to the prediction of attitude, explaining a different aspect of the variance.

Building on the mediational model of leadership, several interventions were developed and tested that were collectively labeled as coaching effectiveness training program (e.g., Smith, Smoll, & Curtis, 1979; Smith, Smoll, & Barnett, 1995; Smoll, Smith, Barnett, & Everett, 1993) which were later developed into the mastery approach to coaching (Smith, Smoll, & Cumming, 2007). In an early example of the intervention research, Smith et al. (1979) conducted a coach intervention that was based on previous findings (Smith et al., 1978) and imbedded in a cognitive-behavioral framework (Bandura, 1977). The intervention design included aspects to help coaches become more aware of their behaviors, create expectancies concerning the likely consequences of coaching behaviors, and develop or enhance their ability to perform desirable behaviors effectively. The intervention was set up as a series of "Does" and "Don'ts" for coaches. For example "Do: REWARD! Do so immediately… Reward *effort* as much as you do results" (p. 62). Although outside observers only noted a significant increase in coaches' use of reinforcement behaviors between the treated and control coaches, players did report significant differences in experimentally treated coaches on various behaviors such as reinforcement, mistake-contingent encouragement, and general technical instruction, and significantly less non-reinforcement, punishment, and punitive technical instruction. Players also reported liking their coach more, wanted to play for their coach more next year, thought the coach was a better teacher, and reported

higher levels of intra-team attraction. Self-esteem was also enhanced in the experimental group.

Smith et al. (1993) conducted another intervention study that focused on improving players' self-esteem. Specific hypothesis were formulated around intervention effects being greatest for children with low self-esteem. The results supported their hypothesis that children who were lower in self-esteem at pre-test experienced greatest gains. Another notable result from this study was that while the children in the trained coaches' group reported liking baseball more and liking their coach more than the control group children, there was no difference between the groups in terms of win-loss record. In a more recent intervention study, Smith et al. (2007) conducted their research in a youth basketball context and included anxiety and motivational climate as key outcomes of the intervention. This last development brought more theoretical focus to the intervention by explicitly including aspects of achievement motivation theories that emphasize the importance of creating a mastery (task)-involving motivational climate (e.g., McArdle & Duda, 2002). Overall the intervention research stemming from Smith and colleagues' work demonstrated evidence that developing coaches results in changing coach behaviors (as measured by observers and athlete reports), which have wide positive ramifications for youth athletes such as enhanced self-esteem, reduced anxiety and drop out, and athletes reporting having more fun.

Multidimensional Model of Leadership

Similar to the mediational model, the multidimensional model of leadership (e.g., Chelladurai, 2007) also identifies leader behaviors as being central to the influence process of coaches. The model depicts a causal pathway that specifies antecedents (situational, leader, and member characteristics), central mechanisms (required, actual, and perceived leader behaviors), and outcomes (satisfaction and performance). In the theoretical model (see Figure 16.2 model A), situational characteristics influence required and preferred behaviors, whereas leader characteristics affect the leader's actual behavior and member characteristics impact preferred and required behaviors. Required, actual, and perceived leader behaviors are then proposed to influence satisfaction and performance. In other words, central mechanisms are hypothesized to mediate the relationship between characteristics and outcomes. Lastly, the multidimensional model of leadership was updated to include transformational leadership (see later section for a description of transformational leadership).

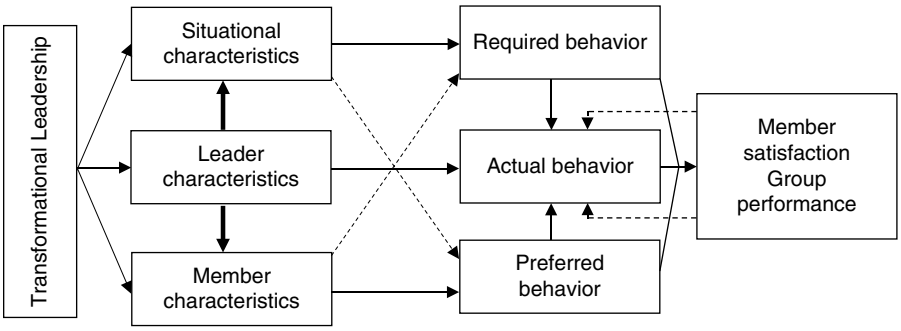
Leadership theories prominent in organizational and mainstream psychology literatures (e.g., Hemphill, 1950; Kahn & Katz, 1952; Lewin, Lippitt, & White, 1939) vastly influenced the development of the multidimensional model of leadership. For example, the contingency approach is a central component of the model (see Fiedler, 1964), and we can see this influence reflected in the congruence hypothesis (Yukl, 1971). The congruence hypothesis specifies that the discrepancy between actual leader behaviors and the followers' preferred behavior or required behavior will determine the outcomes of satisfaction and performance, with the smaller discrepancy yielding more positive athlete outcomes.

Chelladurai and colleagues developed a measure of leadership called the leadership scale for sport (LSS) (Chelladurai & Saleh, 1980). The LSS, designed to measure salient coach leader behaviors, is a 40-item scale split into five behavioral dimensions: training and instruction, democratic behaviors, autocratic behavior, social support, and positive feedback. The following definitions were obtained from Chelladurai (2007, p. 120).

> Training and instruction—Coaching behavior aimed at improving athletes' performance by emphasizing and facilitating hard and strenuous training; instructing them in the skills, techniques, and tactics of the sport; clarifying the relationship among members and structuring and coordinating the members' activities.
>
> Democratic behavior—Coaching behavior that allows greater participation by all the athletes in decisions pertaining to group goals, practice methods, and game tactics.
>
> Autocratic behavior—Coaching behavior that involves independent decision making and stresses personal authority.
>
> Social support—Coaching behavior characterized by a concern for the welfare of individual athletes, positive group atmosphere, and warm interpersonal relations with members.
>
> Positive feedback—Coaching behavior that reinforces an athlete by recognizing and rewarding good performance.

Three different measurement perspectives of the LSS have been developed: (1) athlete report of their coach's behaviors (*actual*); (2) athlete report of their own preferred leader behaviors (*preferred*), and (3) leader self-reported behaviors (*self-report*). All three measurement perspectives have been utilized in the extant literature. The LSS provides a tool that can be used to test the main predictions within the multidimensional model of leadership.

Modal A. Multidimensional model of leadership



Model B. Modified multidimensional model of leadership

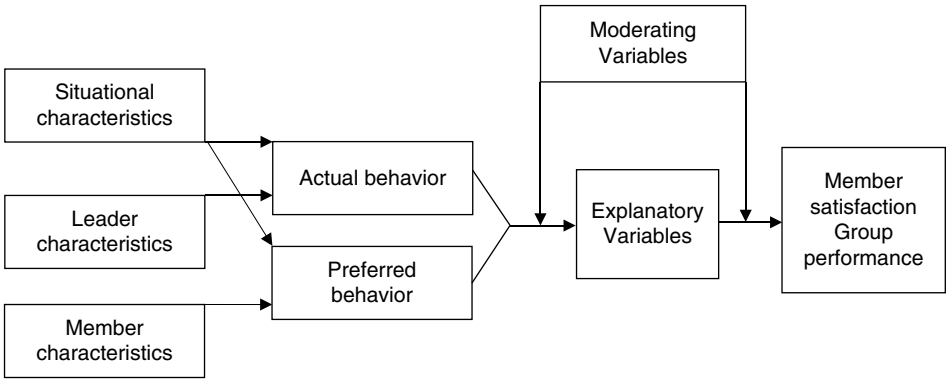


Figure 16.2 The multidimensional model of leadership and the modified multidimensional model of leadership. P. Chelladurai, 2007, Leadership in sports. In G. Tenenbaum and R. C. Eklund (Eds.), *Handbook of Sport Psychology* (3rd ed., pp. 113–135). Hoboken, NJ: Wiley. Reproduced with permission of John Wiley and Sons.The modified model depicted here is not meant as a definitive model but rather for it to be used as a guide to help steer future research in this area. We highlight the importance of including explanatory and moderating variables within the theoretical fabric of the model. We have not explicitly identified explanatory or moderating variables as these will need careful theoretical development and empirical testing. However, the potential explanatory variables are likely to include variables such as confidence and motivation, and the moderating variables will likely consist of individual difference variables such as sex, age, personality, level, and will also include situational and contextual variables. We have also removed the required behavior dimension from the model until proper definitions and measurement issues can be resolved, and thus for the aim of parsimony this aspect has been removed. We recommend that actual behaviors included within the model are broadened beyond those contained within the LSS. Lastly, we have also removed the arrows flowing from leader characteristics to situational and member characteristics as these paths are not clearly defined.

A major contribution Chelladurai made to the leadership literature in sport was the development of the congruence hypothesis. Congruence defines a level of agreement, or fit, between two related constructs, and this congruence is generally hypothesized to affect an outcome (see Edwards, 1994; Kristof, 1996). Within sport, Chelladurai theorized two possible "fit" combinations: (1) preferred-actual congruence (actual leader behaviors from an athlete's perspective); and (2) actual-required congruence (with the measurement of required behavior being unclear, we discuss this later). Research in sport has focused mostly on the preferred-actual congruence, with limited research being conducted on actual-required congruence. This lack of research is probably due to the problems associated with defining and measuring exactly what the leader's required behaviors are.

Findings regarding the preferred-actual congruence hypothesis within sport are mixed and contradictory. Some studies supported the congruence hypothesis (Aoyagi, Cox, & McGuire, 2008; Chelladurai, 1984; Horn & Carron, 1985; Riemer & Chelladurai, 1995; Schliesman, 1987; Shields, Gardner, Light Bredemeier, & Bostro, 1997), whereas other studies did not (e.g., Riemer & Toon, 2001; Chelladurai, Imamura, Yamaguchi, Oinuma, & Miyauchi, 1988). These heterogeneous findings can be explained because of differences in the analytical techniques used to test the congruence hypothesis, differences in leader behaviors examined, differences in samples such as individual or team sport, and individual differences such as age or gender.

With regards to actual-required congruence, as we mentioned earlier, there is very limited research that has

tested this aspect of the multidimensional model of leadership, with only one published study investigating this actual-required congruence (Kao, Chen, Watson, & Halbrook, 2015). Kao et al. (2015) operationalized the required dimension as the athletic director's preference for the coach to display leader behaviors and found support for the congruence hypothesis, but only for positive feedback. There is also an unpublished study that is referred to in Riemer (2007) that tested the actual-required congruence (Chelladurai, 1978). In this study, required behavior was operationalized as the average athlete preference for leader behaviors for each sport, Riemer (2007) reported that no evidence for congruence was obtained. Lastly, Shields et al. (1997) introduced a new dimension of congruence, which they labeled perceptual congruence. Perceptual congruence is the congruence between actual leader (measured by athlete perception) and leader self-report behavior. Light Shields and colleagues also relabeled Chelladurai's preferred-actual congruence (actual leader behavior measured by athlete perceptions) to value congruence. They found that while both value and perceptual congruence predicted cohesion, perceptual congruence was the strongest predictor of cohesion. However, methods used to test congruence hypotheses are problematic and the conclusions that can be drawn from this research are at best tenuous.

To test the congruence hypotheses, researchers have utilized two different methods. The first is to calculate the congruence score by subtracting the preferred behavior from the actual behavior (or vice-versa) and utilizing this discrepancy score in the subsequent analysis as a predictor of an outcome (e.g., satisfaction) or a derivative thereof (e.g., squared discrepancy score). The second is to generate an interaction term between preferred and actual behavior and employ a moderated hierarchical regression analysis (Reimer & Chelladurai, 1995; Chelladurai & Riemer, 1998). Evidence of congruence is purportedly established if the interaction term (product of preferred and actual behavior) predicts variance in an outcome (e.g., satisfaction) over and above the main effects for both variables. The interactional technique was introduced to sport as a potential solution to the problems that were associated with using discrepancy scores (see Reimer & Chelladurai, 1995).

Using discrepancy (or difference) scores to test congruence hypotheses is inappropriate for many reasons that relate to the reliability and the conceptual ambiguity of the difference score variable, and unreasonable constraints on the estimation model (see Edwards, 1994, 2001; Edwards & Parry, 1993; for more extensive discussions about why difference scores are not appropriate). Difference scores were also prevalent in OB/OI research but are not used anymore to test congruence hypotheses. Instead, researchers can use a method that enables congruence to be modeled over three-dimension graphs by combining polynomial regression with response surface methodology. Basically put, a polynomial regression indicates that both variables measuring fit or congruence be modeled (e.g., actual and preferred leader behavior), as well as their interaction (e.g., actual*preferred) and their quadratic term (e.g., actual*actual and preferred*preferred). Predicted values can then be plotted on a three-dimensional graph to display in a visually intuitive way what the estimated model predicts for the lines of fit (when actual = preferred) and misfit (when actual = - preferred). Interesting analyses and illustrations can ensue from polynomial regression and response surface methodologies (Edwards, 2002). Recently, Edwards and Parry (2018) extended the work on polynomial regression by incorporating spline regression into the congruence domain. It is important to note that Edwards does not recommend that spline regression replace polynomial regression, rather he argues that both techniques can be used to investigate congruence. The major difference is that polynomial regression is suited to curvilinear and symmetric surfaces (either side of the congruence axis) whereas spline regression is suited to linear and potentially asymmetric surfaces (Edwards & Parry, 2018). The choice of which technique to use should be based on the hypothesized shape of the surface, that is, whether researchers expect linear or curvilinear relationships at different points on the congruence continuum. In light of the statistical advancements, we urge sport researchers to investigate congruence within sport from a new methodological standpoint. We believe that congruence research is an exciting area for development and offers promise into understanding the relationship between leader behaviors and outcomes.

Along with testing the congruence hypothesis, the LSS has been used to gather descriptive data on athlete preferences, athlete perceptions of frequency of leader behaviors, and it has also been used to test relationships between perceived leader behaviors and athlete outcomes. In terms of athlete preference, interesting differences have emerged in terms of demographic variables such as gender, sport type, playing position (defensive vs. offensive), and ability. Reimer and Chelladurai (1995) reported that, in a sample of National Collegiate Athletic Association (the highest level of College Football in the USA) football players, defensive players reported greater preference for democratic, autocratic, and social support than their offensive counterparts. Riemer and Toon (2001) reported differences in preference based on ability and gender where lower ability athletes (division I vs. division II) prefer positive feedback compared to higher-ability athletes, and male athletes preferred more autocratic behaviors and female athletes preferred more

positive feedback. Relatedly, Martin, Jackson, Richardson, and Weiller (1999) found that females had a stronger preference for democratic behaviors than males. They also found that early adolescent (10–13 years) did not differ in their preferences than older adolescents (14–18 years). In another study, Horn, Bloom, Berglund, and Packard (2011) revealed that the psychological characteristics of trait anxiety and motivation (self-determination) were related to athlete preference. In an effort to explain the contradictory results, Cruz and Kim (2017) found evidence of an interaction between athlete gender and coach gender in determining athlete preference. Specifically, boys with a female coach preferred democratic, autocratic, and social support than boys with a male coach, and girls with a male coach preferred more democratic, autocratic, and social support than with female coaches.

Overall, the results from the preference literature indicate that differences in preferences exist along certain demographic dimensions and coaches may wish to consider gender (of athlete and coach), ability, playing position, athlete psychological character, and the interaction between these variables (cf., Cruz & Kim, 2017). Exactly how and why these differences play out and the consequences of these differences are not clear and could be an interesting avenue for future research. More theoretically guided research is required in this area before firm conclusions and recommendations can be made. For example, there is limited information on the consequences of the preferences; just because an athlete (or a group of athletes) reports a preference for certain behaviors does not mean it increases the athlete's performance. Furthermore, if an athlete indicates preferring a certain behavior and receives lots of this behavior, to what extent will this bring about positive adaptations in the athlete? Moreover, researchers have focused primarily on gender, sport type, ability, playing position, and age as determining differences in levels of preference; future research should probably start focusing on other differentiating factors such as personality dimensions (e.g., see Horn et al., 2011). For example, athletes' level of narcissism, extroversion, or conscientiousness may impact preferences for certain behaviors. Lastly, research has focused on simple main effect paradigms (see Cruz & Kim, 2017, for an exception) with the combination and/or timing of certain behavioral patterns being yet unexplored. For example, if an athlete receives high levels of their preferred behaviors, might this ameliorate against the assumed negative impact of receiving high levels of the non-preferred behaviors?

In terms of perceived levels of behaviors, differences by gender, ability, playing position, scholarship type, and sport type dimensions have been unearthed. Hollembeak and Amorose (2005) reported differences in perceived levels of leader behaviors based on sex, scholarship type, and individual (vs. team) sport. Specifically, males reported receiving higher frequency of autocratic and lower frequency of democratic than females; partial scholarship athletes reported that they received less training and instruction than non-scholarship and full scholarship athletes; and individual sport athletes reported receiving less training and instruction and autocratic behaviors and more democratic behaviors from their coaches. In a similar study, Gardner, Shields, Bredemeier, and Bostrom (1996) reported that males perceived greater levels of autocratic behavior but reported that females perceived that their coaches displayed greater frequencies of training and instruction, democratic behavior, and positive feedback. Moreover, Gardner et al. (1996) reported that junior college athletes perceived their coaches to display more social support than high school athletes. The research evidence to date indicates that there appear to be differences in levels of reported behaviors based on demographic or contextual variables, but the evidence is disparate. We recommend that a systematic review be conducted in order to clarify and organize this literature, and set forth a future research agenda. Theoretical clarifications are sorely needed to advance the field further.

Researchers have also utilized the LSS to test whether levels or frequency of the leader behaviors measured by the LSS predict outcomes (e.g., Amorose & Horn, 2000; Gardner et al., 1996; Hollembeak & Amorose, 2005; Horn & Carron, 1985; Weiss & Friedrichs, 1986). The results have revealed that the frequency of coach behaviors (as measured by the LSS) have been related to variables such as intrinsic motivation (Amorose & Horn, 2000; Hollembeak & Amorose, 2005), needs satisfaction (Hollembeak & Amorose, 2005), and cohesion (Gardner et al., 1996; Shields et al., 1997). Moreover, research has examined the moderating influence of gender. For example, Amorose and Horn (2000) found differences between male and female athletes, in that high levels of training and instruction and low level of autocratic behavior predicted males' intrinsic motivation, whereas females' intrinsic motivation was also predicted by democratic behaviors. The LSS has also been adapted to measure peer leadership (Loughead & Hardy, 2005; Vincer & Loughead, 2010). Loughead and Hardy (2005) compared coach and peer leaders on how much they displayed each of the dimensions in the LSS. The results revealed that coaches were rated as exhibiting greater levels of training and instruction and autocratic behaviors, whereas peer leaders were rated as displaying greater levels of social support, positive feedback, and democratic behaviors than the coach. Vincer and Loughead (2010) examined the relationship between peer leadership and group cohesion, finding that training

and instruction, social support, and democratic behaviors positively impacted cohesion, while autocratic behaviors negatively impacted cohesion.

Based on our review of the literature, we believe that while the multidimensional model of leadership has pushed forward the literature, the model would benefit from being updated. We suggest strengthening the multidimensional model of leadership and so we propose to: (1) revisit the inclusion of transformational leadership as an antecedent; (2) include a broader range of leader behaviors; (3) remove the required behavior dimension; (4) include mediators to explain the relationship between (actual and preferred) behaviors and outcomes; and (5) include moderator variables (see Figure 16.2 Model B)—and we discuss these suggestions in the following section.

Modification 1: Revisit transformational leadership in the model. We suggest that the location of transformational leadership in the model is problematic. Chelladurai (2007) and Riemer (2007) situated transformational leadership as a determinant of situational characteristics, leader characteristics, and member characteristics. Indeed, Riemer (2007) stated that "…transformational leadership behavior (sometimes referred to as charismatic leadership; see e.g., Yammarino et al., 1997) is presumed to influence not only the characteristics of the leader but also those of the member and situation" (p. 62). This statement is conceptually problematic because (1) transformational leadership is a behavioral approach to leadership, and (2) it is unclear who is supposed to behave in a transformational way. The coach's behaviors are depicted downstream in the model, which means that the transformational leadership component of the model cannot emanate from the coach. If the person doing the transformational leadership is not the coach, then presumably it must be the coach's line manager (e.g., performance director or athletic director) or someone else in the sport club or wider organization. While the lack of precision regarding "who" is actually doing the transformational leadership is problematic, a more serious problem relates to transformational leadership being an antecedent of situational, leader, and member characteristics. This is conceptually not possible because individual characteristics such as personality, gender, and age are not malleable and thus cannot be affected by transformational leadership. We therefore suggest moving transformational leadership within the actual behavior dimension of the model to supplement the behaviors identified in the LSS (see Figure 16.2a), leading us to our second suggested modification.

Modification 2: Broaden the behavioral repertoire included in the model. Not only could transformational behaviors be incorporated into the behavioral aspect of the model, but we also believe in the incorporation of further leader behaviors into the model. We recommend that researchers look both within sport research and to other disciplines such as OB/IO (e.g., instrumental leadership, Antonakis & House, 2014) to test the congruence hypothesis. Enlarging the behavioral repertoire investigated using the congruence hypothesis would further increase our understanding of leader behaviors in sport.

Modification 3: Remove the required behavior dimension. Given the problems associated with defining, operationalizing, and measuring the required behavior dimension, we suggest that this aspect of the model is removed. Chelladurai (2007) does not precisely define the required behavior dimension but rather describes the determinants of the required behavior. Thus, ambiguity and confusion surround this construct. This is evidenced in researchers either not including the construct in their research or operationalizing and measuring it in different ways (e.g., Chelladurai, 1978; Kao et al., 2015). Therefore, in the absence of an appropriate definition, we suggest that the construct is removed from the model.

Modification 4: Include theoretically derived explanatory processes. Further work is required to develop theoretical explanations for the congruence hypothesis. That is, researchers need to develop theoretically based explanations as to why congruence should affect the outcomes of interest in the model, members' satisfaction, and group performance. Identifying the key mechanisms within this relationship is crucial to furthering our understanding of the constructs. We encourage researchers to theoretically explicate the influence process of congruence and provide robust empirical tests of these. Mediators of the congruence-outcome relationship could include cognitive (e.g., identification), affective (e.g., positive or negative affect), relational (e.g., trust, leader-member exchange) or motivational (e.g., self-efficacy) constructs. In addition to developing theory around explanatory processes, a key aspect of this theoretical development lies in identifying boundary conditions around the explanatory processes, which leads us to our last suggested modification.

Modification 5: Include moderating variables (boundary conditions) to identify the scope and validity of the theory. Fundamentally, exploring boundary conditions informs under what conditions and for whom theoretical linkages hold or not. The exploration of boundary conditions could refer to contexts, time frames, or individuals that would increase or decrease the effect of congruence on satisfaction and performance. Interestingly, in explaining why they never found support for the congruence hypothesis, Riemer and Toon (2001) hinted that situational variables may indeed intervene in

this section of model, stating that "While the MML suggests that situational factors are antecedent to preferences and perceptions (i.e., actual behavior), they might also impact how preferences and perceptions interact to affect satisfaction" (p. 251). We thus encourage the test of potential moderators within the fabric of the model.

Overall, the research around the multidimensional model of leadership has greatly contributed to the understanding of sport leadership, and there is still today a clear need for more research in this area. The fact that Chelladurai's theorizing remains relevant today, some 40 years after its first appearance, is a testament to the value and depth of his theory. We hope to see more of this line of research flourishing in the sport science literature, especially if methodological and theoretical aspects are diligently taken into account.

Other Leadership Models within Sport

The mediational and multidimensional models of leadership have been the theories most widely researched in sport, and both have made a substantial contribution to knowledge. Other models and approaches have recently appeared in the sport leadership literature. However, these models have received relatively little direct research attention so far. These include Horn's model of coaching effectiveness (Horn, 2008), the motivational model of the coach athlete relationship (Mageau & Vallerand, 2003), and the coach-created motivational climate (Duda & Balaguer, 1999). Similar to the mediational and multidimensional models of leadership, these models adopt a process approach to describing coaching whereby antecedents and consequences of coach behaviors are mapped out. These models have been described elsewhere, and as such we do not provide a description of these here (e.g., Horn, 2008). Nonetheless, we will discuss three other approaches to coach leadership that have been developed in the sport literature, namely, coaching efficacy (Feltz et al., 1999), Côté's development of the coach behavior scale for sport (Côté, Yardley, Sedwick, & Baker, 1999), and transformational leadership (Arthur, Bastardoz, & Eklund, 2017).

Coaching Behavior Scale

Contrary to previous models that emphasized processes, Côté and colleagues developed a measure of coaching behavior named "coaching behavior scale" for sport (CBS-S: Côté et al., 1999). Built on qualitative research with coaches and athletes, the CBS-S represents a behavioral taxonomy of sports coaches. Athletes are asked to rate the frequency that their coach engages in the following seven behaviors (definitions were obtained from Baker, Côté, and Hawes, 2000, pp. 113–114):

> Physical training and planning: behaviors designed to enhance the physiological conditioning of the athlete. Specific behaviors included having a yearly training plan and providing structured workouts.
>
> Goal setting: Behaviors that aid the athlete in setting and achieving personal goals for sport. Specific behaviors included setting long- and short-term goals.
>
> Mental preparation: Behaviors designed to help athletes mentally prepare for their sport. Specific behaviors include providing advice on staying positive and focused.
>
> Technical skills: Behaviors that develop the technical aspects of the athlete's sport. Specific behaviors include the use of positive reinforcement and feedback.
>
> Personal rapport: Behaviors that develop the positive relationship between athlete and coach. Specific behaviors include developing a sense of trust and confidentiality.
>
> Negative personal rapport: Behaviors that develop a negative relationship between athlete and coach. Specific behaviors include yelling when angry and using fear and intimidation.
>
> Competition strategies: Behaviors designed to prepare the athlete for competition. Specific behaviors include ensuring needs are met at competition site and maintaining consistency during competition.

Researchers utilizing the CBS-S have examined the influence of these coach behaviors on athlete outcomes such as satisfaction (Baker, Yardley, & Côté, 2003) and anxiety (Baker, Côté, & Hawes, 2000). Baker et al. (2000) found that negative personal rapport behaviors strongly correlated with anxiety. In a further study, Baker et al. (2003) revealed that sport type (team vs. individual) moderated the relationship between coach behaviors and satisfaction such that stronger effects were evidenced for all the coach behaviors with team sports. The CBS-S contribution to the sport leadership literature lies in adding more coaching behaviors to the behavioral repertoire already existent in sport. In line with our earlier call for additional leader behaviors to be tested within the parameters of the congruence hypothesis, the behaviors identified in the CBS-S would seem to be an example from the sport literature that fits this call. However, the development of the behavioral dimensions contained within the CBS-S was largely

theory-free, and thus, it would prove beneficial to entrench these behaviors and their associated outcomes in a theoretical framework.

Coaching Efficacy and Competency Models

The coaching efficacy model (Feltz et al., 1999; Myers, Feltz, & Wolfe, 2008; Myers, Feltz, Chase, Reckase, & Hancock, 2008) is embedded within Horn's (2002) model of coaching effectiveness and social-cognitive theory. Coaching efficacy is defined as "the extent to which coaches believe they have the capacity to affect the learning and performance of their athletes" (Feltz et al., 1999, p. 765). Coaching efficacy originally comprised four dimensions: motivation efficacy, character building efficacy, game strategy efficacy, and technique efficacy, with a fifth dimension (physical conditioning efficacy) being added later (Myers, Feltz, Chase, Reckase, & Hancock, 2008, p. 1070).

> Motivation efficacy: Confidence a coach has in his or her ability to affect the psychological mood and skills of his or her athletes.
>
> Character building efficacy: Confidence a coach has in his or her ability to positively influence the character development of his or her athletes through sport.
>
> Game strategy efficacy: Confidence a coach has in his or her ability to lead during competition.
>
> Technique efficacy: Confidence a coach has with his or her ability to use instructional and diagnostic skills during practices.
>
> Physical conditioning efficacy: Confidence a coach has in his or her ability to prepare her/his athletes physically for participation in his or her sport.

Feltz et al. (1999) developed the coaching efficacy scale that was designed to measure coaches' self-perceptions of their confidence in each of the dimensions. This scale was subsequently revised by Myers, Feltz, Chase et al. (2008) to create the coaching efficacy scale II-high school teams (CES II-HST). In this scale, coaches are asked to rate their confidence in their ability on the five dimensions on a 10-point Likert scale from 0 (not confident at all) to 9 (extremely confident). The CES II-HST underwent rigorous development and testing procedures that resulted in a well-validated and robust measure of coaching efficacy.

The investigations around coaching efficacy have centered on two main features, antecedents and outcomes of efficacy. Feltz et al. (1999) found that experience as a coach, social support, perceived team ability, and previous season's win-loss record were all sources of coaching efficacy. Moreover, outcomes of coaching efficacy were also identified, revealing differences between high- and low-efficacy coaches. Namely, coaches with higher levels of coaching efficacy (1) showed more praise and encouragement, (2) had less instruction and organization, (3) their players reported having higher levels of satisfaction, and (d) their teams had higher levels of performance. Building on this, Myers, Vargas-Tonsing, and Feltz (2005) reported that perceived team ability, social support, career winning percentage, and years coaching were all important sources of coaching efficacy, which in turn predicted coach behaviors (coach self-reported efficacy enhancing behaviors), team satisfaction, and current years winning percentage. The authors also investigated the moderating role of coaches' and athletes' gender interaction, with two interesting findings: first, the impact of social support on coaching efficacy was significantly greater for female than for male coaches. Second, overall team satisfaction did not differ among female teams trained by male or female coaches, suggesting that the coach's gender did not seem to matter to female athletes.

The coaching efficacy construct underpinned the development of coaching competency (Myers, Feltz, Maier, Wolfe, & Reckase, 2006; Myers, Wolfe, Maier, Feltz, & Reckase, 2006). Coaching competence is defined as athletes' evaluation of their coach's ability to affect athletes' learning and performance and contains the same dimensions as the coaching efficacy construct (Myers, Feltz, et al., 2006). The development of the coaching competency construct followed a similar path to that of the efficacy construct in that a fifth dimension (physical condition) was added later. This scale underwent several refinements (Myers et al., 2006; Myers, Wolfe et al., 2006; Myers, Chase, Beauchamp, & Jackson, 2010; Myers, Beauchamp & Chase, 2011) that resulted in the current version, labeled athletes' perceptions of coaching competency scale II-high school teams (APCCS II-HST) (Myers et al., 2010). The APCCS II-HST asks athletes to rate their coach's competency using the following question: "how competent is your head coach in his/her ability to…" on a five-point scale from complete incompetence to complete competence.

Research on the coaching competency construct identified several interesting associations with athlete outcomes. Myers et al. (2006) found a positive association between coach competency and athlete satisfaction. In a later study using a multilevel analysis, Myers, Beauchamp, and Chase (2011) uncovered a significant relationship between coaching competency (measured at the team level) and athlete satisfaction. Extending on the coach competence and athlete satisfaction relationship, Kao and Tsai (2016) demonstrated that coaching competency mediated the relationship between transformational leadership and athlete satisfaction. In yet another study,

Malete, Chow, and Feltz (2013) examined the relationship between coaching efficacy, coaching competency, and antisocial behaviors in a collectivist-oriented culture (Botswana). The results revealed that coaches' game strategy competence was positively related to antisocial behaviors, whereas coaches' character building competence was not. Interestingly, coaching efficacy was not related to antisocial behaviors.

To sum up the coaching efficacy literature, researchers have developed and validated a measure of coaching efficacy along with antecedents (e.g., coach experience, social support, and win-loss record) and consequences (e.g., coach behavior and team satisfaction) of coaching efficacy being identified. Gender was demonstrated as an important variable when considering coaching efficacy. In terms of the coach competency literature, there is a well-established relationship between coach competence and athlete satisfaction, as measured by athlete perceptions of coach competence (Kao & Tsai, 2016; Myers et al., 2006; Myers, Wolfe et al., 2006) and some emerging evidence with regards to coaching competency and athletes' perceptions of transformational leadership (Kao & Tsai, 2016) and athlete perceptions of antisocial behaviors (Malete et al., 2013). However, there is scope to broaden the research agenda and examine a wider range of antecedents and outcome variables within this important, yet relatively new area of research.

Transformational Leadership

Transformational leadership is a style of leadership that is described as inspiring, developmental, values-based, and intellectually simulating (cf., Arthur, Bastardoz, & Eklund, 2017). This style of leadership was originally theorized to cover four broad dimensions of leader behaviors: idealized influence, inspirational motivation, intellectual stimulation, and individualized consideration (Bass, 1985). Transformational leadership is often described in relation to transactional leadership, which is a style of leadership that is based on quid pro quo exchanges underpinned by rewards and sanctions. Transformational leadership is generally hypothesized to predict variance in outcomes over and above the variance accounted for by transactional leadership, what is commonly referred to as the augmentation hypothesis (for a review and meta-analysis, see Judge & Piccolo, 2004). Very prevalent in OB/IO research for decades, transformational leadership theory was only relatively recently introduced to the sport domain by Zacharatos, Barling, and Kelloway (2000). After a relatively slow start, research on transformational leadership in sport has undergone exponential growth since around 2010 (cf., Arthur & Tomsett, 2015; Arthur, Bastardoz, & Eklund, 2017). The research literature within sport thus far has largely emulated the positive results observed in organizational psychology. That is, coaches and/or captains who are rated as being more transformational by their athletes appear to also be those who lead teams on which athletes tend to report strong perception of group cohesion (Callow, Smith, Hardy, Arthur, & Hardy, 2009; Cronin, Arthur, Hardy, & Callow, 2015; Smith, Arthur, Hardy, Callow, & Williams, 2013), more positive communication (Smith et al., 2013), are willing to make more sacrifices for the team (Cronin et al., 2015), invest extra effort (Arthur, Woodman, Ong, Hardy, & Ntoumanis, 2011), have greater levels of need satisfaction and well-being (Stenling & Tafvelin, 2014), are more satisfied (Kao & Tsai, 2016), and have higher levels of performance (Bormann & Rowold, 2016; Bormann, Schulte-Coerne, Diebig, & Rowold, 2016; Charbonneau, Barling, & Kelloway, 2001).

The research in sport has also begun to uncover some mechanisms by which transformational leadership impacts outcomes. For example, communication (Smith et al., 2013) and sacrifice (Cronin et al., 2015) mediate the effect of transformational leadership on cohesion. Charbonneau et al. (2001) identified intrinsic motivation as a mediator of the relationship between transformational leadership and performance, and Stenling and Tafvelin (2014) found that need satisfaction mediates the impact of transformational leadership on athletes' well-being. Researchers have also begun to examine contextual or situational moderators. For example, Arthur et al. (2011) demonstrated that athlete narcissism moderated the relationship between transformational leadership and extra effort. Specifically, Arthur and colleagues reported that the effect of the transformational leader behaviors of fostering acceptance of group goals and high performance expectations on athlete self-reported motivation was weaker for those scoring higher in narcissism than those scoring lower in narcissism. In a three-way interactional design, Bormann et al. (2016) examined the interaction between transformational leadership, team performance, and win orientation on player performance. The results of this study demonstrated the importance of environmental contingencies (team performance) and individual differences (player's motivation for winning) in determining the relationship between transformational leadership and player's individual performance levels. Within the limitations of transformational theory (see Endogeneity section), we encourage future research that will untangle the underlying influence process and boundary conditions of the transformational leadership construct.

Drawing from the OB/IO literature and the accumulated evidence in sport, transformational leadership can be trained. Developmental interventions targeted at leader's behavior change were associated with enhanced

follower outcomes across a range of contexts that includes military (Arthur & Hardy, 2014; Dvir et al., 2002; Hardy et al., 2010), banking (Barling et al., 1996), physical education teachers (Beauchamp, Barling, & Morton, 2011), and health care (Mullen & Kelloway, 2009). To our knowledge, only one intervention study was conducted in sport (Vella, Oades, & Crowe, 2013), an important first step in the sport literature; yet, the design prevents drawing any hasty conclusion (i.e., the control group received no controlled treatment). Nevertheless, we strongly encourage researchers to develop transformational leadership interventions targeted specifically at a sport context.

A range of different measurement instruments has been used to measure transformational leadership within sport. The differentiated transformational leadership inventory (Hardy et al., 2010) and the multifactor leadership questionnaire (MLQ) (Bass & Avolio, 1995) are the scales predominantly used, and other measures including the TLI (Podsakoff, MacKenzie, Moorman, & Fetter, 1990) are used to a lesser extent. Transformational leadership has been operationalized both as a global and as a differentiated construct (i.e., treating the four different dimensions as separate), with the relative merits of each approach being discussed elsewhere (e.g., Arthur & Tomsett, 2015). Yet, creating composite (or index) scores of conceptually distinct factors is not appropriate despite the high correlations between the factors (often used to justify creating global index scores, see Arthur et al., 2017). We also refer readers to our Endogeneity section about the harmful effects of using scales in cross-sectional, common-method designs.

In addition to quantitative inquiries, researchers have also conducted qualitative investigations (e.g., Mills & Boardley, 2016; Newland, Newton, Podlog, Legg, & Tanner, 2015; Smith et al., 2017). In contrast to the quantitative literature, researchers using qualitative methods have typically done so in elite and professional contexts (but not only; see Newland et al., 2015). In their study involving the All Blacks rugby team, Hodge, Henry, and Smith (2014) reported that transformational leadership and aspects of the vision, support, and challenge model (Arthur, Hardy, & Woodman, 2012) emerged in their data of a world class sports team. Furthermore, Mills and Boardley (2016) investigated a sample of England premier division football managers, and Smith et al. (2017) focused on a sample consisting of professional English cricketers. In the latter study, Smith and colleagues recruited professional cricket players from First Class Counties in England and Wales emanating from teams where their coach and captain had previously been rated as transformational. Their findings offer hindsight in terms of *how* and *what* transformational leaders do in elite sport. Even limited, qualitative investigation into the transformational leadership construct has to date also yielded support for its relevance within the sport environment.

In relation to transformational leadership theory, the Vision, Support, and Challenge model (VSC) has also been developed as an applied meta-cognitive model of the primary mechanisms by which transformational leaders are proposed to have their effect (Arthur & Lynn, 2017; Arthur, Hardy, & Woodman, 2012). Introduced in a military training context, Hardy et al. (2010) developed the VSC model in order to simplify the theory and make it more accessible to leaders. The VSC model has been used to underpin two interventions studies within a military context (Arthur & Hardy, 2014; Hardy et al., 2010), but as yet the model is untested in sport. The VSC model depicts an explanatory process of leadership whereby VSC are key mechanisms by which leader effects are transmitted. The VSC model therefore has potential to shed light within the sport leadership literature by contributing toward developing knowledge on the explanatory processes within leadership research.

Recently, Mills and Boardley (2017) developed an implicit association test to measure what they referred to as transformational leadership integrity. They defined transformational leadership integrity as "…the consistency in thought and action to the principles associated with both: (1) True transformational leadership, and (2) pseudo-transformational leadership" (p. 34). Mills and Boardley (2017)—building on the notions of true, authentic, and pseudo-transformational leadership—suggest that transformational leadership integrity should not be measured by follower reports because leaders may conceal their integrity and attitudes but still display overt behaviors that are consistent with transformational leadership. They suggest measuring integrity via indirect implicit association tests, whereby reaction times to salient stimuli are purported to tap the underling attitude. Although at its very early stages within sport, the application of implicit measures may potentially inform the transformational leadership literature in sport.

Overall, the study of transformational leadership in sport has made a substantial contribution to our understanding of leadership. However, as we describe in the next section, there are serious limitations related to the typical cross-sectional research design and tautological definitions. Based on a recent review (Arthur et al., 2017), we use transformational leadership to highlight these issues; yet, these issues also have bearing across other leadership models developed and applied in sport. Thus, we encourage researchers to reflect on how these issues may affect their theories and empirical knowledge related to sport leadership.

Endogeneity

The previous sections highlighted the long history of sport leadership research and offered hindsight gleaned from decades of thinking and empirical tests. Despite intuitively appealing hypotheses, our field lacks proper empirical assessment of the theories. In the current section, we question the current methodological standards that inherently impede our understanding of the leadership phenomenon. We also offer solutions that will hopefully help researchers in sport develop more causal knowledge. We start by discussing the term endogeneity, a prevalent "disease" (see Antonakis, 2017) that was first introduced in social sciences about 40 years ago, the period referred to as the beginning of the "identification revolution" in economics. The consequences of the endogeneity problem have been discussed recently in some management disciplines (e.g., Bettis, Gambardella, Helfat, & Mitchell, 2014; Guide & Ketokivi, 2015; Hamilton & Nickerson, 2003; Reeb, Sakakibara, & Mahmood, 2012).

In simple terms, endogeneity is a threat to the statistical validity of the results; that is, the findings based on quantitative inquiries affected by endogeneity are confounded to some degree. To understand the true meaning of endogeneity, we review the basics of statistical analysis. Without going into too much detail, a critical assumption of Ordinary Least Squares (OLS), of which Analyses of Variance (ANOVA) is a special case, and Maximum Likelihood (ML) estimators relate to the exogeneity of regressors. Assume that we gather data to test the effect of a leadership style (x_i) of a coach on an outcome variable such as followers' satisfaction with the coach (y_i). We would use the following equation to test the relationship: $y_i = b_0 + b_1 x_i + e_i$. The exogeneity assumption requires that the error term e_i, representing all unmeasured (or omitted) causes affecting the dependent variable y_i, be uncorrelated with included regressor(s) such that the correlation(x_i, e_i) approximates zero (Kennedy, 2003). The OLS estimator derives parameters b_0 and b_1 in order to minimize the sum of squared residuals (equal to the model predicted values minus the observed values) and by construction automatically makes e_i orthogonal to x_i (Ketokivi & McIntosh, 2017). Should the correlation (x_i, e_i) not be close to zero, the exogeneity (also referred to as orthogonality) assumption will not be satisfied and thus endogeneity would be present (Antonakis et al., 2010; Antonakis, Bendahan, Jacquart, & Lalive, 2014b). Because the OLS estimator assumes a null correlation between predictor(s) and unmeasured causes, endogeneity affects parameter estimates and renders the coefficients b_0 and b_1 inconsistent; as such, the estimated coefficient of interest b_1 does not reflect the true effect of a coach's leadership style (x_i) on followers' satisfaction with the coach (y_i). Note also that researchers cannot even interpret the correlation (x_i, y_i) when endogeneity is present (Antonakis et al., 2010).

Researchers must thus ensure that relevant omitted causes, which would otherwise be pooled in the error term e_i, are modeled so that e_i does not correlate with included predictors. Endogeneity is a matter of extent and so will be more or less a threat to the consistency of the results depending on how serious it is (Antonakis, 2017; Ketokivi & McIntosh, 2017). The higher the correlation (x_i, e_i), the higher the biases will be in estimated coefficients, which ultimately prevent knowledge formation (Antonakis et al., 2010). However, unmeasured causes are by definition not measured, and thus, we neither observe nor can compute the correlation (x_i, e_i) (Roberts & Whited, 2013). We can only infer this correlation by comparing an estimate thought to be correct (i.e., a consistent estimator) to one that potentially omits the important causes (i.e., an efficient estimator). Thus, any discussion about the effect of endogeneity should be framed in terms of "choices" or "dilemma" rather than as a definite "problem" (Ketokivi & McIntosh, 2017).

Yet, the problem is very serious because endogeneity-plagued estimates simply cannot be used to inform leaders and policy-makers (Antonakis, Bendahan, Jacquart, & Lalive, 2010). If policy is what we aim to influence, then dealing decisively with the endogeneity problem should be an imperative. This is important for journal editors and authors because quantitative studies affected by endogeneity tend to receive fewer attention and citations (Antonakis, Bastardoz, Liu, & Schriesheim, 2014a). We will see in a later section that the method of choice to deal with endogeneity allows researchers to somehow approximate whether endogeneity is a major source of concern in the data. We discuss the major potential reasons leading to endogeneity in the following section.

Reasons for Violation of the Exogeneity Assumption

Potential features of research design leading to endogeneity are common across many disciplines such as leadership or more generally management. Our goal here is to discuss what we have identified as critical causes leading to endogeneity in sport leadership research, which include: (1) omitted variables bias; (2) common-method variance; (3) simultaneity and reverse causality; (4) measurement errors; and (5) omitted selection. We do not aim to prove that these causes lead to inconsistent estimates; we refer interested readers to other publications that have discussed these issues in more detail (Antonakis et al., 2010, 2014b). To illustrate the five causes of endogeneity, we refer to a simple

example where a coach's leadership style (e.g., transformational leadership) is assumed to predict players' satisfaction with the coach.

1. Omitted Variable Bias

An estimated relationship between a regressor and a criterion will be endogenous if there exists a third variable correlating both with the regressor and the criterion that is not included in the regression model. Finding examples for an unmeasured predictor of players' satisfaction that is correlated with coaches' transformational leadership is straightforward, and one could think of other leadership styles that are part of the full-range model. Transactional (provision of rewards and sanctions) and instrumental (leadership based on expertise) leadership styles are both correlated with a transformational leadership style and will also be significantly related to followers' satisfaction (Antonakis & House, 2014). Omitting these leadership styles can change (1) the magnitude of the effect from significant to non-significant or vice-versa, and (2) the direction of the effect from positive to negative or vice-versa. Other omitted variables may include other leadership styles such as task-oriented leadership (see Banks, McCauley, Gardner, & Guler, 2016 for an incremental validity test of competing styles), trust or other affect-related attitudes felt by the subordinates (e.g., the more players like their coach, the more satisfied they will be with her and the more positive their rating of her leadership), or the personality of the player (e.g., narcissistic players will not be satisfied and rate negatively the coach's leadership style).

Because the omitted variable is probably the most important cause for inconsistent estimates, we strongly encourage researchers to measure relevant control variables. Non-significant control variables, particularly if they do not change estimates of key regression (Antonakis et al., 2010), can safely be dismissed from the regression because they would not cause endogeneity (Jacquart, Cole, Gabriel, Koopman, & Rosen, 2017). Although controlling for all variables will never be practically possible, any reasoned attempt to measure the true effect of a coach's leadership should at least include (1) one or more competing leadership style, (2) some coach's personality dimensions, and (3) be free from attributions bias.

2. Common-Method Variance

Common-method variance has been defined by Richardson, Simmering, and Sturman (2009) to indicate "systematic error variance shared among variables measured with and introduced as a function of the same method and/or source" (p. 763). In other words, it refers to a situation where variables at different stages of the estimated model (e.g., independent variable, mediator, and/or dependent variable) are measured using the same method and/or the same source, leading to measured relationships being vastly inflated (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Podsakoff & Organ, 1986). In our example, imagine that a researcher gathers data by asking players to fill two 6-item, 7-point Likert scales about their coach's transformational behavior and their own satisfaction with the coach. This example indicates both a common-method (leadership style and satisfaction are measured with scales that have similar response scales) and common-source (both data points are provided by the same participant) variance, and it actually portrays very well the current literature on transformational leadership in sports (Arthur et al., 2017).

Various cognitive mechanisms explain why this phenomenon is problematic. First, a halo effect generally tends to bias attributions, because the general impressions that players have of their coach will taint players' ratings. Trust, liking, attractiveness, or effectiveness are examples that could affect players' general impressions of the coach (Antonakis, Bastardoz, Jacquart, & Shamir, 2016). Second and similarly, if a coach is cognitively categorized as leader-like (using implicit leadership theories; see Lord, Foti, & De Vader, 1984), raters may attribute prototypical characteristics in line with the category (e.g., good or bad coach) and may even "fill-in-the-blank" to ensure ratings are consistent with the categorization (Antonakis & House, 2014; Cantor & Mischel, 1977). Third, knowledge of the performance of the coach may also trigger alignment in ratings, what is called "performance-cue" effects (Jacquart & Antonakis, 2015; Lord, Binning, Rush, & Thomas, 1978). Because a coach is successful, it is likely that players will both rate her as transformational and be satisfied with her; here the performance of the coach (or the team) would play as an omitted variable (see omitted variable section). Such attributional bias was the primary reason why "skeptics," in the 1980s, severely criticized the leadership field, arguing that leadership was nothing more than a social construction (Meindl & Ehrlich, 1987). Finally, individuals usually strive to keep consistency in their evaluation (Podsakoff et al., 2003; Salancik & Pfeffer, 1977), so that if I rate my coach as poorly stimulating and not inspiring or considerate, it is unlikely that I will report being satisfied with her.

Whereas some have argued that common-method effects are largely overstated and may be more of an urban legend (Spector, 2006), we are of the opinion that its effects may potentially be pervasive. Unfortunately, we do not have a magic wand. As we have discussed, the effects of common-method, common-source variance are manifold and cannot simply be solved with some popular statistical procedure, such as modeling a latent factor or by performing a Harman single factor

test (see Antonakis et al., 2010; Podsakoff et al., 2003). We do not encourage the use of split-sample designs because reducing the sample size by a factor of two entails increases in estimates' sample bias, reduced efficiency, lower power and so a higher chance of committing a Type II error rate (i.e., not rejecting the null hypotheses when it should be rejected). Rather, we call for carefully planned study designs. Authors should in advance wonder how they could measure their variable using other methods (e.g., coding the coach transformational behaviors) or other sources (e.g., asking close relatives such as spouses or friends about the target's player satisfaction with his or her coach).

3. Simultaneity and Reverse Causality

Simultaneous effects arise when the effect of a variable x on a variable y almost instantaneously provoke an effect of y on x, whereas reverse causality describes a situation where the dependent variable y is in fact a predictor of the independent variable x (Antonakis et al., 2010). In our example, it is possible that players who are satisfied with their coach will be more attentive to their coach's advice and respect the team instructions, which would in turn foster coaches to perform such behaviors as being developmental or trying to be inspirationally motivating (i.e., dimensions of transformational leadership). This feedback loop example is typical of leadership studies, in that situational or contextual aspects can sometimes be the main drivers of leaders' behaviors (e.g., Shamir & Howell, 1999)

Because most of the transformational leadership research is cross-sectional (Arthur et al., 2017), dismissing simultaneous effects or reverse causality can become a tricky endeavor. In fact, a common explanation used by researchers to justify the direction of causality relate to the order of measurement (i.e., using x_{t-1} to predict y_t). Although it is a necessary first step, this is not sufficient to justify causality because it may well be that y_{t-1} predicts (and so is correlated with) x_{t-1} and will be correlated with y_t as well (Fischer, Dietz, & Antonakis, 2017). Although we hope to see well-designed longitudinal studies that could provide credible hints about the temporal order of effect, we can only advise researchers willing to truly untangle the causality link in their constructs to use experimental procedures.

4. Measurement Errors

Endogeneity arises due to measurement errors because constructs that are measured imperfectly are actually modeled as if they were perfectly measured (i.e., without errors). This source of endogeneity is typical of social sciences where the majority of our constructs are "latent," in that they are neither directly observable nor measurable (Antonakis et al., 2014b). Instead, we develop items and scales to approximate constructs of interest. Our observations of published studies indicate a general trend to form composite variables with items stemming from the same scale (be it a dimension or factor) and use this aggregate variable in an OLS regression analysis. This procedure is problematic because the derived coefficient b_1 was estimated assuming no measurement errors in the variable. The extent to which unmodeled measurement errors affect estimates consistency depends on the reliability of measures: more unreliable measures are more problematic and lead to severely inconsistent estimates, whereas almost perfectly reliable measure can be considered free from endogeneity threat and estimates can be deemed as good (Ketokivi & McIntosh, 2017). Because no rules of thumb can be broad enough to be applicable to any sample (Lance, Butts, & Michels, 2006), we will not venture into giving one single threshold applicable for every situation or sample. We would though suggest that reliabilities above .90 should not pose a major endogeneity threat, whereas reliabilities below .70 would almost surely make measurement errors an issue (Note: we advise against using those values as broad rules of thumb because the effects of measurement errors may vary across many design and study aspects).

To correctly model the relation between a latent variable and its proposed items, a method of choice emerged about 50 years ago: Structural Equation Modeling (SEM) (Jöreskog, 1967). SEM allows the differentiation of item variance that pertains to the construct (i.e., the part of variance in the item that is due to the latent variable) and error variance (i.e., the part of unexplained variance in the item) (Kline, 2016). We can only recommend this method to any researchers dealing with latent constructs measured with multiple items. Another potential estimation procedure is called "errors-in-variable" regression. This technique allows researchers to use composite variables and to model the reliability of the composite variable, using composite reliability, test-retest reliabilities, or by deriving theoretical estimate (Antonakis et al., 2010; Bollen, 1989). Although calls have been made to develop better (i.e., more reliable) scales, the capabilities of SEM or errors-in-variable regression allow researchers to model even low reliability scale. Note that even if reliabilities are above .90, we still encourage researchers to correct for measurement errors: pragmatically, SEM or errors-in-variables regressions are easily accessible in most statistical software nowadays.

5. Omitted Selection

Endogeneity can appear when the selection mechanism is not random, and so observations high on the predictor variable differ significantly from observations that are low on the predictor variable. Imagine that coaches high on transformational leadership have been carefully

selected, developed, and are given plenty of resources within their club, whereas coaches low on transformational leadership emerged from inefficient and poor sport organizations. Such selection mechanism would also probably predict the players' satisfaction with their coach. So, if selection is not specifically measured and included in the regression equation, it will be pooled in the error term e_i and will engender inconsistent estimates.

Even though we recognize that we usually work with convenient samples, we encourage researchers whenever possible to select random representative samples (Antonakis et al., 2010) instead of self-selected ("snowball") samples that are prone to various biases (Marcus, Weigelt, Hergert, Gurt, & Gelléri, 2016). Researchers (and reviewers alike) suspecting endogeneity due to non-random assignment should at best model a proxy of the selection mechanism (e.g., one could use fixed-effects for clubs) and at worse discuss the impact of an endogenous sampling procedure (Antonakis, 2017).

An Assessment of the Sport Leadership Literature

Sources of endogeneity findings are not exclusive to the sport leadership literature; indeed, some fields are currently at a turning point with respect to how they approach science (e.g., some top journals reject quantitative manuscripts on the spot if they report cross-sectional studies without discussing endogeneity and causality threats; see Antonakis, 2017; Guide & Ketokivi, 2015). We believe that our field should not lag behind in the process and embrace this path as well. Antonakis and colleagues (2016) argue that "[r]esearchers use questionnaire measures probably out of convenience, simply because they have been trained to do so, or because everyone else does it. It is a 'quick and dirty' way to obtain data" (p. 307). We are firmly convinced that cross-sectional studies that use perceptual measures from a single source—ironically both the prototypical study leading to endogeneity and the prototypical study published in the transformational leadership literature in sport (Arthur et al., 2017)—have reached their heyday and will not be relied upon in the future.

Even though our tone may seem critical, we are not saying that the accumulated empirical knowledge thus far is incorrect (though it may, or may not, be). However, a shift in researchers' designs and methods is sorely needed if we want to produce causally relevant knowledge. Because, in the end, it is causality that we should truly care about; only causal knowledge can inform policy-makers, sport organizations, future coaches, and youth educators with respect to the prescribed leadership behaviors that one should adopt in a specific context. In the following pages, we offer different ways for researchers in sport leadership to increase the validity and the causal claims' strength of their results, starting with the gold standard: randomized experiments.

Ways to Ensure Internal Validity of Studies

Randomized Experiments

Randomized experiments are the "gold standard" in terms of internal validity and causality because when properly performed, the changes observed on the outcome variable(s) can only be due to the difference in treatments administered. The logic behind experiments is the idea of counterfactuals: What would we observe on the dependent variable for the treated observations had they not received the treatment? The control group serves as the counterfactual for the treatment group (and vice-versa).

At the core of this method lies the randomization of observations (generally individuals but it can be any entity) to group in order to create control and treatment groups that are, initially equal on average. Assuming that the randomization worked as intended, the two groups will be equal on any observable or unobservable covariates that may also predict the outcome variable (e.g., men, women, extroverted, introverted, narcissistic, intelligent). So if we observe an effect on the outcome, the only reason that could explain this difference is the treatment; or in other words, the internal validity of the results is ensured because there is no confusion that could affect or explain our results (Shadish, Cook, & Campbell, 2002). Thus, even if being a women may cause the outcome to increase, there will be an equal amount of women in both groups, and so, being a women does not correlate with the treatment variable.

To discuss our example about testing the effect of coaches' transformational leadership on satisfaction with coaches, we could relatively easily test this idea in a field experiment. We could recruit some current coaches, randomize them in two groups (and even pre-measure some covariates of interest to correct for unbalanced matching), and offer two different trainings on leadership. In order to provide a fair comparison (Cooper & Richardson, 1986) and avoid potential demand effects (Antonakis, 2017), coaches assigned to the control condition should receive a credible treatment that would ensure their motivation and their willingness to participate in the study is affected in a similar way as coaches in the experimental treatment. Such control treatment could be a general leadership development program. Leaders in the treatment group would receive a training targeted at developing transformational behaviors (Dvir, Eden, Avolio, & Shamir, 2002; Hardy et al., 2010). Weeks or months later—letting sufficient time for an

effect to appear—we could collect players' satisfaction with their coach using single items or scales (measuring dependent variables with scales is not a problem; see Antonakis et al., 2016). We would estimate the following equation: $y_i = b_0 + b_1 t_i + e_i$ where t_i would indicate whether a coach was in the control or treatment group. If the random assignment creates equal groups on average, the exogeneity assumption is respected because the correlation between the explanatory variable t_i and the error term e_i is close to zero.

As a way to make experiments even more consequential, we encourage researchers to use incentivized performance whenever this may affect the pattern of results (Antonakis, 2017). In the example so far, we are of the idea that incentivizing the performance would not significantly alter the behavior or satisfaction of player. But imagine now that the outcome of interest is the performance of the team: in this context, the performance of players should be incentivized at the individual or group level to ensure that any effect of the transformational leadership style is relevant beyond participants' self-interest to maximize their well-being (Antonakis, d'Adda, Weber, & Zehnder, 2015). Indeed, incentivized experiments are the backbone in economics that ensure participants make consequential decisions in their experiments (see Zehnder, Herz, & Bonardi, 2017).

Finally, ensuring the external validity of finding is critical for randomized experiments. To be relevant for policy-makers or sports organizations, implications derived from experiments have to be generalizable to other samples in other contexts. All things being equal, field experiments increase the external validity in comparison to laboratory experiments because they take place in an environment that resembles real life (see Eden, 2017 for a review of field experiments). However, external validity often comes at the cost of reduced internal validity because the high ecological environment in which the field experiment takes place cannot be fully controlled by the researcher, and so, other variables may affect the results in one direction or the other. In fact, the context of the experiment may be far less important for the generalizability of the findings than the reliability of the operationalized manipulation, what Highhouse (2009) termed "domain representativeness."

Quasi-Experiments

Quasi-experiments are "almost" experiments in that a treatment affects one group but not another (Shadish et al., 2002). The subtle, though important, difference with randomized experiment is that observations are not randomly assigned to conditions, which can create biased counterfactuals. If the researcher is aware of this selection issue, this threat can be minimized. We only cover here two quasi-experimental designs we believe could be easily applied to the sport leadership literature, namely, the Regression Discontinuity Design (RDD) and the Difference-in-Difference model. Other experimental procedures that try to model an endogenous (i.e., not random) selection mechanism include the Heckman selection model (Heckman, 1977) and the propensity score matching (Caliendo & Kopeinig, 2008; Li, 2013).

The quasi-experimental design that very closely approximates the randomized experiment is the regression discontinuity design, or RDD (Cook, 2008). This design has strong internal validity because observations are assigned to different conditions based on an observable and quantifiable variable (Bastardoz, Jacquart, & Antonakis, 2017; Mellor & Mark, 1998). Imagine that your university ethics approval committee refuses that you randomize participants to a transformational or general leadership development program. In this case, you could measure coaches' leadership abilities before the experiment and then assign coaches to the conditions based on their score on the leadership assessment (and also provide the theorized efficient training to those who need it most). Those with a score below the cutoff, that could be set at the mean, the median, or any other relevant value, would receive the transformational leadership program, whereas those above the cutoff would receive the general leadership program. Because the cutoff delineates observations in the control from observations in the treatment group, the cutoff is generally referred to as the discontinuity (hence the design's name). To correctly model the selection mechanism, researchers should simply include the scores of the leadership assessment as a predictor of the outcome, which can be any variable that is relevant to both treated and control groups (e.g., it can be players' satisfaction with coach, or it can be a post-treatment assessment of leadership abilities). Researchers can also apply this design whenever a random shock such as a crisis, a death, or a change in ownership happens, and here the assignment variable is time.

The second quasi-experimental procedures is known under various names: untreated control group design with pre- and post-test for psychologists (Shadish et al., 2002) or difference-in-difference model in economics (Angrist & Pischke, 2008; Antonakis et al., 2010). This procedure compares how different groups react to different treatments over time and whether their reaction differs. In a sense, this design resembles the RDD because it tries to create an appropriate counterfactual group against which to compare the treated group (Jacquart et al., 2017). Suppose you are given access to two basketball clubs in a similar setting, and you wish to test whether your transformational leadership development program works. Because assigning randomly participants within clubs could not be an option

(e.g., clubs are too small and ripple effects could not be avoided), you could implement the different programs in the different clubs. A difference-in-difference model would compare the difference in players' satisfaction (satisfaction post-treatment minus satisfaction pre-treatment) in the treatment group compared to the control group. Although this procedure seems in theory appealing, it requires that both clubs be similar on various dimensions (e.g., size, budget, selection, quality of infrastructure, to name a few), which may be a too strong assumption at times (for a similar experiment, see Greenberg, 1990).

Of course, we perfectly understand that all constructs are not amenable to experimental manipulation, and we also see limits in the applicability of quasi-experimental procedures. But we encourage researchers to think creatively about the applicability of this causally relevant methods and not to fall prey to convenience and conformity bias that could hasten the publication process. To avoid any misunderstanding, we want to emphasize that available methods should not drive the exploration of theoretically relevant questions. Ketokivi and McIntosh (2017) rightfully argue that "(c)hoosing models and explanatory variables based on whether they can be credibly manipulated in an experiment seems like putting the cart before the horse" (p. 5). For some constructs or research questions, it will simply not be possible or appropriate to run experiments. Also, in the early stages of theory development and testing, it will be more appropriate to first establish the tenability of a model with cross-sectional and/or prospective paradigms, using preliminary evidence to proceed further to causal testing. Cross-sectional paradigms are still very much appropriate, but in these cases, researchers will have to use an estimation procedure that appropriately tests models and gets consistent estimates. This statistical method relying on instrumental variable is briefly presented next.

Tackle Endogeneity with Instrumental Variables

Two-stages-least-squares estimation is a very useful procedure that allows consistent estimation of the effect of an endogenous variable on an outcome variable. This statistical method requires a variable called an *instrument*—an exogenous variable—that is used to eliminate endogeneity (Antonakis et al., 2010; Bascle, 2008; Cameron & Trivedi, 2005). An important caveat is that finding and measuring good instrumental variable(s) may in some cases be very tricky. To qualify as a good instrumental variable, a variable z_i should respect three conditions (Antonakis et al., 2014b; Ketokivi & McIntosh, 2017): (1) z_i should vary randomly in nature or at least be exogenous to the outcome variable y_i; (2) z_i should only affect the outcome variable y_i through its effect on the endogenous variable x_i (i.e., z_i should be unaffected by unmeasured variables predicting the outcome variable y_i; the exclusion condition); and (3) z_i should be strongly and significantly correlated with x_i, the endogenous variable that is being instrumented (the relevance condition).

Assume the following true relationships are:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$x_i = \alpha_0 + \alpha_1 z_i + \nu_i$$

As its name indicates, the estimation procedure is performed in two stages. In the first stage, the endogenous variable x_i is regressed on the instrumental variable z_i so as to get OLS estimates of α_0 and α_1. A usual rule of thumb to assess whether the relevance condition is respected is that the F-test of the first-stage equation be higher than 10 (Staiger & Stock, 1997). If this is the case, researchers can then compute the predicted values $\hat{x}_i$ which in this simple case refers to α_0 plus the product of α_1 with z_i. In the second stage, the outcome variable y_i is regressed on the predicted values $\hat{x}_i$ which gives the 2SLS estimators (i.e., b_0 and b_1) of the true coefficients β_0 and β_1 (note, a correction to standard errors is required to ensure appropriate inferences). Because a good instrumental variable z_i should be uncorrelated with ε_i (the exclusion condition), the predicted value $\hat{x}_i$ will be too, and so endogeneity threats can be avoided. If one has multiple instrumental variables, an important requirement is that at least one instrument per endogenous variable used in the first-stage regression should be excluded from the second-stage regression (order condition; Antonakis et al., 2014b; Wooldridge, 2002).

To portray our explanation of instrumental variable, let's return to our example. From our assessment of the sources of endogeneity, we have identified transformational leadership (x_i) as potentially endogenous to the player's satisfaction (y_i). Because we anticipated this state of affairs before gathering the data, we also collected coaches' personality, sex, and intelligence (different z_i) that should theoretically be strong instruments. Meta-analyses have indicated that transformational leadership is correlated with personality factors such as extraversion, openness to experience and emotional stability (Bono & Judge, 2004; Judge, Bono, Ilies, & Gerhardt, 2002), gender (Eagly, Johannesen-Schmidt, & Van Engen, 2003), and intelligence (Judge, Colbert, & Ilies, 2004). Furthermore, these variables vary to some extent randomly in nature and should not be affected by unmeasured factors explaining players' satisfaction, so the correlation between z_i and ε_i is close to zero. Also, we

make the assumption that the only reason why coaches' personality, gender, and intelligence would affect players' satisfaction is through the display of transformational behaviors (this assumption may at times be too strong). Thus, in the first step, we would regress players' transformational ratings of coaches on the proposed instruments. We would test whether the appropriateness of these instruments within our particular sample using the F-test of this regression (note that the strength of instruments may vary across samples). Based on the estimates of the first stage, we would then compute the predicted values $\hat{x}_i$ based on the first-stage model for coaches' transformational leadership. In this second step, we would regress the player's satisfaction (y_i) on coaches' transformational leadership predicted values $\hat{x}_i$. In the second step, we could include some instruments such as personality dimensions, gender on intelligence (especially if they are predictors of players' satisfaction beyond their effect on transformational leadership) as long as at least one instrument is excluded from the second-stage equation.

This estimation procedure can be very useful when it comes to estimating mediation models (think of z_i as experimental condition(s) and x_i as mediator variables), because instruments can help disentangle the true causal effect of the mediator (which is generally always endogenous) on the outcome variable (Antonakis et al., 2014b; Shaver, 2005). In experimental research, manipulated conditions generally make for very good instruments because they are theoretically exogenous and should be strong predictors of the mediators. As long as instruments satisfy the exclusion and relevance conditions, researchers can be creative and cast a wide net in their search for instruments. In the presence of strong instrumental variables, one can perform a Hausman (1978) test to check whether the coefficient b_1 from the OLS estimation significantly differs from the coefficient b_1 stemming from the 2SLS estimation; if that is the case, this generally indicates that the endogenous variable is indeed endogenous (Antonakis et al., 2010, 2014b).

Yet we want to add a word of caution: weak instruments—correlating only weakly with (i.e., being poor predictors of) the endogenous variable—can be a cure worse than the disease and be very misleading. Weak instruments will cause more problems than they solve (Ketokivi & McIntosh, 2017). In some instance, researchers will be aptly advised to stick to OLS estimates, when they do not have strong instrumental variables despite cautious measurement and search (Murray, 2006) and when their theoretical assessment of their statistical model does not suggest strong endogeneity threats (Ketokivi & McIntosh, 2017). We are fully aware that finding good instruments will at times be difficult, which is probably the main limitation of the instrumental variable method. We believe that finding good instruments is the responsibility of authors, not editors or reviewers. If authors cannot find good instruments, this is not a sufficient reason to ignore endogeneity issues. Should an ex-ante assessment of endogeneity indicate serious threats, we encourage researchers to think carefully about the design and variables that will be measured. Planning carefully is key, and good science is probably best served this way.

Our aim is not to set unrealistically high standards that even the best-intentioned researcher could not apply in his or her own research; rather, we call for authors to take heed of and critically discuss how endogeneity can affect their estimated coefficients and threaten their statements about causality. Endogeneity is a matter of degree, and in some instances endogeneity will be a meaningful problem that cannot be disregarded (Antonakis, 2017). To conclude, we again echo Ketokivi and McIntosh (2017) in their assessment that "endogeneity should not be thought of as a yes/no issue. If we want to turn it into a simple dichotomy, the answer is clear: endogeneity is always a problem. But this is just stating the obvious. Instead, we should seek to examine whether endogeneity is so severe that it plausibly constitutes a problem insofar as the objectives of the inquiry are concerned" (p. 10).

The Way to Get the Gist of Leaders' Behaviors

Another possibility to limit endogeneity issues that have a rich history in the sport leadership literature refers to the coding of objective behavior. The mediational model of sport leadership (Smoll & Smith, 1989; Smoll, Smith, Curtis, & Hunt, 1978), alongside the development of the coach behavior assessment system (Smith, Smoll, & Hunt, 1977), emphasized the appraisal of how leaders in sports behave. In addition to the CBAS, other coach observational systems have been developed such as the Arizona State University observation instrument (Lacy & Darst, 1984) and the coach analysis intervention system (Cushion, Harvey, Muir, & Nelson, 2012). The observational methodology in sport has also been applied to code the behaviors of expert coaches. For example, Tharp and Gallimore (1976) and Gallimore and Tharp (2004) observed and coded the behaviors of the basketball coach John Wooden; Bloom, Crumpton, and Anderson (1999) observed another basketball coach, Jerry Tarkanian; and Becker and Wrisberg (2008) observed yet another basketball coach, Pat Summit (for a recent review of observational measure used in sport, see Cope, Partington, & Harvey, 2017). The aforementioned examples being notable exceptions unfortunately, the majority of the research in sport leadership followed

the trend also found in other disciplines (Van Knippenberg & Sitkin, 2013) toward the proliferation of follower report scales, probably because they are cheap, easy to use, and lots of data can be collected at the same time (Jacquart et al., 2017). Without an intent to step backward, we wish to re-encourage the use of objectively coded data. Leaders' behaviors coded by properly trained raters will generate variables freed from common-method variance in a context that has high ecological validity. Also, behavioral ratings are less tainted by implicit leadership theories than general leadership ratings (Gioia & Sims Jr, 1985).

For researchers who cannot get access to changing rooms or training pitches where unobtrusive coding is easily performed, the development of the Internet and social media may offer content across various channels: press conferences, club statements, official communication, video-based interviews or training session, or even using archival data (Barnes, Dang, Leavitt, Guarana, & Uhlmann, 2015). Computer-aided text analysis can efficiently replace human coders (Short, Broberg, Cogliser, & Brigham, 2010), and we also foresee a trend toward more big-data analyses (Tonidandel, King, & Cortina, 2018).

An important aspect to ensure unbiased ratings is that coders should be unaware of the outcome of leaders' behaviors, because coders may attribute specious behaviors to leaders simply by being knowledgeable about the outcome (Antonakis et al., 2016; Arthur et al., 2017). Also, we encourage any researchers developing a new coding scheme to ingrain the behaviors into a clear theoretical framework. The I/O-OB fields have much to offer in this aspect, because various theories could inform our understanding of sport leadership such as trait activation theory (Tett & Burnett, 2003), leader-follower distance (Antonakis & Atwater, 2002; Shamir, 1995), or moral foundations theory (Graham, Haidt, & Nosek, 2009). Nevertheless, any new behavioral theory should clearly delineate the leaders' behaviors from their effects, otherwise tautological definitions—where the leaders' behaviors are by definition true—will continue to bestrew our field. We discuss this in the next section.

Tautological Definitions

Tautological definitions have recently been under the radar in the leadership literature (Antonakis et al., 2016; Van Knippenberg & Sitkin, 2013). A tautology indicates "unnecessary repetition, usually in close proximity, of the same word, phrase, idea or argument" (Oxford English Dictionary, 2007). Because a tautological definition is almost true by definition, it cannot be falsifiable. Tautological definitions are problematic because they do not increase our understanding of phenomena (MacKenzie, 2003), and poorly defined constructs impede the development of research streams (Antonakis, 2017; Antonakis et al., 2016). Indeed, the transformational leadership construct is plagued with tautological statements equating transformational and effective leadership, and research on transformational leadership in sport is not immune from this (Arthur et al., 2017). Furthermore, transformational leadership, the very name of which indicates some transformation (for the better), is a highly loaded definition (Antonakis et al., 2016; Van Knippenberg & Sitkin, 2013).

On top of tautological definitions, operationalizations of constructs reflect this confounded theorizing (Antonakis, 2017). When Bass (1985) introduced the transformational leadership construct in management research, no definition of the construct was given, and he used the multifactor leadership questionnaire scale (Bass & Avolio, 1995) to describe how and what a transformational leader does (Antonakis et al., 2016). These circumstances explain why most MLQ items confound behaviors of transformational leaders with outcomes (see Arthur et al., 2017; Van Knippenberg & Sitkin, 2013 for tangible examples). It is one thing to say that transformational leader should transform, inspire, or develop followers; it is yet another to understand what leaders should do and how they should behave to be transforming, inspiring, or developmental. Thus, we call for construct definitions in sport leadership freed from tautological theorizing, echoing other calls for better definitions in the broader social sciences (Podsakoff, MacKenzie, & Podsakoff, 2016; see also the chapter on mental toughness for useful guidelines on defining constructs). Such definition will be carefully developed (i.e., researchers take a firm stance to indicate what is and what is not part of the construct), be entrenched in the broader leadership literature (because many insights can be gleaned from previous work), and clearly delineate the nomological network of constructs such as antecedent, moderators, and outcomes.

Conclusion

We provided a review of the state of the science with regard to leadership research in sport. There is no doubt that coaches play a pivotal role in developing and preparing athletes and team. The world of the coach is probably as complex as it is important to grasp, and no single model or theory will possibly portray all these complexities. However, the development of sound theory accompanied by a systemic approach and robust testing of carefully developed hypotheses will undoubtedly shed some light on this complex and elusive environment. The field in general has made big steps toward identifying behaviors coaches related to satisfied and high-performing athletes and teams; yet, much remains

to be done, and what lies ahead is exciting. Indeed, we have identified challenges that leadership researchers in sport will face, and our hope is that the whole field seizes the opportunity. As a collective, we have to be at the forefront of applying the latest theories and methodologies or we may lose credit, importance, and funding. The major recommendations for sport leadership arising from our review are:

- Reconnecting with observational methodologies.
- Modifications to the multidimensional model of leadership:
  1) Reposition transformational leadership.
  2) Include a broader range of leader behaviors.
  3) Removal of the required behavior dimension.
  4) Include mediators to explain the process by which congruence affects the model outcomes.
  5) Include moderators to indicate when the relationship between congruence and outcomes will be weaker, stronger, or nullified.
- Re-examine the congruence hypothesis using polynomial regression and response surface methodology.
- Tackle endogeneity issues when designing studies or using an instrumental variable approach.
- Take advantage of experiments and quasi-experimental designs to establish causality.
- Define leadership constructs rigorously so as to avoid tautologies.

References

Amorose, A. J., & Horn, T. S. (2000). Intrinsic motivation: Relationships with collegiate athletes' gender, scholarship status, and perceptions of their coaches' behavior. *Journal of Sport and Exercise Psychology*, *22*(1), 63–84. doi:org/10.1123/jsep.22.1.63

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton, NJ: Princeton University Press.

Antonakis, J. (2017). On doing better science: From thrill of discovery to policy implications. *The Leadership Quarterly*. doi:10.1016/j.leaqua.2017.01.006

Antonakis, J., & Atwater, L. (2002). Leader distance: A review and a proposed theory. *The Leadership Quarterly*, *13*(6): 673–704. doi.org/10.1016/S1048-9843(02)00155-8

Antonakis, J., Bastardoz, N., Jacquart, P., & Shamir, B. (2016). Charisma: An ill-defined and ill-measured gift. *Annual Review of Organizational Psychology and Organizational Behavior*, *3*(1). doi:10.1146/annurev-orgpsych-041015-062305

Antonakis, J., Bastardoz, N., Liu, Y., & Schriesheim, C. A. (2014a). What makes articles highly cited? *The Leadership Quarterly*, *25*(1): 152–179. doi:10.1016/j.leaqua.2013.10.014

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly*, *21*(6): 1086–1120. http://dx.doi.org/10.1016/j.leaqua.2010.10.010

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014b). Causality and endogeneity: Problems and solutions. In D. V. Day (Ed.), *The Oxford handbook of leadership and organizations* (pp. 93–117). New York, NY: Oxford University Press.

Antonakis, J., d'Adda, G., Weber, R., & Zehnder, C. (2015). *Just words? Just speeches? On the economic value of charismatic leadership*. Working paper of the OB department, University of Lausanne.

Antonakis, J., & House, R. J. (2014). Instrumental leadership: Measurement and extension of transformational–transactional leadership theory. *The Leadership Quarterly*, *25*(4), 746–771. http://dx.doi.org/10.1016/j.leaqua.2014.04.005

Aoyagi, M. W., Cox, R. H., & McGuire, R. T. (2008). Organizational citizenship behavior in sport: Relationships with leadership, team cohesion, and athlete satisfaction. *Journal of Applied Sport Psychology*, *20*(1), 25–41. doi:10.1080/10413200701784858

Arthur, C. A., Bastardoz, N., & Eklund, R. (2017). Transformational leadership in sport: Current status and future directions. *Current Opinion in Psychology*, *16*, 78–83. doi:/10.1016/j.copsyc.2017.04.001

Arthur, C. A., & Hardy, L. (2014). Transformational leadership: A quasi-experimental study. *Leadership & Organization Development Journal*, *35*(1), 38–53. doi:10.1108/LODJ-03-2012-0033

Arthur, C. A., Hardy, L., & Woodman, T. (2012). Realisng the olympic dream: Vision, support, and challenge. *Reflective Practice Intrenational and Multidisciplinary Perspectives*, *13*(3), 399–406. doi:10.1080/14623943.2012.670112

Arthur, C. A., & Lynn, A. (2017). Transformational leadership and the role of the coach. In R. Thelwell, C. Harwood, & I. Greenlees (Eds.), *The psychology of sports coaching: Research and practice* (pp. 187–203). Abingdon: Routledge.

Arthur, C. A., & Tomsett, P. (2015). Transformational leaderhsip behaviour in sport. In S. D. Mellalieu, & S. Hanton (Eds.), *Contemporary advances in sport psychology* (pp. 175–201). Abingdon: Routledge.

Arthur, C. A., Woodman, T., Ong, C. W., Hardy, L., & Ntoumanis, N. (2011). The role of athlete narcissism in moderating the relationship between coaches' transformational leader behaviors and athlete motivation. *Journal of Sport & Exercise Psychology*, *33*(1), 3–19. https://doi.org/10.1123/jsep.33.1.3

Banks, G. C., McCauley, K. D., Gardner, W. L., & Guler, C. E. (2016). A meta-analytic review of authentic and transformational leadership: A test for redundancy. *The Leadership Quarterly*, *27*(4), 634–652. https://doi.org/10.1016/j.leaqua.2016.02.006

Baker, J., Côté, J., & Hawes, R. (2000). The relationship between coaching behaviours and sport anxiety in athletes. *Journal of Science and Medicine in Sport/Sports Medicine Australia*, *3*(2), 110–119. doi:10.1016/S1440-2440(00)80073-0

Baker, J., Yardley, J., & Côté, J. (2003). Coach behaviors and athlete satisfaction in team and individual sports. *International Journal of Sport Psychology*, *34*(3), 226–239. doi:10.1177/1747954116676117

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215.

Bandura, A. (1986). *Social foundations of thought and action*. New York, NY: Prentice Hall.

Barling, J., Weber, T., & Kelloway, E. K. (1996). Effects of transformational leadership training on attitudinal and financial outcomes: A field experiment. *Journal of Applied Psychology*, *81*(6), 827–832. doi:10.1037/0021-9010.81.6.827

Barnes, C. M., Dang, C. T., Leavitt, K., Guarana, C. L., & Uhlmann, E. L. (2015). Archival data in micro-organizational research: A toolkit for moving to a broader set of topics. *Journal of Management*, *44(4)*, 1453–1478. doi:10.1177/0149206315604188

Bascle, G. (2008). Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization*, *6*(3), 285–327. doi/10.1177/1476127008094339

Bass, B. M. (1985). *Leadership and performance beyond expectations*. New York, NY: The Free Press.

Bass, B. M., & Avolio, B. J. (1995). *MLQ Multifactor Leadership Questionnaire for Research: Permission Set*. Redwood City, CA: Mindgarden.

Bastardoz, N., Jacquart, P., & Antonakis, J. (2017). *When ANOVA gets it wrong: A re-introduction to the Regression Discontinuity design*. Paper presented at the Academy of Management Proceedings.

Beauchamp, M. R., Barling, J., & Morton, K. L. (2011). Transformational teaching and adolescent self-determined motivation, self-efficacy, and intentions to engage in leisure time physical activity: A randomised controlled pilot trial. *Applied Psychology-Health and Well Being*, *3*(2), 127–150. doi:10.1111/j.1758-0854.2011.01048.x

Becker, A. J., & Wrisberg, C. A. (2008). Effective coaching in action: Observations of legendary collegiate basketball coach Pat Summitt. *Sport Psychologist*, *22*(2), 197–211. https://doi.org/10.1123/tsp.22.2.197

Bettis, R., Gambardella, A., Helfat, C., & Mitchell, W. (2014). Quantitative empirical analysis in strategic management. *Strategic Management Journal*, *35*(7), 949–953. DOI: 10.1002/smj.2278

Bloom, G. A., Crumpton, R., & Anderson, J. E. (1999). A systematic observation study of the teaching behaviors of an expert basketball coach. *The Sport Psychologist*, *13*(2), 157–170.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bono, J. E., & Judge, T. A. (2004). Personality and transformational and transactional leadership: A meta-analysis. *Journal of Applied Psychology*, *89*, 901–910. doi:10.1037/0021-9010.89.5.901

Bormann, K. C., & Rowold, J. (2016). Transformational leadership and followers' objective performance over time: Insights from german basketball. *Journal of Applied Sport Psychology*, *28*(3), 367–373. doi:10.1080/10413200.2015.1133725

Bormann, K. C., Schulte-Coerne, P., Diebig, M., & Rowold, J. (2016). Athlete characteristics and team competitive performance as moderators for the relationship between coach transformational leadership and athlete performance. *Journal of Sport & Exercise Psychology*, *38*(3), 268–281. doi:10.1123/jsep.2015-0182

Caliendo, M., & Kopeinig, S. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. doi:10.1111/j.1467-6419.2007.00527.x

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge: Cambridge University Press.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, *35*(1), 38. http://dx.doi.org/10.1037/0022-3514.35.1.38

Callow, N., Smith, M. J., Hardy, L., Arthur, C. A., & Hardy, J. (2009). Measurement of transformational leadership and its relationship with team cohesion and performance level. *Journal of Applied Sport Psychology*, *21*(4), 395–412. doi:10.1080/10413200903204754

Charbonneau, D., Barling, J., & Kelloway, E. K. (2001). Transformational leadership and sports performance: The mediating role of intrinsic motivation. *Journal of Applied Social Psychology*, *31*(7) doi:10.1111/j.1559-1816.2001.tb02686.x

Chelladurai, P. (1978). *A contingency model of leadership in athletics*. Unpublished doctoral dissertation, University of Waterloo, Canada.

Chelladurai, P. (1984). Discrepancy between preferences and perceptions of leadership-behavior and satisfaction

of athletes in varying sports. *Journal of Sport Psychology*, *6*(1), 27–41. https://doi.org/10.1123/jsp.6.1.27

Chelladurai, P. (2007). Leadership in sports. In G. Tenenbaum and R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 113–135). Hoboken, NJ: Wiley.

Chelladurai, P., Imamura, H., Yamaguchi, Y., Oinuma, Y., & Miyauchi, T. (1988). Sport leadership in a cross-national setting: The case of Japanese and Canadian university athletes. *Journal of Sport and Exercise Psychology*, *10*(4), 374–389. https://doi.org/10.1123/jsep.10.4.374

Chelladurai, P., and Riemer, H. A. (1998). Measurement of leadership in sport. In J. L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 227–253). Morgantown, WV: Fitness Information Technology.

Chelladurai, P., & Saleh, S. D. (1980). Dimensions of leader behavior in sports development of a leadership scale. *Journal of Sport Psychology*, *2*(1), 34–45. https://doi.org/10.1123/jsp.2.1.34

Cook, T. D. (2008). "Waiting for life to arrive": A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, *142*(2), 636–654. doi:10.1016/j.jeconom.2007.05.002

Cooper, W. H., & Richardson, A. J. (1986). Unfair comparisons. *Journal of Applied Psychology*, *71*(2), 179. http://dx.doi.org/10.1037/0021-9010.71.2.179

Cope, E., Partington, M., & Harvey, S. (2017). A review of the use of systematic observation in coaching research between 1997 and 2016. *Journal of Sport Sciences*, *35*(20), 2042–2050. doi:10.1080/02640414.2016.1252463

Côté, J., Yardley, J. K., Sedwick, W., & Baker, J. (1999). An exploratory examination of the coaching behaviour scale for sport. *Avante*, *5*, 82–92.

Cronin, L. D., Arthur, C. A., Hardy, J., & Callow, N. (2015). Transformational leadership and task cohesion in sport: The mediating role of inside sacrifice. *Journal of Sport & Exercise Psychology*, *37*(1), 23–36. doi:10.1123/jsep.2014-0116

Cruz, A. B., & Kim, H. D. (2017). Leadership preferences of adolescent players in sport: Influence of coach gender. *Journal of Sports Science & Medicine*, *16*(2), 172–179.

Curtis, B., Smith R., & Smoll, F. (1979). Scrutinizing the skipper— study of leadership behaviors in the dugout. *Journal of Applied Psychology*, *64*(4), 391–400. doi:10.1037//0021-9010.64.4.391

Cushion, C., Harvey, S., Muir, B., & Nelson, L. (2012). Developing the coach analysis and intervention system (CAIS): Establishing validity and reliability of a computerised systematic observation instrument. *Journal of Sports Sciences*, *30*(2), 203–218. doi:10.1080/02640414.2011.635310

Duda, J. L., & Balaguer, I. (1999). Toward an integration of models of leadership with a contemporary theory of motivation. In R. Lidor & M. Bar-Eli (Eds.), *Sport psychology: Linking theory practice* (pp. 213–230). Morgantown, WV: Fitness Information Technology.

Dvir, T., Eden, D., Avolio, B. J., & Shamir, B. (2002). Impact of transformational leadership on follower development and performance: A field experiment. *Academy of Management Journal*, *45*(4), 735–744. doi:10.2307/3069307

Eagly, A. H., Johannesen-Schmidt, M. C., & Van Engen, M. L. (2003). Transformational, transactional, and laissez-faire leadership styles: A meta-analysis comparing women and men. *Psychological Bulletin*, *129*, 569–591. doi:10.1037/0033-2909.129.4.569

Eden, D. (2017). Field experiments in organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, *4*, 91–122. https://doi.org/10.1146/annurev-orgpsych-041015-062400

Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. *Journal of Management*, *20*(3), 683–689.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods*, *4*(3), 265–287. http://dx.doi.org/10.1177/109442810143005

Edwards, J. R. (2002). Alternatives to difference scores: Polynomial regression analysis and response surface methodology. In F. Drasgow & N. Schmitt (Eds.), *The Jossey-Bass business & management series. Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 350–400). San Francisco, CA: Jossey-Bass

Edwards, J., & Parry, M. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal*, *36*(6), 1577–1613. doi:10.2307/256822

Edwards, J. R., & Parry, M. E. (2018). On the use of spline regression in the study of congruence in organizational research. *Organizational Research Methods*, *21*(1), 68–110. doi:10.1177/1094428117715067

Feltz, D., Chase, M., Moritz, S., & Sullivan, P. (1999). A conceptual model of coaching efficacy: Preliminary investigation and instrument development. *Journal of Educational Psychology*, *91*(4), 765–776. doi:10.1037//0022-0663.91.4.765

Fiedler, F. E. (1964). A contingency model of leadership effectiveness. *Advances in Experimental Social Psychology*, *1*, 149–190.

Fischer, T., Dietz, J., & Antonakis, J. (2017). Leadership process models: A review and synthesis. *Journal of Management*, *43*(6), 1726–1753. DOI: 10.1177/0149206316682830

Gallimore, R., & Tharp, R. (2004). What a coach can teach a teacher, 1975–2004: Reflections and reanalysis of John Wooden's teaching practices. *Sport Psychologist*, *18*(2), 119–137. 10.1123/tsp.18.2.119

Gardner, D., Shields, D., Bredemeier, B., & Bostrom, A. (1996). The relationship between perceived coaching behaviors and team cohesion among baseball and softball players. *Sport Psychologist*, *10*(4), 367–381. https://doi.org/10.1123/tsp.10.4.367

Gioia, D. A., & Sims Jr., H. P. (1985). On avoiding the influence of implicit leadership theories in leader behavior descriptions. *Educational and Psychological Measurement*, *45*(2), 217–232. doi/abs/10.1177/001316448504500204

Gould, D. R., & Wright, E. M. (2012). The psychology of coaching. In S. Murphey, S. (Ed.), *The Oxford handbook of sport and performance psyhcology* (pp. 343–363). New York, NY: Oxford University Press.

Graham, J., Haidt, J., & Nosek, B. A.(2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029. doi:10.1037/a0015141

Greenberg, J. (1990). Employee theft as a reaction to underpayment inequity: The hidden cost of pay cuts. *Journal of Applied Psychology*, *75*(5), 561. doi:10.1037/0021-9010.75.6.667

Griffith, C. R. (1925). Psychology and its realtion to athletic competition. *Physical Education Review*, *30*, 193–199.

Guide, V. D. R., & Ketokivi, M. (2015). Notes from the editors: Redefining some methodological criteria for the journal. *Journal of Operations Management*, *37*, v–viii.

Hamilton, B. H., & Nickerson, J. A.(2003). Correcting for endogeneity in strategic management research. *Strategic Organization*, *1*(1), 51–78. doi:10.1177/1476127003001001218

Hardy, L., Arthur, C. A., Jones, G., Shariff, A., Munnoch, K., Isaacs, I., & Allsopp, A. J. (2010). The relationship between transformational leadership behaviors, psychological, and training outcomes in elite military recruits. *Leadership Quarterly*, *21*(1), 20–32. doi:10.1016/j.leaqua.2009.10.002

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, *46*(6), 1251–1271.

Heckman, J. J. (1977). *Sample selection bias as a specification error (with an application to the estimation of labor supply functions)*. National Bureau of Economic Research Cambridge, MA.

Hemphill, J. K. (1950). Relations between the size of the group and the behavior of "superior" leaders. *The Journal of Social Psychology*, *32*(1), 11–22.

Highhouse, S. (2009). Designing experiments that generalize. *Organizational Research Methods*, *12*(3), 554–566. doi:/abs/10.1177/1094428107300396

Hodge, K., Henry, G., & Smith, W. (2014). A case study of excellence in elite sport: Motivational climate in a world champion team. *Sport Psychologist*, *28*(1), 60–74. doi:10.1123/tsp.2013-0037

Hollembeak, J., & Amorose, A. J. (2005). Perceived coaching behaviors and college athletes' intrinsic motivation: A test of self-determination theory. *Journal of Applied Sport Psychology*, *17*(1), 20–36. doi:10.1080/10413200590907540

Horn, T. S. (2002). Coaching effectiveness in the sport domain. In T. S. Horn (Ed.), *Advances in sport psychology* (pp. 309–354). Champaign, IL: Human Kinetics.

Horn, T. S. (2008). Coaching effeciveness in the sport domain. In T. S. Horn (Ed.), *Advances in sport psychology* (3rd ed., pp. 239–267). Champaign, IL: Human Kinetics.

Horn, T. S., Bloom, P., Berglund, K. M., & Packard, S. (2011). Relationship between collegiate athletes' psychological characteristics and their preferences for different types of coaching behavior. *The Sport Psychologist*, *25*(2), 190–211.

Horn, T., & Carron, A. v. (1985). Compatability in coach-athlete relationships. *The Journal of Sports Psychology, 7(2), 137–149.*

Jacquart, P., & Antonakis, J. (2015). When does charisma matter for top-level leaders? Effect of attributional ambiguity. *Academy of Management Journal*, *58*(4), 1051–1074.

Jacquart, P., Cole, M. S., Gabriel, A. S., Koopman, J., & Rosen, C. (2017). Studying leadership: Introduction to research design and methods. In D. V. Day, & J. Antonakis (Eds.), *The nature of leadership* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*(4), 443–482.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: A qualitative and quantitative review. *Journal of Applied Psychology*, *87*(4), 765.

Judge, T. A., Colbert, A. E., & Ilies, R. (2004). Intelligence and leadership: A quantitative review and test of theoretical propositions. *Journal of Applied Psychology*, *89*(3), 542. doi:10.1037/0021-9010.89.3.542

Judge, T. A., & Piccolo, R. F. (2004). Transformational and transactional leadership: A meta-analytic test of their relative validity. *Journal of Applied Psychology*, *89*(5), 755–768.

Kahn, R. L., & Katz, D. (1952). *Leadership practices in relation to productivity and morale*. Ann Arbor, MI: Institute for Social Research, University of Michigan.

Kao, S., Chen, Y., Watson, J. C., & Halbrook, M. (2015). Relationships between the congruence of required and perceived leadership behavior and satisfaction in athletes. *Psychological Reports*, *117*(2), 391–405.

Kao, S., & Tsai, C. (2016). Transformational leadership and athlete satisfaction: The mediating role of coaching competency. *Journal of Applied Sport Psychology*, *28*(4), 469–482. doi:https://doi.org/10.1080/10413200.2016.1187685

Kennedy, P. (2003). *A guide to econometrics*. Cambridge, MA: MIT Press.

Ketokivi, M., & McIntosh, C. N.(2017). Addressing the endogeneity dilemma in operations management research: Theoretical, empirical, and pragmatic considerations. *Journal of Operations Management*, *52*, 1–14.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford Press.

Kristof, A.L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology*, *49*, 1–15.

Lance, C. E., Butts, M. M., & Michels, L. C.(2006). The sources of four commonly reported cutoff criteria what did they really say? *Organizational Research Methods*, *9*(2): 202–220.

Lacy, A. C., & Darst, P. W. (1984). Evolution of a systematic observation system: The ASU coaching observation instrument. *Journal of Teaching in Physical Education*, *3*(3), 59–66.

Lewin, K., Lippitt, R., & White, R. K. (1939). Patterns of aggressive behavior in experimentally created "social climates." *The Journal of Social Psychology*, *10*(2), 269–299.

Li, M. (2013). Using the propensity score method to estimate causal effects: A review and practical guide. *Organizational Research Methods*, *16*(2), 188–226. doi/abs/10.1177/1094428112447816.

Lord, R. G., Binning, J. F., Rush, M. C., & Thomas, J. C. (1978). The effect of performance cues and leader behavior on questionnaire ratings of leadership behavior. *Organizational Behavior and Human Performance*, *21*(1), 27–39.

Lord, R. G., Foti, R. J., & De Vader, C. L. (1984). A test of leadership categorization theory: Internal structure, information processing, and leadership perceptions. *Organizational Behavior and Human Performance*, *34*(3), 343–378.

Loughead, T., & Hardy, J. (2005). An examination of coach and peer leader behaviors in sport. *Psychology of Sport and Exercise*, *6*(3), 303–312. doi:10.1016/j.psychsport.2004.02.001

MacKenzie, S. B. (2003). The dangers of poor construct conceptualization. *Journal of the Academy of Marketing Science*, *31*(3), 323–326. doi/abs/10.1177/0092070303031003011

Mageau, G., & Vallerand, R. (2003). The coach-athlete relationship: A motivational model. *Journal of Sports Sciences*, *21*(11), 883–904. doi:10.1080/02640410310001401374

Malete, L., Chow, G. M., & Feltz, D. L. (2013). Influence of coaching efficacy and coaching competency on athlete-level moral variables in Botswana youth soccer. *Journal of Applied Social Psychology*, *43*(10), 2107–2119. doi:10.1111/jasp.12164

Marcus, B., Weigelt, O., Hergert, J., Gurt, J., & Gelléri, P. (2016). The use of snowball sampling for multi source organizational research: Some cause for concern. *Personnel Psychology*, *70*(3), 635–673.

Martin, S., Jackson, A., Richardson, P., & Weiller, K. (1999). Coaching preferences of adolescent youths and their parents. *Journal of Applied Sport Psychology*, *11*(2), 247–262. doi:10.1080/10413209908404203

McArdle, S., & Duda, J. K. (2002). Implications of the motivational climate in youth sports. *Children and Youth in Sport: A Biopsychosocial Perspective*, *2*, 409–434.

Meindl, J. R., & Ehrlich, S. B. (1987). The romance of leadership and the evaluation of organizational performance. *Academy of Management Journal*, *30*(1), 91–109.

Mellor, S., & Mark, M. M. (1998). A quasi-experimental design for studies on the impact of administrative decisions: Applications and extensions of the regression-discontinuity design. *Organizational Research Methods*, *1*(3), 315–333.

Mills, J. P., & Boardley, I. D. (2016). Expert premier league soccer managers' use of transformational leadership behaviours and attitude towards sport integrity: An intrinsic case study. *International Journal of Sports Science & Coaching*, *11*(3), 382–394. doi:10.1177/1747954116645011

Mills, J. P., & Boardley, I. D. (2017). Development and initial validation of an indirect measure of transformational leadership integrity. *Psychology of Sport and Exercise*, *32*, 34–46. doi.org/10.1016/j.psychsport.2017.05.005

Mullen, J. E., & Kelloway, E. K. (2009). Safety leadership: A longitudinal study of the effects of transformational leadership on safety outcomes. *Journal of Occupational and Organizational Psychology*, *82*(2), 253–272. doi:10.1348/096317908X325313

Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. *The Journal of Economic Perspectives*, *20*(4), 111–132. doi:10.1257/jep.20.4.111

Myers, N. D., Beauchamp, M. R., & Chase, M. A. (2011). Coaching competency and satisfaction with the coach: A multi-level structural equation model. *Journal of Sports Sciences*, *29*(4), 411–422. doi:10.1080/02640414.2010.538710

Myers, N. D., Chase, M. A., Beauchamp, M. R., & Jackson, B. (2010). Athletes' perceptions of coaching competency scale II-high school teams. *Educational and Psychological Measurement*, *70*(3), 477–494. doi:10.1177/0013164409344520

Myers, N. D., Feltz, D. L., Chase, M. A., Reckase, M. D., & Hancock, G. R. (2008). The coaching efficacy scale II-high school teams. *Educational and Psychological*

Measurement, *68*(6), 1059–1076. doi:10.1177/0013164408318773

Myers, N. D., Feltz, D. L., & Wolfe, E. W. (2008). A confirmatory study of rating scale category effectiveness for the coaching efficacy scale. *Research Quarterly for Exercise and Sport*, *79*(3), 300–311. doi.org/10.1080/02701367.2008.10599493

Myers, N. D., Feltz, D. L., Maier, K. S., Wolfe, E. W., & Reckase, M. D. (2006). Athletes' evaluations of their head coach's coaching competency. *Research Quarterly for Exercise and Sport*, *77*(1), 111–121.

Myers, N. D., Vargas-Tonsing, T. M., & Feltz, D. L. (2005). Coaching efficacy in intercollegiate coaches: Sources, coaching behavior, and team variables. *Psychology of Sport and Exercise*, *6*(1), 129–143. doi:10.1016/j.psychsport.2003.10.007

Myers, N.D., Wolfe, E., Maier, K., Feltz, D., & Reckase, M. (2006). Extending validity evidence for multidimensional measures of coaching competency. *Journal of Sport & Exercise Psychology*, *28*, S139–S139.

Newland, A., Newton, M., Podlog, L., Legg, W. E., & Tanner, P. (2015). Exploring the nature of transformational leadership in sports: A phenomenological examination with female athletes. *Qualitative Research in Sport, Exercise and Health*, *7*(5), 663–687. doi:10.1080/2159676X.2015.1007889

Oxford English Dictionary (2007). *Oxford English Dictionary Online*. Oxford: Oxford University Press.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of applied psychology*, *88*(5), 879. doi:10.1037/0021-9010.88.5.879

Podsakoff, P. M., MacKenzie, S. B., Moorman, R. H., & Fetter, R. (1990). Transformational leader behaviors and their effects on followers' trust in leader, satisfaction, and organizational citizenship behaviors. *The Leadership Quarterly*, *1*(2), 107–142. doi:10.1016/1048-9843(90)90009-7

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2016). Recommendations for creating better concept definitions in the organisational, behavioural, social sciences. *Organisational Research Methods*, *19*, 159–203. doi:10.1177/1094428115624965

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*(4), 531–544. doi.org/10.1177/014920638601200408

Reeb, D., Sakakibara, M., & Mahmood, I. P. (2012). From the editors: Endogeneity in international business research. *Journal of International Business Studies*, *43*(3): 211–218. doi.org/10.1057/jibs.2011.60. doi:10.1007/s40279-015-0345-4

Richardson, H. A., Simmering, M. J., & Sturman, M. C.(2009). A tale of three perspectives: Examining post hoc statistical techniques for detection and correction of common method variance. *Organizational Research Methods*, *12*(4), 762–800. doi:10.1177/1094428109332834.

Riemer, H. A. (2007). Multidimensional model of coach leadership. In S. Jowette & D. Lavallee (Eds.), *Social psychology in sport* (pp. 57–73). Champaign, IL: Human Kinetics.

Riemer, H. A., & Chelladurai, P. (1995). Leadership and satisfaction in athletics. *Journal of Sport & Exercise Psychology*, *17*(3), 276–293. doi.org/10.1123/jsep.17.3.276

Riemer, H., & Toon, K. (2001). Leadership and satisfaction in tennis: Examination of congruence, gender, and ability. *Research Quarterly for Exercise and Sport*, *72*(3), 243–256. doi:10.1080/02701367.2001.10608957

Roberts, M. R., & Whited, T. M. (2013). Endogeneity in empirical corporate finance. *Handbook of the Economics of Finance* (Vol. *2*, pp. 493–572). Elsevier.

Salancik, G. R., & Pfeffer, J. (1977). An examination of need-satisfaction models of job attitudes. *Administrative Science Quarterly*, *22(3)*, 427–456.

Schliesman, E. S. (1987). Relationship between the congruence of preferred and actual leader behavior and subordinate satisfaction with leadership. *Journal of Sport Behavior*, *10*(3), 157.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton, Mifflin and Company.

Shamir, B. (1995). Social distance and charisma: Theoretical notes and an exploratory study. *The Leadership Quarterly*, *6*(1), 19–47. doi.org/10.1016/1048-9843(95)90003-9

Shamir, B., & Howell, J. M. (1999). Organizational and contextual influences on the emergence and effectiveness of charismatic leadership. *The Leadership Quarterly*, *10*(2), 257–283. doi.org/10.1016/S1048-9843(99)00014-4

Shaver, J. M. (2005). Testing for mediating variables in management research: Concerns, implications, and alternative strategies. *Journal of Management*, *31*(3), 330–353. oi/abs/10.1177/0149206304272149

Shields, D. L. L., Gardner, D. E., Bredemeier, B. J. L., Bostro, A. (1997). The relationship between leadership behaviors and group cohesion in team sports. *Journal of Psychology* , *13*(2), 196-210. doi: 10.1080/00223989709601964.

Short, J. C., Broberg, J. C., Cogliser, C. C., & Brigham, K. H. (2010). Construct validation using computer-aided text analysis (CATA): An illustration using entrepreneurial orientation. *Organizational Research Methods*, *13*(2), 320–347. doi/abs/10.1177/1094428109335949

Smith, M. J., Arthur, C. A., Hardy, J., Callow, N., & Williams, D. (2013). Transformational leadership and

task cohesion in sport: The mediating role of intrateam communication. *Psychology of Sport and Exercise*, *14*(2), 249–257.

Smith, M. J., Young, D. J., Figgins, S. G., & Arthur, C. A. (2017). Transformational Leadership in Elite Sport: A Qualitative Analysis of Effective Leadership Behaviors in Cricket. *The Sport Psychologist*, *31*(1), 1-15. doi:10.1123/tsp.2015-0077

Smith, R., Zane, N., Smoll, F., & Coppel, D. (1983). Behavioral-assessment in youth sports – coaching behaviors and childrens attitudes. *Medicine and Science in Sports and Exercise*, *15*(3), 208–214.

Smith, R. E., & Smoll, F. L. (2007). Social-cognitive approach to coaching behaviors. In S. Jowette & D. Lavallee (Eds.), *Social psychology in sport* (pp. 75–90). Champaign, IL: HUman KInetics.

Smith, R., Smoll, F., & Barnett, N. (1995). Reduction of childrens sport performance anxiety through social support and stress-reduction training for coaches. *Journal of Applied Developmental Psychology*, *16*(1), 125–142. doi:10.1016/0193-3973(95)90020-9

Smith, R. E., Smoll, F. L., & Cumming, S. P. (2007). Effects of a Motivational Climate Intervention for Coaches on Young Athletes' Sport Performance Anxiety. *Journal of Sport and Exercise Psychology*, *29*(1), 39-59. doi:10.1123/jsep.29.1.39

Smith, R. E., Smoll, F. L., & Curtis, B. (1979). Coach effectiveness training: A cognitive-behavioral approach to enhancing relationship skills in youth sport coaches. *Journal of Sport Psychology*, *1*(1), 59–75.

Smith, R. E., Smoll, F. L., & Hunt, E. (1977). System for behavioral-assessment of athletic coaches. *Research Quarterly*, *48*(2), 401–407.

Smoll, F. L., & Smith, R. E. (1989). Leadership behaviors in sport: A theoretical model and research paradigm. *Journal of Applied Social Psychology*, *19*(18), 1522–1551. doi:10.1111/j.1559-1816.1989.tb01462.x

Smoll, F. L., Smith, R. E., Barnett, N. P., & Everett, J. J. (1993). Enhancement of children's self-esteem through social support training for youth sport coaches. *Journal of Applied Psychology*, *78*(4), 602–610. doi.org/10.1037/0021-9010.78.4.602

Smoll, F. L., Smith, R. E., Curtis, B., & Hunt, E. (1978). Toward a mediational model of coach-player relationships. *Research Quarterly*, *49*(4), 528–541. doi/abs/10.1080/10671315.1978.10615567

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods*, *9*(2), 221–232. doi:10.1177/1094428105284955

Spielberger, C. D. (1966). Theory and research on anxiety. *Anxiety and Behavior*, *1*, 3–20.

Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586. doi:10.2307/2171753

Stenling, A., & Tafvelin, S. (2014). Transformational leadership and well-being in sports: The mediating role of need satisfaction. *Journal of Applied Sport Psychology*, *26*(2), 182–196. doi:10.1080/10413200.2013.819392

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, *88*(3), 500–517. doi.org/10.1037/0021-9010.88.3.500

Tharp, R. G., & Gallimore, R. (1976). What a coach can teach a teacher. *Psychology Today*, *9*(8), 75–78.

Tonidandel, S., King, E. B., & Cortina, J. M. (2018). Big data methods: Leveraging modern data analytic techniques to build organizational science. *Organizational Research Methods*, *21*(3), 525–547.

Van Knippenberg, D., & Sitkin, S. B. (2013). A critical assessment of charismatic— transformational leadership research: Back to the drawing board? *The Academy of Management Annals*, *7*(1), 1–60. doi.org/10.1080/19416520.2013.759433

Vella, S. A., Oades, L. G., & Crowe, T. P. (2013). A pilot test of transformational leadership training for sports coaches: Impact on the developmental experiences of adolescent athletes. *International Journal of Sports Science and Coaching*, *8*(3), 513–530. doi/abs/10.1260/1747-9541.8.3.513

Vincer, D. J. E., & Loughead, T. M. (2010). The relationship among athlete leadership behaviors and cohesion in team sports. *Sport Psychologist*, *24*(4), 448–467. doi.org/10.1123/tsp.24.4.448

Weiss, M. R., & Friedrichs, W. D. (1986). The influence of leader behaviors, coach attributes, and institutional variables on performance and satisfaction of collegiate basketball teams. *Journal of Sport Psychology*, *8*(4), 332–346. doi.org/10.1123/jsp.8.4.332

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, MA: MIT Press.

Yammarino, F. J., Dubinsky, A. J., Comer, L. B., & Jolson, M. A. (1997). Women and transformational and contingent reward leadership: A multiple-levels-of-analysis perspective. *Academy of Management Journal*, *40*(1), 205–222.

Yukl, G. (1971). Toward a behavioral theory of leadership. *Organizational Behavior and Human Performance*, *6*(4), 414–440.

Zacharatos, A., Barling, J., & Kelloway, E. K. (2000). Development and effects of transformational leadership in adolescents. *Leadership Quarterly*, *11*(2) doi:10.1016/S1048-9843(00)00041-2

Zehnder, C., Herz, H., & Bonardi, J.-P. (2017). A productive clash of cultures: Injecting economics into leadership research. *The Leadership Quarterly*, *28*(1), 65–85. doi.org/10.1016/j.leaqua.2016.10.004