Longitudinal Stability of the Leadership Scale for Sports

Richard B. Fletcher and Margaret H. Roberts

School of Psychology
Massey University, Albany, New Zealand

This study evaluated the invariance properties of the Leadership Scale for Sport in a sample of 219 female netball players over four time points within a 10-week playing season. Support was found for Chelladurai and Saleh’s (1980) hypothesized 5-factor structure of the Leadership Scale for Sport. Furthermore, differential stability and partial invariance was found for the Leadership Scale for Sport when all four time periods were included. Players perceived slight changes in their coach’s autocratic behavior and social support over the season; however, the three other leadership dimensions showed larger changes. The motivational aspects of training and instruction and positive feedback behavior were perceived to increase, while democratic behavior simultaneously decreased in the second half of the season. Furthermore, perceptions of leadership within teams showed a high level of homogeneity with the exception of positive feedback behavior.

Key words: leadership, longitudinal stability, sport

INTRODUCTION

The importance of business leaders in determining organizational success or failure has long been emphasized, leading to the emergence of a large volume of leadership theories from organizational psychology. Similarly, the leadership role of the coach in sporting contexts has also been emphasized as making a significant contribution to athletes’ sporting success or failure. The role of the coach has both technical and interpersonal components which are aimed towards maximizing athletes’ performance. The technical aspects of sport include training athletes and developing game strategies, whereas the interpersonal role of the coach includes supporting and motivating athletes and having an awareness of their personal strengths and limitations to produce the best possible performance. Despite the importance placed on the leadership role of the coach, only a handful of sporting leadership theories have been given empirical attention.

Three models of sporting leadership dominate this literature: (a) the Normative Model of Decision Styles in Coaching (Chelladurai & Haggerty, 1978); (b) the Mediational Model of Leadership in Sport (Smoll & Smith, 1989); and (c) the Multidimensional Model of Leadership in Sport (MML; Chelladurai & Saleh, 1980). One of the most utilized measures within the MML literature is the Leadership Scale for Sport (LSS; Chelladurai & Saleh, 1980). While research...
has investigated the stability of the measure over two time periods (Chelladurai & Saleh, 1980),
evidence of its longitudinal factorial stability is yet to be determined. With this in mind, the
primary intention of this article is to evaluate the longitudinal stability of MML’s accompanying
psychometric measure, the LSS. The longitudinal nature of this research provides an opportunity
to undertake a secondary evaluation of leadership behavior, enabling us to gain an indication of
whether athletes’ perceptions of leadership behavior change over time.

The Multidimensional Model of Leadership

Chelladurai’s (1990) MML is one of the most significant models of sporting leadership, and has
generated extensive empirical attention. The MML proposes that group performance and member satisfaction are a function of the congruence of required, actual, and preferred leadership behavior. These leadership behaviors are considered to vary in response to the antecedent variables of situational characteristics, which includes contextual factors such as the nature of the sport and associated social norms; leader characteristics, which include the individual differences of the coach; and member characteristics, which encompass the individual differences of the players. The antecedent variables of member characteristics and situational characteristics determine the required leader behavior and preferred leader behavior. Actual leader behavior is considered to be responsive to the member and situational characteristics, as well as the required and preferred leadership behavior. Chelladurai (1980) asserts that the congruence between preferred, required, and perceived leadership behavior determine the level of the outcome variable of member satisfaction and group performance.

Chelladurai and Saleh (1980) developed the 40-item LSS to assess leadership behavior and evaluate the hypothesized relationships within the MML. Each of the five subscales of the LSS is considered to represent a dimension of leadership behavior: democratic, autocratic, training and instruction, social support, and positive feedback. Chelladurai and Saleh (1980) suggest that the democratic behavior subscale (nine items) and autocratic behavior subscale (five items) are factors that reflect the decision style of the coach. Democratic behavior assesses the extent to which the coach allows the athletes to participate in decision making and goal setting. Autocratic behavior reflects an authoritarian decision-making style and also includes the closeness between coach and player in terms of their relationship. Training and instruction (13 items) represents direct tasks of the coach, such as assisting athletes to develop skills and learn the tactics of the sport. Social support behavior (eight items) represents the coach’s ability to satisfy the interpersonal needs of the athletes, either directly or indirectly through creating a supportive atmosphere amongst members. Positive feedback behavior (five items) represents the coach’s ability to recognize and express appreciation of members’ efforts and complement performance.

The psychometric qualities of the LSS suggest moderate to high internal consistency reliability estimates for the training and instruction, democratic behavior, social support, and positive feedback subscales of the LSS. Lower reliability estimates have generally been obtained using the autocratic behavior subscales, although this finding has been inconsistent across studies. Higher internal consistency reliability has been obtained for the perception version when compared to the preference version across multiple studies, although acceptable reliability has been found for both versions (Chelladurai & Riemer, 1998).

Chelladurai and Saleh (1980) assessed the stability of the LSS with 53 physical education students with a four-week interval between the two administrations. Adequate test–retest reliability
was found for training and instruction ($\alpha = .72$), democratic behavior ($\alpha = .82$), autocratic behavior ($\alpha = .76$), social support ($\alpha = .71$), and positive feedback ($\alpha = .79$).

Chelladurai and Reimer (1998) evaluated the construct validity of the perception and preference versions of the LSS using confirmatory factor analysis (CFA) on data collected from collegiate football players ($n = 217$ for the perception version, and $n = 317$ for preference version). Both models offered support for the hypothesized 5-factor structure of the LSS and suggested a good fit to the population (Root Mean Square Error of Approximation [RMSEA] .062 and .060 respectively). However, the inclusion of the Tucker-Lewis Index (TLI) and Bollen’s fit index (D2) suggested a stronger fit for the perception model when compared to the preference model (TLI .83 and .77, D2 .85 and .78, respectively). Although these fit indices are lower, these results provide basic evidence for the validity of the 5-factor structure of the LSS. However, as some of the fit indices were unsatisfactory, the authors recommended that the subscale structure should be evaluated using other sets of data.

An important methodological issue is that researchers have used techniques which have the inherent assumption of the homogeneity of leadership behavior, such as averaging the perceptions of players to determine actual leadership behavior (Chelladurai & Riemer, 1998). Chelladurai and Riemer suggest that the use of average player perceptions as a measure of actual leader behavior can be supported if the level of variance is low. However, the use of average perceptions is dubious if variance in player perceptions is substantial, as this would suggest heterogeneity in athletes’ perceptions of leader behavior.

Finally, the cross-sectional nature of previous research is a significant limitation, as it holds the inherent assumption that leadership behavior is static over the course of a season. However, Weinberg and McDermott (2002) found that coaches placed high importance on their ability to adapt their behavior to changes in the situational context. Therefore, the timing of data collection could significantly influence the observed results.

In addition to the methodological issues, the psychometric properties of the LSS have not been fully explored. This is crucial in determining whether the observed relationships between constructs are real or merely an artifact of measurement error (Kline, 1993). One method is the use of longitudinal structural equation modeling techniques to enable the researcher to determine whether observed changes over time are due to changes within the subjects, changes within the construct, or unstable measurement properties of the tool (Vandenberg & Lance, 2000). Thus, the aims of this research were to examine the psychometric properties of the LSS by: (a) testing the factor structure of the LSS using confirmatory factor analysis; and (b) testing the longitudinal psychometric properties by analyzing the temporal stability of the LSS in terms of structural and differential stability. Of secondary interest, was to (c) evaluate whether perceptions of leadership behavior are stable over a season, and if these perceptions are homogenous within teams.

METHOD

Participants

Participants consisted of 219 senior level female netball players from 25 teams in Auckland, New Zealand, competing over a 10-week playing season. Players were of regional and national standard and ranged between 6 to 12 players per team. It is not uncommon in New Zealand to have older players participating at this level, and thus, the structure of the teams in terms of age
ranged from 14 to 47 years, with an average age of 23 years ($SD = 6.36$): only 27 players were older than 30 years of age. Although data on the gender of the coach was not taken, most netball coaches were female. The teams practiced once a week, and generally, the coaches were active players. Netball involves two teams attempting to pass a ball through a hoop, however, unlike basketball, players are not allowed to step while holding the ball and, therefore, must pass to another player to move the ball to the goal. Netball in New Zealand is the most-played female sport in New Zealand, and the national team is ranked within the top three teams in the world.

**Measure**

The perception version of the LSS (Chelladurai & Saleh, 1980) was used in this study, as we were interested in how players perceive their coaches’ leadership style over a playing season. The perception version of the LSS is a 40-item measure that includes the following five subscales: training and instruction (13 items), democratic behavior (nine items), autocratic behavior (five items), social support (eight items), and positive feedback behavior (five items). Items are scored on a 5-point Likert scale anchored from 1 (**always**) to 5 (**never**). Given that the nature of the relationship between players and coach is less formal than the North American environment, we opted for the Chelladurai & Saleh’s original version, as it incorporated questions that were relevant to the New Zealand sample. For example “My coach invites athletes home” is pertinent to use with this sample, as the coaches were more often than not player-coaches, and thus, their interactions with players is different from the North American context and may lead to such social interactions.

Informed consent was obtained from all participants prior to data collection. Members of each team completed the LSS prior to practice and within five days of the next competitive match. Data were collected at four intervals during a 10-week period: initial baseline, one, six, and 10 weeks.

**Evaluation of Missing Data**

As data were collected at the practice session preceding the next game, there were players that did not always attend the session. The number of participants at each practice session varied over the course of the season, and thus, towards the end of the season, the number of responses did decrease, although not to the point where data analyses were compromised. Over the four data collection points, a total of 8%, 22%, 27%, and 36% (respectively) of the data were missing. The inappropriate treatment of missing data in longitudinal research can potentially bias the estimates yielded during analysis. To determine the most appropriate way of proceeding with missing data, the researcher must first assess whether their data is missing at random (MAR) or missing completely at random (MCAR). Data that is MAR implies that the cause of missing data is systematic and correlated with the variable containing missing data. On the other hand, data that is MCAR implies that the cause of missing data is a random process and is uncorrelated with the variable containing missing data (Little & Ruben, 1987).

Following the methods outlined by Conroy, Metzler, and Hofer (2003), data at each time point was found not to significantly correlate with total scores on the LSS. Data was therefore considered to be MCAR, and as such, participant attrition was not considered to be a function of prior scores on the LSS.
Confirmatory Factor Analysis

This study first evaluated a CFA model using AMOS 5.0 (Arbuckle, 2003) to determine the most appropriate relationships between the five factors of leadership in sport identified by Chelladurai and Saleh (1980). A first-order CFA model hypothesized five correlated first-order factors (training and instruction, democratic behavior, autocratic behavior, social support, and positive feedback behavior) over the four time points, to assess the degree to which this explained the observed scores on the LSS.

Evaluation of Structural Stability

All analyses were conducted at the subscale level, as the sample size led to convergence issues, which is not uncommon when attempting to test multi-subscale measures across waves of data (see Conroy et al., 2003). Structural stability refers to the longitudinal consistency of measurement properties of a measure. Establishing structural stability for the LSS would suggest that the five dimensions of leadership behavior hypothesized by Chelladurai and Saleh (1980) were being consistently measured over time.

Evidence of structural stability was obtained using Meredith’s (1993) hierarchy of increasingly stringent tests of invariance: (a) configural invariance, (b) weak factorial invariance, and (c) strong factorial invariance. Structural stability of the LSS can be concluded when the progression through these nested models does not cause significant changes in model fit (Cheung & Rensvold, 2002; Schutz, 1998).

The least stringent test of configural invariance requires equivalent numbers and patterns of factor loadings across time points. Weak factorial invariance requires the factor loadings to be equivalent across time points and evaluates the degree to which factor–variable regressions are identical. Strong factorial invariance includes the addition of equality constraints to the intercepts of each manifest variable to the latent constructs across time points. The progression through each stage of invariance testing is contingent on the establishment of invariance using the previous less-stringent constraints. Evidence of structural stability is established when progression through each stage of testing does not result in significant changes in model fit (Meredith, 1993).

Marsh (1993) noted that if the same measures are used at each wave of data collection, then the corresponding errors tend to correlate. Failure to account for these correlated errors can lead to a positive inflation of the beta path coefficients. In light of this, errors associated with the observed variable were correlated with the corresponding error across each time point throughout invariance testing.

Evaluation of Differential Stability

Differential stability is similar to test–retest reliability, as it is a correlation of the scores obtained over multiple testings. However, differential stability provides an indication of whether an individual retains their relative position in relation to other individuals within a team over time. Evidence of differential stability can be established through assessing the standardized covariance coefficients between each team (Conroy et al., 2003). The model which achieves the highest
level of factorial invariance is utilized to estimate the differential stability coefficients between the latent factors across each time period.

**Evaluation of Model Fit**

The chi-square statistic is a test of the null hypothesis that the specification of factor loadings, factor variances, co-variances, and error variances for the model are valid. In light of this, a non-significant result is desirable, as it suggests a well-fitting model (Byrne, 2001). As the chi-square statistic is sensitive to sample size, Vanderberg and Lance (2000) recommend that the chi-square statistic should be interpreted only in conjunction with several other practical fit indexes.

The comparative-fit index (CFI) ranges from 0 to 1.00 and indicates how much better the model fits when compared to a null model in which all observed variables are uncorrelated. Adequate fit is indicated when a value greater than .90 is observed (Byrne, 2001).

The TLI yields values between 0 to 1.00. The TLI is not systematically related to sample size and is useful when comparing nested models, as it appropriately rewards model parsimony. A TLI of .95 or greater indicates good fit, with a value of .90 representing the lower bound of what is acceptable. A change of less than .01 between nested models means the null hypothesis should be retained (Cheung & Rensvold, 2002).

The RMSEA indicates how well the model may hold in the population. A RMSEA of ≤ .06 indicates good model fit, whereas values ranging from .08 to .10 indicate mediocre fit (Hu & Bentler, 1999). The 90% confidence interval accompanying the RMSEA indicates the precision of the yielded value when evaluating model fit (Byrne, 2001).

**RESULTS**

Model Fit

The hypothesized model for the LSS indicated adequate fit to the data (Table 1) and was commensurate with previous research findings. Significant chi-square statistics were obtained over

<table>
<thead>
<tr>
<th>TABLE 1</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goodness-of-Fit Indices for the First Order Models for the LSS Over Four Time Points</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Df</strong></td>
</tr>
<tr>
<td>First Order</td>
</tr>
<tr>
<td>Time 1</td>
</tr>
<tr>
<td>Time 2</td>
</tr>
<tr>
<td>Time 3</td>
</tr>
<tr>
<td>Time 4</td>
</tr>
</tbody>
</table>

*Note. Df = degrees of freedom; χ² = chi square; TLI = Tucker Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval. *p < .001.
the four time points; however, emphasis was placed on the additional goodness-of-fit statistics, due to the previously mentioned limitations in using the chi-square statistic. The TLI and CFI were higher than .8 for times one, two, and four, indicating adequate but marginal fit; however, this is below the recommended threshold for good fit of > .95 (Hu & Bentler, 1999). The RMSEA of .063 for time one and time two indicated a good fit of the observed model to the population. The goodness-of-fit statistics at time three indicated lower model fit (see Table 1). The TLI and CFI were low; however, the RMSEA of .097 suggested an adequate fit. Tight (± .01) confidence intervals accompanied the RMSEA for each time point, indicating a high level of precision in the RMSEA in reflecting the fit of the model to the population.

Descriptive Statistics

In Table 2 are the descriptive statistics for the five leadership scales averaged across all teams over the four time points. The scores for each subscale are best interpreted across time rather than the magnitude between subscales, as item numbers in each scale varied, with the lowest number of items in the rewarding subscale. In general, the means across all teams show an increase over the season, with the associated standard deviations reflecting the variability between the teams to be moderately large. This is in line with the high Intraclass Correlation Coefficients (ICCs) within teams, which suggested less variability within teams (see below). The five subscales of the LSS showed moderate to high internal consistency across all four time points (Cronbach’s $\alpha =$ .75 to .96, mean $\alpha =$ .89; see Table 2), which, with the exception of democratic behavior, were reasonably stable over time.

Data Aggregation

As the members of each team have the same coach, perceptions of leadership is nested with teams. To determine whether leadership is a team-level construct, participants’ data were aggregated into teams. One-way analyses of variances (ANOVAs) were undertaken for the total scores obtained at each time-point on the LSS to determine if variance was greater between teams or within teams. If homogeneity of responses between teams exists, the aggregation of individual responses to represent team characteristics is justified (Moritz & Watson, 1998). The resulting $F$ values from
TABLE 3
Intraclass Correlation Coefficient for the Subscales of the LSS

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
<th>Time 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training and instruction</td>
<td>.80</td>
<td>.84</td>
<td>.87</td>
<td>.81</td>
</tr>
<tr>
<td>Democratic</td>
<td>.78</td>
<td>.84</td>
<td>.76</td>
<td>.84</td>
</tr>
<tr>
<td>Autocratic</td>
<td>.88</td>
<td>.92</td>
<td>.89</td>
<td>.91</td>
</tr>
<tr>
<td>Social support</td>
<td>.81</td>
<td>.82</td>
<td>.84</td>
<td>.79</td>
</tr>
<tr>
<td>Rewarding</td>
<td>.71</td>
<td>.75</td>
<td>.86</td>
<td>.71</td>
</tr>
</tbody>
</table>

All ANOVAs were statistically significant \( p < .05 \), providing evidence that there is stronger agreement within than between teams in perceptions of leadership behavior.

A limitation of the ANOVA procedure and accompanying \( F \) test is that it does not provide information on the strength of the team effect. An ICC was calculated to indicate the level of agreement within teams. A high ICC indicates a strong level of agreement amongst team members (Schutz, 1998). Table 3 indicates that the ICC’s for all five subscales increase between time one and time two, with this trend continuing at time three for the training and instruction, social support, and positive feedback behavior subscales. The lowest level of agreement between team members was found for positive feedback behavior, with ICCs for three of the four time periods between \( r = .71 \) and \( .75 \). Overall, these results indicate a high level of agreement between team members in their perceptions of leadership behavior; however, this level of agreement fluctuates somewhat over time.

### Structural Stability

Model fit was assessed by comparing fit indices of each progressive model to an appropriate independence model (see Table 4). Progression through each subsequent test of invariance indicated deterioration in goodness of fit for each subscale. In four out of the five models (excluding autocratic), the chi-square statistic yielded significant results when weak invariance criteria were imposed. In light of the previously mentioned caution regarding the interpretation of the chi-square statistic, emphasis was placed on the \( \Delta \)CFI, as recommended by Cheung and Rensvold (2002). The consideration of this indicator suggested that weak factorial invariance existed for the training, democratic, social support, and positive feedback behavior subscales of the LSS. The exception, however, was the autocratic subscale which reached strong factorial invariance.

### Differential Stability

The constraints of weak factorial invariance were maintained, and differential stability coefficients for each model between each time period were calculated (Table 5). The training, democratic, autocratic, and social support subscales all met or exceeded the lower acceptable bound of reliability \( r = \geq .70 \) suggested by Nunnally and Bernstein (1994). The positive feedback behavior subscale, with the exception of time one to time two \( (.82) \) exhibited borderline differential stability \( (.67 \text{ to } .72) \). Overall, these results suggest an acceptable level of differential stability for four of the five dimensions of leadership as measured by the LSS.
### Table 4
Goodness-of-Fit Indices for Sequential Longitudinal Invariance Tests

<table>
<thead>
<tr>
<th>Scale</th>
<th>df</th>
<th>$\chi^2$</th>
<th>df diff.</th>
<th>$\Delta \chi^2$</th>
<th>TLI</th>
<th>CFI</th>
<th>$\Delta CFI$</th>
<th>RMSEA (90% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independence</td>
<td>1378</td>
<td>8689.6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.156 (.153–.159)</td>
</tr>
<tr>
<td>Configural</td>
<td>1190</td>
<td>2078.4</td>
<td>—</td>
<td>—</td>
<td>.86</td>
<td>.88</td>
<td>—</td>
<td>.058 (.054–.063)</td>
</tr>
<tr>
<td>Weak</td>
<td>1226</td>
<td>2132.3</td>
<td>36</td>
<td>53.9*</td>
<td>.86</td>
<td>.88</td>
<td>.00</td>
<td>.062 (.058–.066)</td>
</tr>
<tr>
<td>Strong</td>
<td>1265</td>
<td>2340.4</td>
<td>39</td>
<td>208.1*</td>
<td>.84</td>
<td>.85</td>
<td>.03</td>
<td>.062 (.058–.066)</td>
</tr>
<tr>
<td><strong>Democratic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independence</td>
<td>666</td>
<td>4854.7</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.169 (.165–.174)</td>
</tr>
<tr>
<td>Configural</td>
<td>534</td>
<td>845.1</td>
<td>—</td>
<td>—</td>
<td>.91</td>
<td>.93</td>
<td>.00</td>
<td>.052 (.045–.058)</td>
</tr>
<tr>
<td>Weak</td>
<td>558</td>
<td>858.7</td>
<td>24</td>
<td>13.6</td>
<td>.91</td>
<td>.93</td>
<td>.00</td>
<td>.050 (.043–.056)</td>
</tr>
<tr>
<td>Strong</td>
<td>585</td>
<td>955.7</td>
<td>27</td>
<td>97*</td>
<td>.90</td>
<td>.91</td>
<td>.02</td>
<td>.054 (.048–.060)</td>
</tr>
<tr>
<td><strong>Autocratic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independence</td>
<td>210</td>
<td>2563.01</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.226 (.218–.234)</td>
</tr>
<tr>
<td>Configural</td>
<td>134</td>
<td>183.8</td>
<td>—</td>
<td>—</td>
<td>.97</td>
<td>.98</td>
<td>—</td>
<td>.041 (.025–.055)</td>
</tr>
<tr>
<td>Weak</td>
<td>146</td>
<td>185.4</td>
<td>12</td>
<td>1.6</td>
<td>.98</td>
<td>.98</td>
<td>.00</td>
<td>.035 (.016–.050)</td>
</tr>
<tr>
<td>Strong</td>
<td>161</td>
<td>235.3</td>
<td>15</td>
<td>49.9*</td>
<td>.96</td>
<td>.97</td>
<td>.01</td>
<td>.046 (.033–.058)</td>
</tr>
<tr>
<td><strong>Social Support</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independence</td>
<td>528</td>
<td>4653.5</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.189 (.184–.194)</td>
</tr>
<tr>
<td>Configural</td>
<td>410</td>
<td>742.9</td>
<td>—</td>
<td>—</td>
<td>.90</td>
<td>.92</td>
<td>—</td>
<td>.061 (.054–.068)</td>
</tr>
<tr>
<td>Weak</td>
<td>431</td>
<td>773.5</td>
<td>21</td>
<td>30.6</td>
<td>.90</td>
<td>.92</td>
<td>.00</td>
<td>.060 (.053–.067)</td>
</tr>
<tr>
<td>Strong</td>
<td>453</td>
<td>1120.5</td>
<td>22</td>
<td>347*</td>
<td>.81</td>
<td>.84</td>
<td>.02</td>
<td>.082 (.076–.088)</td>
</tr>
<tr>
<td><strong>Rewarding</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independence</td>
<td>210</td>
<td>2678.0</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.232 (.048–.072)</td>
</tr>
<tr>
<td>Configural</td>
<td>134</td>
<td>248.0</td>
<td>—</td>
<td>—</td>
<td>.93</td>
<td>.95</td>
<td>—</td>
<td>.062 (.050–.074)</td>
</tr>
<tr>
<td>Weak</td>
<td>146</td>
<td>261.4</td>
<td>12</td>
<td>13.4</td>
<td>.93</td>
<td>.95</td>
<td>.00</td>
<td>.060 (.048–.072)</td>
</tr>
<tr>
<td>Strong</td>
<td>161</td>
<td>347.3</td>
<td>15</td>
<td>85.9*</td>
<td>.90</td>
<td>.92</td>
<td>.03</td>
<td>.073 (.062–.083)</td>
</tr>
</tbody>
</table>

*Note.* $df =$ degrees of freedom; diff. = difference; TLI = Tucker Lewis index; CFI = comparative fit index; RMSEA = root-mean-square error of approximation; CI = confidence interval.

$p < .05.$

### Table 5
Differential Stability Coefficients for the LSS Under Weak Invariance Criteria

<table>
<thead>
<tr>
<th></th>
<th>Time 1–2</th>
<th>Time 2–3</th>
<th>Time 3–4</th>
<th>Time 1–3</th>
<th>Time 2–4</th>
<th>Time 1–4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>.88</td>
<td>.81</td>
<td>.77</td>
<td>.79</td>
<td>.80</td>
<td>.78</td>
</tr>
<tr>
<td>Democratic</td>
<td>.93</td>
<td>.84</td>
<td>.78</td>
<td>.77</td>
<td>.78</td>
<td>.75</td>
</tr>
<tr>
<td>Autocratic</td>
<td>.91</td>
<td>.87</td>
<td>.89</td>
<td>.81</td>
<td>.91</td>
<td>.90</td>
</tr>
<tr>
<td>Social support</td>
<td>.91</td>
<td>.90</td>
<td>.88</td>
<td>.86</td>
<td>.89</td>
<td>.87</td>
</tr>
<tr>
<td>Rewarding</td>
<td>.82</td>
<td>.71</td>
<td>.72</td>
<td>.67</td>
<td>.71</td>
<td>.67</td>
</tr>
</tbody>
</table>

### Latent Mean Stability

The evaluation of latent mean stability offers the best indication of differences in perceptions of leadership behavior over time. However, evaluation of latent means was not undertaken, as strong factorial invariance was not found for the LSS. To gain an indication of changes in perceptions
of leadership a repeated measures multivariate analysis of variance (MANOVA) was undertaken. The omnibus $F$ test failed to reject the null hypothesis that there is no significant difference in means of each subscale of the LSS over time. As the assumption of sphericity was violated, the Greenhouse-Geisser statistic was interpreted. The Greenhouse-Geisser corrects for the violation of sphericity, therefore reducing the risk of making a type 1 error (Dancey & Reidy, 2002). Evaluation of each subscale of the LSS independently revealed that perceptions of leadership significantly changed over time for the social support ($F = 18.69, p < .001$), democratic ($F = 12.03, p < .001$), autocratic ($F = 3.73, p < .001$), rewarding positive feedback ($F = 24.97, p < .001$), and training behavior ($F = 29.25, p < .001$) subscales.

Although these tests show that there is a significant change in perceptions of leadership behavior over the season, it does not show specifically when these changes occurred. Figure 1 was evaluated to pinpoint which time points for each subscale may reflect significant changes in leadership behavior. Mean scores for each subscale on the LSS were evaluated, and two-tailed paired sample $t$ tests were evaluated at points of interest. The significance level was specified at .01, to lower the chance of committing a type 1 error.

Athletes perceived autocratic behavior to be the lowest frequency leadership behavior (Figure 1), and perceptions of this behavior appeared to slowly increase over the season. It appears that the most change in perceived autocratic behavior occurs at between time two and time three, and a paired samples $t$ test shows that this change was significant ($t_{152} = 2.76, p = .007$).

Social support behavior was perceived to be the next most infrequent leadership behavior, and team means for perceptions of social support behavior decreased across the season. Statistical significance was found for the perceived decrease in social support behavior between time one ($M = 19.9, SD = 7.3$) and time two ($M = 21.2, SD = 8.0$), ($t_{149} = −7.15, p < .001$). Social support behavior continued to decrease at a slower rate between time two ($M = 21.2, SD = 8.0$) and time four ($M = 22.0, SD = 8.5$), ($t_{146} = −2.76, p = .007$). However, the high standard deviations obtained for this subscale suggest a large dispersion in scores for perceptions of social support.

Democratic behavior was initially perceived as the third most demonstrated leadership behavior in the first half of the season; however, this increase became the most frequent behavior at time

![FIGURE 1](image)
three and time four. Statistical significance was found for the increase in perceived democratic behavior between time two \( (M = 20.5, SD = 7.6) \) and time three \( (M = 23.1, SD = 10.2) \), \( t_{152} = -4.38, p < .001 \) [two-tailed]).

Training and instruction behavior alongside positive feedback behavior were the most frequently perceived leadership behaviors at time one, whereas perceptions of these behaviors decreased towards the end of the season. Statistical significance was found for the decrease in perceived training and instruction behavior between time two and time three \( (t_{147} = -6.71, p < .001 \) [two-tailed]). The observed increase in mean perceptions of training and instruction between time three \( (M = 30.0, SD = 11.9) \) and time four \( (M = 28.2, SD = 11.0) \) was not statistically significant \( (t_{151} = 2.15, p < .03) \).

Athletes perceived little change in positive feedback behavior between time one \( (M = 8.4, SD = 3.8) \) and time two \( (M = 9.0, SD = 4.0) \); and between time three \( (M = 10.9, SD = 5.0) \) and time four \( (M = 10.1, SD = 4.8) \). However, perceptions of positive feedback behavior showed a significant decrease between time two and time three \( (t_{147} = -6.71, p < .001 \) [two-tailed]).

**DISCUSSION**

The results from this study support Chelladurai and Saleh’s (1980) hypothesized 5-factor structure of the LSS in a team of elite female athletes. Subsequent longitudinal invariance tests revealed a minimum of weak factorial invariance for the LSS. Support was found for the hypothesis that perceptions of leadership behavior are similar within teams, and that these perceptions change over the course of the season.

Three out of the four time points yielded equal or better goodness-of-fit statistics than those found by Chelladurai and Riemer (1998). The weaker model fit at time three was interesting to note, and it is difficult to know why this occurred, although it is worth noting the reliability estimates were all strong for this time point. A potential reason for this poorer model fit might be that this was a crucial point in the season where play-offs loomed, and perhaps this may have affected the players’ concentration when attending to form filling. As it stands, there were three time points where the LSS psychometric properties were well supported.

Generally, the reliability estimates for the various subscales were high across all the time points. Although some studies have reported lower estimates of reliability for the autocratic behavior subscale (Chelladurai & Saleh, 1980; Chelladurai & Carron, 1981), the results from this study are in line with those from Hastie (1993, 1995). A potential reason for the similarity in these lower estimates might be the similarity of sample participants (both female samples) and the type of sports played (netball and volleyball).

**Structural Stability**

Progression through Meredith’s (1993) hierarchy of increasingly stringent tests of invariance suggested a minimum of weak factorial invariance for the 5-factor structure of the LSS, with autocratic behavior reaching strong factorial invariance.

From a psychometric perspective, this is the first study to evaluate the psychometric properties of the LSS over the course of a competition season, and the results are important in enabling
researchers to confidently evaluate changes in perceptions of leadership behavior longitudinally using this measure. The weak factorial invariance achieved when testing the LSS over four time points provides good evidence that the LSS is consistently measuring perceptions of leadership behavior over time.

**Differential Stability**

This study found support for the differential stability of all five subscales in the LSS across the 10-week competition period. High differential stability is essential when evaluating change over time, as it indicates the stability or consistency of individual differences over time relative to other team members (Schutz, 1998). The high level of differential stability obtained for the LSS suggests that, although perceptions of leadership behavior changed over the course of the season, individuals’ perceptions of this behavior relative to other team members was consistent.

**Perceptions of Leadership**

The results of this study found that athletes perceive leadership behavior to change over the 10-week competition period, in particular, training and instruction, positive feedback, and democratic behavior. Athletes also perceived autocratic behavior to be the least-frequent leadership behavior displayed by their coach. The low level of perceived autocratic behavior in this study is congruent with previous studies which have shown female athletes prefer (Sherman Fuller, & Speed, 2000) and perceive (Riemer & Toon, 2001) low levels of autocratic behavior. Furthermore, the level of perceived autocratic behavior is similar to perceptions in other studies using team sports athletes (Sherman et al., 2000; Riemer & Chelladurai, 1995).

The low levels of perceived autocratic behavior may reflect the nature of the sport, as netball is a sport where players must think on their feet and react to the changing dynamics of the game. Therefore, an autocratic style of decision making from the coach may be inappropriate, given the experience level of the players and the high variability of the task. The increase in perceived autocratic behavior at time three may reflect the stress of the competition at that time, causing the coach to more frequently use her authority to direct the team.

Although democratic behavior at time one and time two was perceived to be the third most frequent leadership behavior, perceptions of this behavior significantly increased at time three to become the most frequently perceived leadership behavior. Furthermore, perceptions of democratic behavior between time three and time four showed no significant changes. A democratic decision-making style was preferred by Australian athletes (Sherman et al., 2000), and female athletes when compared to males (Riemer and Toon, 2001). Furthermore, coaches interviewed in a qualitative study expressed a strong preference for democratic behavior (Weinberg & McDermott, 2002). However, research has not yet examined whether these preferences change in response to different situational factors or over time.

The Normative Model of Decision Styles for Coaching (Chelladurai & Haggerty, 1978) suggests that the coach should use a participative style of decision making when the coach does not have high relative information, the problem is not complex, when team acceptance is critical, and when the team is integrated. Therefore, the lower levels of democratic behavior demonstrated at the start of the season may reflect the coach preferring their expert knowledge, as they may have
The emphasis on training and instruction early in the season may reflect the coach trying to build cohesion in the newly formed team and develop skills in preparation for the increased competition later in the season. Chelladurai and Saleh (1980) suggest that positive feedback behavior and training and instruction are linked, as positive feedback behavior is considered to build the confidence of team members as they develop new skills. This behavior could also be emphasized early in the season to build relationships and enhance teamwork within the newly formed team.

The requirements for training and instruction could decrease in the second half of the season, as the team may have honed their teamwork and skills earlier in the season. Furthermore, the team is likely to have experienced a number of wins over the season and are eagerly anticipating the increased level of competition during play-offs. The combination of the experience of working
as a team and participating in competitions is likely to enhance the efficacy and motivation of the team, and therefore, team members are likely to require less training and instruction behavior from the coach to maintain their motivation.

Homogeneity in Perceptions of Leadership Behavior

Previous researchers using the LSS to evaluate perceived leadership behavior have averaged player perceptions with the underlying assumption of the homogeneity of leadership behavior within teams. However, there has been conflicting evidence as to whether this assumption is valid (Chelladurai & Riemer, 1998).

The calculated ICCs in this study suggest that there is a strong level of agreement between team members in their perceptions of leadership behavior. Although this level of agreement fluctuates slightly over time, it remains strong enough to justify data aggregation from individual players to determine perceived leadership behavior for all subscales except positive feedback behavior. The lower level of agreement amongst team members in relation to positive feedback behavior may suggest that the coach demonstrates differing levels of positive feedback behavior amongst team members.

These findings offer some support for the suggestion by coaches in Weinberg and McDermott’s (2002) study where coaches suggested they vary their behavior in response to individual needs. The findings of this study suggest that the level of variation in perceptions of leadership behavior are low within teams and that the coaches appears to vary their positive feedback behavior to a greater extent. Chelladurai and Saleh (1980) suggest that positive feedback behavior is a motivational aspect of leadership behavior and is determined by the situational and member characteristics. Therefore, the lower ICCs found for perceptions of positive feedback behavior when compared to those found for other leadership behaviors may reflect the coach’s demonstration of differing amounts of positive feedback behavior between team members as their motivation fluctuates.

Future Directions

While this study found evidence for the longitudinal stability and invariance of the LSS, it is important to note that the nested structure of the data should be considered when interpreting the results. This suggests that more sophisticated approaches that combine structural equation modeling (SEM) and hierarchical linear modeling (HLM) might be needed next, to build on the findings in this study. Our smaller sample of teams meant that the minimum level two units (i.e., teams) required for HLM/SEM approaches was insufficient for such analyses.

Conclusions

This is the first longitudinal study using the LSS, and, as such, offers a unique contribution to existing leadership research. Prior to this study, the stability of the psychometric properties of the LSS had not been established, despite being used extensively to assess leadership within sport. The finding of weak factorial invariance enables researchers to feel confident that the LSS
is measuring perceptions of leadership consistently over time. Furthermore, the high level of
differential stability found for the LSS suggests that, although absolute values for perceptions of
leadership behavior change, individuals maintain their relative position within the team in their
perceptions of leadership behavior.

The finding that perceptions of leadership change over time is an important contribution to
leadership research, as the majority of studies in business and sport have been cross sectional in
nature. Furthermore, the high level of structural stability found for the LSS suggests these are
true reflections of athlete’s perceptions of leadership behavior and not due to measurement error.

This study also offers an important insight into the nature of team perceptions of leadership
behavior. Team members showed a high level of agreement in their perceptions of leadership
behavior, which is suggestive of a high degree of homogeneity of leadership behavior. However,
where perceptions of leadership varied, the high differential stability of the LSS suggests that ath-
etes maintained their relative position amongst team members in their perceptions. The finding
of homogeneity in perceptions provides researchers with some justification in aggregating team
level data in future research using the LSS.

The finding in this study that sporting leadership behavior changes over time has implications
for broader leadership theory. The cross-sectional nature of most business leadership research is a
significant limitation in understanding the nature of leadership. At present, few studies have eval-
uated leadership behavior longitudinally, instead offering the inherent assumption that leadership
behavior is static within a particular context (Kroec, Lowe, & Brown (2004). This assumption
is potentially a contributing cause for the seemingly contradictory findings in much of the lead-
ership research. Through conducting longitudinal studies into leadership behavior, researchers
can begin to understand which behaviors enhance productivity at different stages of the journey
towards the current organizational goal.

REFERENCES


Byrne, B. M. (2001). Structural equation modelling with AMOS: Basic concepts, applications, and programming.
Mahwah, NJ: Lawrence Erlbaum Associates.


6–9.

exercise psychology measurement (pp. 227–253). Morgantown: Fitness Information Technology.

Journal of Sport Psychology, 2, 34–45.

Structural Equation Modelling, 9, 233–255.

appraisals. Structural Equation Modelling, 10, 401–422.

UK: Prentice Hall.


