



138  
684  
THS

1  
000  
63660767

This is to certify that the  
dissertation entitled

ATHLETES' EVALUATIONS OF THEIR HEAD COACH'S  
COACHING COMPETENCIES: A MULTILEVEL CONFIRMATORY  
FACTOR ANALYSIS

presented by

Nicholas Daniel Myers

has been accepted towards fulfillment  
of the requirements for the

Dual-Major  
Doctoral

degree in

Department of Kinesiology &  
Department of Counseling,  
Educational Psychology,  
Special Education

Edward W. K. / Daniel L. Feltz  
Major Professors' Signatures

12/01/04

Date

MSU is an Affirmative Action/Equal Opportunity Institution

**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.  
**MAY BE RECALLED** with earlier due date if requested.

| DATE DUE     | DATE DUE | DATE DUE |
|--------------|----------|----------|
| 6,500 652208 |          |          |
| MAY 08 2010  |          |          |
|              |          |          |
|              |          |          |
|              |          |          |
|              |          |          |
|              |          |          |
|              |          |          |
|              |          |          |
|              |          |          |

**ATHLETES' EVALUATIONS OF THEIR HEAD COACH'S COACHING  
COMPETENCIES: A MULTILEVEL CONFIRMATORY FACTOR ANALYSIS**

**By**

**Nicholas Daniel Myers**

**A DISSERTATION**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**DOCTOR OF PHILOSOPHY-DUAL MAJOR**

**Department of Kinesiology**

**Department of Counseling, Educational Psychology, and Special Education**

**2005**

## **ABSTRACT**

### **ATHLETES' EVALUATIONS OF THEIR HEAD COACH'S COACHING COMPETENCIES: A MULTILEVEL CONFIRMATORY FACTOR ANALYSIS**

**By**

**Nicholas Daniel Myers**

**This study (a) provided initial validity evidence for the Coaching Competency Scale (CCS), and (b) introduced multilevel confirmatory factor analysis (MCFA) as an appropriate methodology to use when data are meaningfully nested and an evaluation of the factor structure of a set of indicators is desired. Data were collected from intercollegiate men's ( $g = 8$ ) and women's ( $g = 13$ ) soccer and women's ice hockey teams ( $g = 11$ ). Results offered some support for the proposed multilevel multidimensional conceptualization of coaching competency, the internal consistency reliabilities of the coaching competency estimates, and a relationship between motivation competency and satisfaction with the coach within teams. Validity concerns were observed for the original rating scale structure and the relationship between motivation competency and satisfaction with the coach between teams. Results were interpreted to guide future research with the CCS, to provide recommendations for revisions to the instrument, and to assist researchers in physical education and exercise science in understanding when and how to apply MCFA to their data.**

## DEDICATION

This work is dedicated, in part, to my partner, Ahnalee. Thank you for hanging in there over the past five years, babe. I look forward to our life together.

This work is also dedicated, in part, to my family and friends. Please know that although you have often been pushed aside over the last few years, you have never been forgotten. Thank you for your love, companionship, and strength. I take you with me wherever I go.

## ACKNOWLEDGEMENTS

**Deb Feltz:** Thank you for your mentorship, guidance, and support over the last five years. And, thank you for providing an opportunity to an applicant who had little “traditional” background five years ago. Nobody has had a larger impact on my graduate training.

**Ed Wolfe:** Thank you for your mentorship, excellent teaching in complex subjects, and active advisement over the last three years. Your decision to write a note on my final exam three years ago, asking me if I ever thought about pursuing a degree in MQM, has changed my career. Thank you for taking the time that such insight and curiosity require.

**Kim Maier:** Thank you for your excellent teaching in complex subjects, willingness to consult, and encouragement in my job search. Although we have not known each other for long, I hope that our paths will cross at regular intervals down the road.

**Mark Reckase:** Thank you for your excellent teaching in complex subjects, and honesty in our interactions. And, thank you for packaging your deep knowledge and experience in probes that are available to people, like me, with considerably less knowledge and experience. You provide a model that many students in MQM aspire to emulate.

## TABLE OF CONTENTS

|                                                             |      |
|-------------------------------------------------------------|------|
| LIST OF TABLES.....                                         | vii  |
| LIST OF FIGURES.....                                        | viii |
| KEY TO SYMBOLS AND ABBREVIATIONS.....                       | ix   |
| <br><b>CHAPTER 1</b>                                        |      |
| <b>INTRODUCTION AND REVIEW OF COACHING EFFECTIVENESS</b>    |      |
| LITERATURE.....                                             | 1    |
| Nature of the Problem.....                                  | 1    |
| Review of Coaching Effectiveness Literature.....            | 3    |
| Establishing an Initial Validity Framework for the CCS..... | 6    |
| Interpretive framework.....                                 | 7    |
| Instrument development.....                                 | 7    |
| Instrumentation.....                                        | 8    |
| External model.....                                         | 9    |
| Internal model.....                                         | 9    |
| Statement of Purpose.....                                   | 12   |
| Research Questions.....                                     | 12   |
| <br><b>CHAPTER 2</b>                                        |      |
| INTRODUCTION OF MCFA.....                                   | 13   |
| Consequences of Ignoring Multilevel Data Structures.....    | 14   |
| Steps in a MCFA.....                                        | 15   |
| A Technical Synopsis.....                                   | 15   |
| Sample size.....                                            | 19   |
| Judging the fit of MCFA.....                                | 19   |
| An Application of Muthén's Steps.....                       | 21   |
| <br><b>CHAPTER 3</b>                                        |      |
| METHOD.....                                                 | 23   |
| Sample.....                                                 | 23   |
| Procedure.....                                              | 23   |
| Measures.....                                               | 24   |
| Coaching competence.....                                    | 24   |
| Satisfaction with the coach.....                            | 24   |
| Treatment of Data.....                                      | 25   |
| Missing data.....                                           | 25   |
| Outliers.....                                               | 25   |
| Normality.....                                              | 26   |
| Analyses.....                                               | 26   |
| Rating scale.....                                           | 26   |

|                                                                                                             |    |
|-------------------------------------------------------------------------------------------------------------|----|
| Internal models.....                                                                                        | 30 |
| Internal consistency reliability.....                                                                       | 31 |
| Forming measures to test questions 4 and 5.....                                                             | 32 |
| Testing questions 4 and 5.....                                                                              | 33 |
| Model estimation and fit.....                                                                               | 36 |
| Reliability estimates.....                                                                                  | 36 |
| <br>CHAPTER 4                                                                                               |    |
| RESULTS.....                                                                                                | 37 |
| Did Athletes Employ the Rating Scale Structure in the Manner that the Authors<br>Intended?.....             | 37 |
| To What Degree did the Proposed Internal Models fit the Data?.....                                          | 40 |
| Step 1.....                                                                                                 | 40 |
| Step 2.....                                                                                                 | 40 |
| Step 3.....                                                                                                 | 44 |
| Step 4.....                                                                                                 | 48 |
| How Reliable were the Rank Ordering of Coaching Competency Estimates?.....                                  | 55 |
| Were Coaching Competency Estimates Positively Related to Satisfaction with the<br>Coach Within Teams?.....  | 55 |
| Were Coaching Competency Estimates Positively Related to Satisfaction with the<br>Coach Between Teams?..... | 58 |
| <br>CHAPTER 5                                                                                               |    |
| DISCUSSION.....                                                                                             | 60 |
| <br>APPENDICES.....                                                                                         | 67 |
| <br>REFERENCES.....                                                                                         | 69 |
| <br>FOOTNOTES.....                                                                                          | 75 |

## LIST OF TABLES

|                                                                                                                             |       |
|-----------------------------------------------------------------------------------------------------------------------------|-------|
| Table 1 Original and Post Hoc Rating Scale Structures.....                                                                  | 38-39 |
| Table 2 Item Characteristics for the CCS.....                                                                               | 41    |
| Table 3 Pooled Within-Teams Correlations and Covariances.....                                                               | 42    |
| Table 4 Scales Between-Teams Correlations and Covariances.....                                                              | 43    |
| Table 5 Model-Data Fit Statistics.....                                                                                      | 47    |
| Table 6 Within-Teams and Between-Teams Estimates.....                                                                       | 50    |
| Table 7 Comparisons of Factor Loadings and Correlations Among Factors by Gender...                                          | 52    |
| Table 8 Comparisons of Factor Loadings and Correlations Among Factors by Sport.....                                         | 53    |
| Table 9 Comparisons of Factor Loadings and Correlations Among Factors by Year.....                                          | 54    |
| Table 10 Correlations Between Competency Judgments and Satisfaction with the Coach<br>at the Individual and Team-Level..... | 56    |
| Table 11 Hierarchical Linear Models where Satisfaction was the Dependent Variable...                                        | 59    |

## LIST OF FIGURES

|                                                                                    |    |
|------------------------------------------------------------------------------------|----|
| <i>Figure 1.</i> Horn's working model of coaching effectiveness.....               | 2  |
| <i>Figure 2.</i> Multidimensional internal model of the CCS.....                   | 11 |
| <i>Figure 3.</i> Multidimensional MUML model of CCS.....                           | 18 |
| <i>Figure 4.</i> Coherence example for a well-fitting five category structure..... | 29 |

## KEY TO SELECT ABBREVIATIONS

Any format acceptable to the departments may be used.

$\alpha$ : Cronbach's coefficient alpha  
AERA: American Educational Research Association  
APA: American Psychological Association  
CAIC: consistent Akaike information criterion  
CBAS: coaching behavior assessment system  
CBC: character building competence  
CBQ: coaching behavior questionnaire  
CCS: coaching competency scale  
CEQ: coaching evaluation questionnaire  
CES: coaching efficacy scale  
CFA: confirmatory factor analysis  
CFI: comparative fit index  
 $\delta$ : item difficulty  
DQS: decision style questionnaire  
EFA: exploratory factor analysis  
FIML: full maximum likelihood  
GSC: game strategy competence  
ICC: intraclass correlation coefficient  
LM: Lagrange multiplier test  
LSS: leadership scale for sports  
MC: motivation competence  
MCFA: multilevel confirmatory factor analysis  
MNSQ: outfit mean square fit statistic  
MUML: Muthén's maximum likelihood  
NASPE: National Association for Sport and Physical Education  
NCME: National Council on Measurement in Education  
RMSEA: root mean square error of approximation  
RSM: rating scale model  
 $S^*_B$ : scaled between group covariance matrix  
 $\Sigma_B$ : between group population covariance matrix  
 $S_{PW}$ : within group covariance matrix  
SRMR: standardized root mean squared residual  
 $S_T$ : total covariance matrix  
 $\Sigma_W$ : within group population covariance matrix  
 $\tau$ : threshold  
TC: technique competence  
 $\theta$ : person ability  
TLI: Tucker-Lewis index

## CHAPTER 1

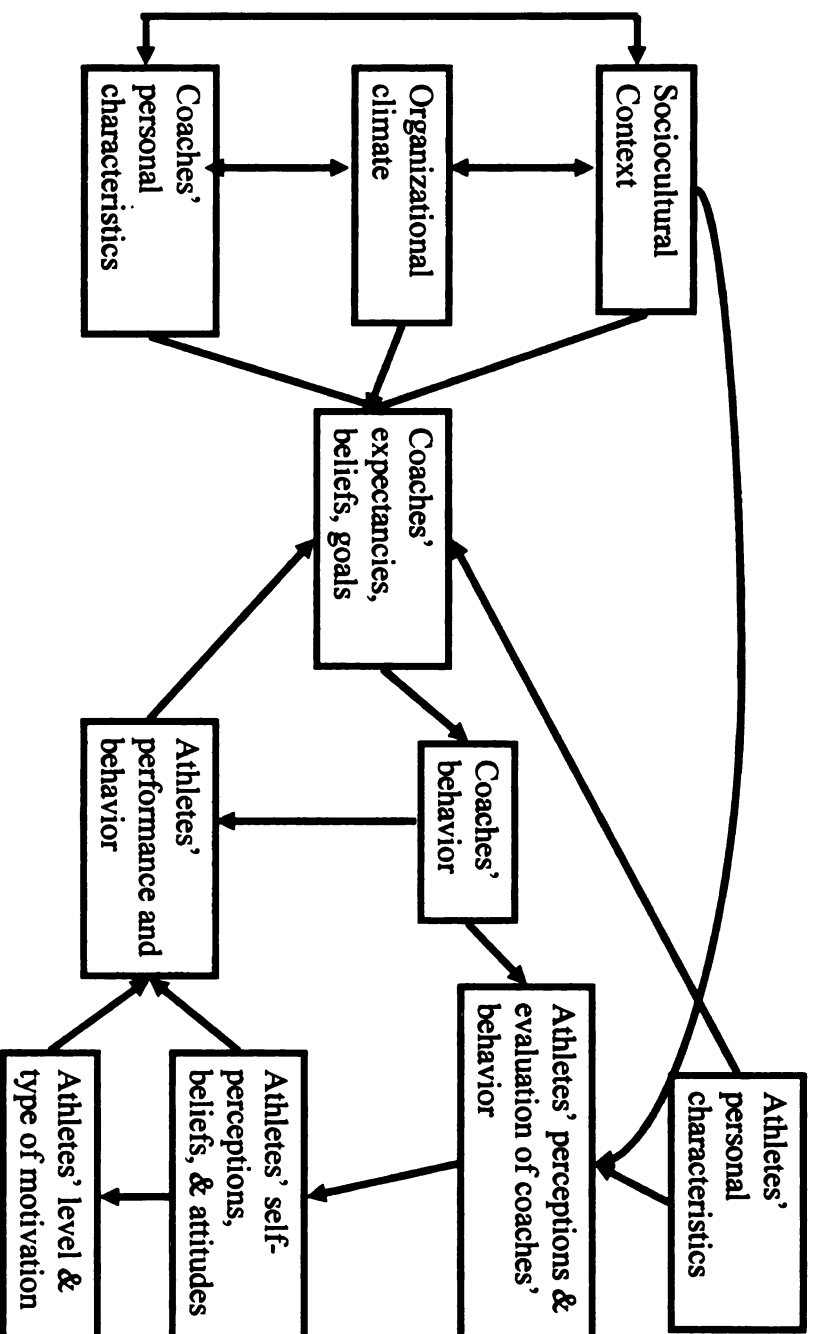
### INTRODUCTION AND REVIEW OF COACHING EFFECTIVENESS LITERATURE

#### *Nature of the Problem*

To date, much of the research in sport leadership has been directed toward identifying particular coaching styles that elicit successful performance and/or positive psychological responses from athletes (Horn, 2002). The two most prominent models of leadership effectiveness in sport, the Multidimensional Model of Leadership (Chelladurai, 1978) and the Mediation Model of Leadership (Smoll & Smith, 1989), have served as frameworks for much of the related research. Recently, Horn combined elements of both models to form a working model of coaching effectiveness.

Horn's (2002) model of coaching effectiveness, as displayed in Figure 1, is founded on at least three assumptions. First, both antecedent factors (e.g., coach's personal characteristics, organizational climate) and personal characteristics of athletes influence a coach's behavior indirectly through a coach's expectancies, beliefs, and goals. Second, a coach's behavior directly affects athletes' perceptions and evaluations of a coach's behavior. Third, athletes' perceptions and evaluation of a coach's behavior mediate the influence that a coach's behavior has on athletes' self-perceptions (e.g., self-efficacy) and attitudes (e.g., satisfaction with a coach), which in turn directly affects athletes' motivation and performance. Because athletes' perceptions and evaluation of a coach's behavior are believed to play a critical role in coaching effectiveness, providing a tool to assess athletes' evaluations of key coaching competencies is important to the continued improvement of coaching, and to the further development of coaching effectiveness models.

Figure 1. Horn's working model of coaching effectiveness.



### *Review of Coaching Effectiveness Literature*

There are many instruments, for which validity evidences already exists, designed to measure a coach's behavior.<sup>1</sup> The Coaching Behavior Assessment System (CBAS; Smith, Smoll, & Hunt, 1977), the Leadership Scale for Sports (LSS; Chelladurai & Saleh, 1978, 1980), and the Decision Style Questionnaire (DQS; Chelladurai & Arnott, 1985) are the most prominent of such instruments. As reviewed by Horn (2002), these instruments have also been used to assess athletes' perceptions of their coach's behavior (e.g., how often does your coach use positive reinforcement with athletes?) and/or decision styles (e.g., what type of decision style would your coach employ in order to select a team captain?). However, none of these instruments measure athletes' evaluations of their coach's behavior (e.g., how competent do you believe your coach is in teaching the skills of your sport?). While each of these instruments has contributed mightily to the understanding of coaching behavior over the last few decades, Smoll and Smith (1989) themselves note that, "the ultimate effects that coaching behavior exerts are mediated by the meaning that players attribute to them" (p. 1527), effects these instruments do not measure.

Unlike the CBAS, the LSS and the DQS, the Coaching Evaluation Questionnaire (CEQ; Rushall & Wiznuk, 1985) and the Coaching Behavior Questionnaire (CBQ; Kenow & Williams, 1992) were designed to assess athletes' evaluative reactions to specific aspects of their coach's behavior. Rushall and Wiznuk suggest that the CEQ allows athletes to evaluate a coach on his or her personal qualities, personal and professional relationships, organizational skills, and performance as a teacher and a coach. The stated purpose for collecting these measures is to provide immediate feedback

to coaches in order to enhance their ability to relate more effectively to athletes. Although the items are intended to measure separate constructs, scores are to be totaled across items or formed at the item-level. Psychometric evidence for the CEQ appears to be limited to item-level test-retest reliability. The CEQ rarely appears in the literature.

The CBQ allows athletes to evaluate what they perceive to be their coach's typical behavior, specifically his or her negative activation and supportiveness/emotional composure during competition against a top opponent. The purpose of collecting these measures appears to be to provide coaches with information on how a coach's negative activation/supportiveness and emotional composure affects their athletes' performance and relevant psychological states during competition against a top opponent. Recent psychometric work on CBQ provided evidence for the proposed two-factor structure, the internal consistency reliability, and the external aspect of the validity of measures from the instrument (Williams et al., 2003). Evidence for the use of the CBQ for the intended purposes is accumulating. The CBQ certainly provides a valuable tool to measure athletes' evaluative reactions to aspects of their coach's behavior, but it measures a fairly specific and narrow subset of coaching behavior (negative activation and supportiveness, and emotional composure) and in a rather targeted scenario (competition against a top opponent).

Three competency domains that are stipulated in the *National Standards for Athletic Coaches* (National Association for Sport and Physical Education [NASPE], 1995) and are not fully covered by the CBQ are (a) growth, development and learning of athletes, (b) psychological aspects of coaching, and (c) skills, tactics and strategies. Within the growth, development and learning of athletes, an expected competency is that

a coach provide instruction to develop specific motor skills. Within the psychological aspects of coaching domain, expected competencies include that a coach demonstrate effective motivational skills; and, that a coach conduct practices and competitions to enhance social/emotional growth and promote good sportsmanship in athletes. Within the skills, tactics and strategies domain, an expected competency is that a coach applies level-appropriate competitive strategies. An instrument that more fully measures athletes' evaluations of their head coach in these three competency domains is warranted.

The Coaching Efficacy Scale (CES; Feltz, Chase, Moritz, & Sullivan, 1999) was developed to measure a coach's belief in his or her ability to influence the learning and performance of his or her athletes. The specific factors measured—instructional technique, motivation, character building, and game strategy—purposely overlap with the expected competency domains articulated above. The CES has demonstrated psychometric qualities that provide an acceptable level of confidence in the efficacy measures produced (Feltz et al.; Lee, Malet, & Feltz, 2002; Myers, Wolfe, & Feltz, in press). Resulting measures have correlated with observed and reported coaching behavior (Feltz et al.; Myers, Vargas-Tonsing, & Feltz, in press), leadership styles (Sullivan & Kent, 2003), and athletes' self-efficacy, satisfaction with the coach, and team performance (Myers, Vargas-Tonsing, et al.; Vargas-Tonsing, Warners, & Feltz, 2003).

However, according to models of coaching effectiveness, why a coach's beliefs, like coaching efficacy, are related to athletes' self-perceptions and performance is due to the influence that these beliefs exert on a coach's behavior (see Figure 1). But, the influence that a coach's behavior exerts on athletes' self-perceptions, motivation, and performances is mediated, at least in part, by athletes' evaluations of their coach's

behavior. Thus, athletes' evaluations of their coach's competence in domains measured by the CES—instructional technique, motivation, game strategy, and character building—should be related to athletes' self-perceptions and attitudes, which in turn should be related to athletes' motivation and performance. But, because the CES was not originally designed to measure athletes' evaluations of their head coach's competencies, new evidence should be provided to demonstrate the validity of the use of CES measures for this purpose (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). From this point forward, when the CES is referred to as an instrument to measure coaching competencies it will be referred to as the Coaching Competency Scale (CCS).

#### *Establishing an Initial Validity Framework for the CCS*

Validity refers to the degree to which evidence and theory support the intended uses and interpretations of scores from a particular test (AERA, APA, & NCME, 1999). Validation is an ongoing, integrated, and value-laden judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on various modes of measurement (Messick, 1989). Important steps in the validation process for the CCS include providing a conceptual framework for scores derived from the instrument (i.e., an interpretive framework), revealing the instrument's development process, explaining how the construct is operationalized within the instrument (i.e., instrumentation), specifying how proposed components of coaching competency exert influence on responses to the items and how

the components are related to one another (i.e., an internal model), and predicting how measures of coaching competencies relate to other variables (i.e., an external model).

*Interpretive framework.* High school and lower division collegiate athletes of team sports comprise populations for which this version of the CCS is intended. The purpose of the instrument is to measure athletes' evaluation of their head coach's ability to affect the learning and performance of athletes. Measures are intended to be used to test relationships in models of coaching effectiveness (e.g., how do athletes' evaluations of their head coach's competencies affect athletes' self-perceptions, beliefs, and attitudes?) and to serve as a tool in the evaluation of coaching education programs (e.g., assess the ability of education programs to alter athletes' perceptions of their head coach's competencies). Contexts in which the CCS can be used include field, laboratory, and other educational settings. Scores on the CCS are conceptualized as being norm-referenced because interpretations are intended to compare a head coach's coaching competency scores to other coach's scores within a specified population, or to track changes in a coach's competency scores across time.

*Instrument development.* The development process of the CES (Feltz et al., 1999) served as a basis for beginning the validation process for using the instrument for this new purpose and is summarized here. The CES was developed during a 5-week seminar involving 11 coaches who had varying levels of experience in coaching and were graduate students in sport psychology. The *National Standards for Athletic Coaches* (NASPE, 1995), preliminary work on a coaching efficacy scale (Park, 1992), and a review of the coaching education literature, provided a framework for group discussions on the key components of coaching efficacy. Themes that emerged from the group's

discussions were reduced to: teaching technique, implementing game strategies, motivating athletes, and developing athletes' character. These themes represented many of the primary roles that previous research has suggested coaches perform (Gould, 1987).

The dimensions that emerged from the seminar led to the generation of 41 items. Items included the stem: "How confident are you in your ability to." The rating scale employed was a 10-point Likert scale with categories ranging from 0 (*not at all confident*) to 9 (*extremely confident*). Nine collegiate and scholastic coaches acted as content experts to evaluate the relevance of the items on a scale that ranged from 1 (*essential*) to 3 (*not essential*). Feedback from this group content led the research team to conclude that all items were potentially important indicators of coaching efficacy. However, 17 of the original items were later dropped after considering the results of factor analyses. All of the retained items (24) were determined relevant for the CCS. However, the stem was changed to: "How competent is your head coach in his or her ability to." And, the anchors of the rating scale were changed to 0 (*not at all competent*) and 9 (*extremely competent*).

*Instrumentation.* A substantive aspect of how coaching competence is operationalized within the CCS is the degree to which athletes employed the rating scale structure in the way the authors intended or "systematically" (Zhu, Updyke, & Lewandowski, 1997). Previous research with the CES (Myers, Wolfe, & Feltz, in press) and long-standing recommendations for measuring attitudes (Likert, 1932) suggest that this rating scale structure likely contains too many categories and that collapsing data to create fewer category distinctions should be considered in order to achieve an optimal

rating scale structure for the CCS. Criteria for determining this will be discussed in the Methods section.

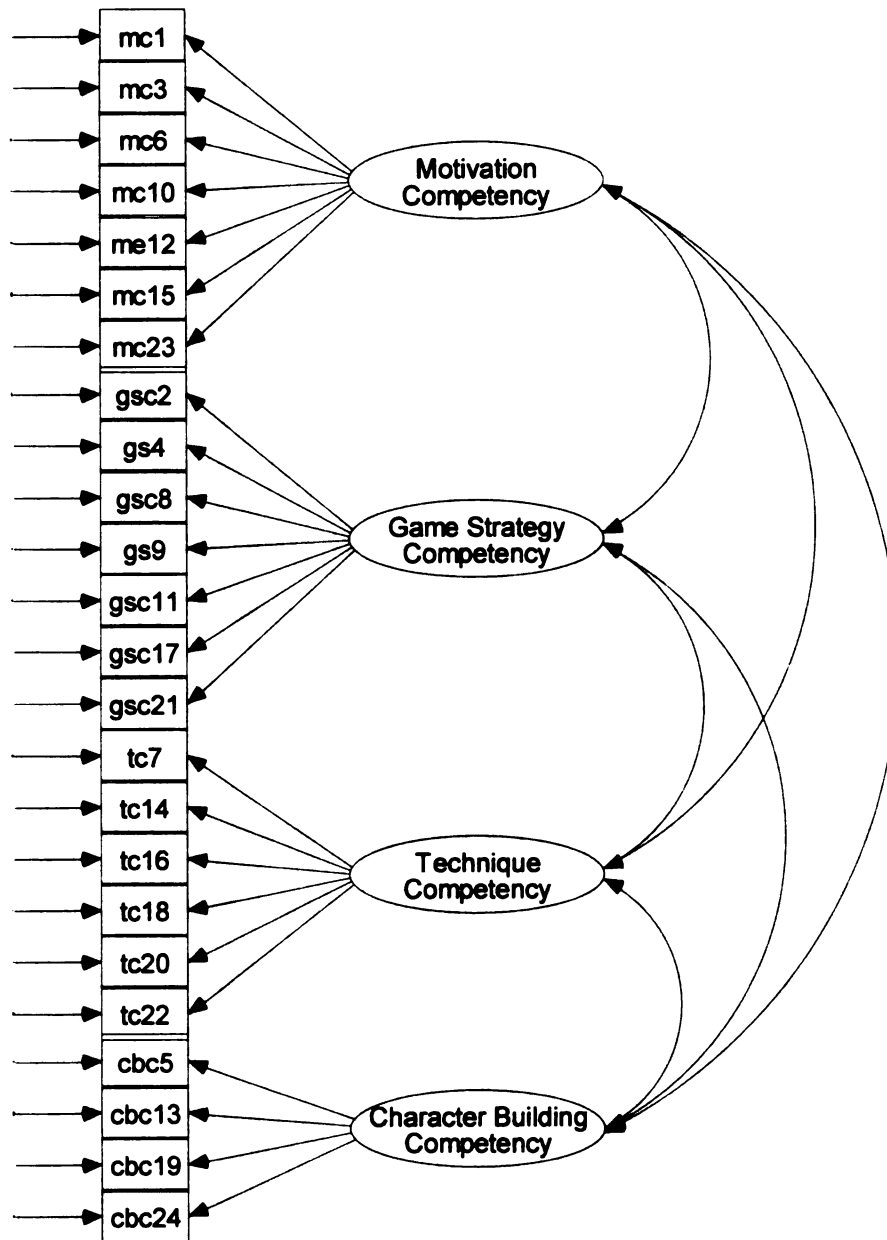
*External model.* According to Horn's (2002) model of coaching effectiveness (see Figure 1), a coach's behavior, athletes' characteristics (e.g., age, gender, psychological characteristics), and the sociocultural context directly affect athletes' evaluations of their head coach's competency. In turn, athletes' evaluations of their head coach's competency directly affect athletes' self-perceptions (e.g., efficacy beliefs) and attitudes (e.g., satisfaction with the coach). In accordance with the model, to establish initial and limited evidence for the external aspect of validity, I hypothesized that athlete and team measures of coaching competencies would be positively related to athlete and team satisfaction with the head coach, respectively.

*Internal model.* Feltz et al. (1999) put forth both an oblique multidimensional and a hierarchical internal model of the CES. The hierarchical model fit the data significantly worse than the oblique multidimensional model. Because the multidimensional factors tend to be moderately to highly correlated with one another (Feltz et al.; Lee et al., 2002), subsequent psychometric work has focused on comparing the fit of the multidimensional model to a unidimensional model (Myers, Wolfe, et al., in press). These authors concluded that the multidimensional model tends to fit the data better than a unidimensional model, but that either model can be justified depending on the purpose of the study.

The fit of both an oblique multidimensional model (see Figure 2) and a unidimensional model were explored for the CCS. Motivation competence (MC) is specified to influence responses to seven items and is defined as athletes' evaluations of

their head coach's ability to affect the psychological mood and skills of athletes. Game strategy competence (GSC) is specified to influence responses to seven items and is defined as athletes' evaluations of their head coach's ability to lead during competition. Technique competence (TC) is specified to influence responses to six items and is defined as athletes' evaluations of their head coach's instructional and diagnostic abilities. Character building competence (CBC) is specified to influence responses to four items and is defined as athletes' evaluations of their head coach's ability to influence the personal development and positive attitude toward sport in their athletes. In the unidimensional model, total coaching competence (TCC) is specified to influence responses to all 24 items and is defined as athletes' evaluations of their head coach's ability to affect the learning and performance of athletes. To determine the fit of these internal models the nesting of athletes within head coaches need be taken into account. Multilevel confirmatory factor analysis (MCFA) is an appropriate methodology when data are meaningfully nested and an evaluation of the factor structure of an instrument is desired (Muthén, 1989, 1994).

Figure 2. Multidimensional model of the CCS.



### *Statement of Purpose*

The purposes of this study were to (a) provide initial evidence for key aspects of the validity framework described, and (b) introduce MCFA as an appropriate methodology under the previously stated circumstances.

### *Research Questions*

This study provided initial evidence for key aspects of the validity framework described by examining the following questions:

1. Did athletes employ the rating scale structure in the manner that the authors intended?
2. To what degree did various multilevel internal models fit the data?
3. How reliable were the rank orderings of coaching competency estimates?
4. Were coaching competency estimates positively related to satisfaction with the coach within teams?
5. Were coaching competency estimates positively related to satisfaction with the coach between teams?

Additionally, in Chapter 2, this study provides a modest primer of applying MCFA methodology when athletes are nested in teams.

## CHAPTER 2

### INTRODUCTION OF MCFA

This introduction should not be construed as original thinking by the author. Rather, it is based on the author's understanding of Muthén's pioneering work (1989; 1994) and previous syntheses of MCFA in the education literature (Kaplan, 2000; Hox, 2002). The rationale for providing it is to put a relatively new and complex methodology which may have important applications in sport and exercise science, in a context that may be more familiar and accessible to this audience. A working understanding of confirmatory factor analysis (CFA) is assumed.

The call to separate substantive levels of variance, such as within students and between classrooms, has been voiced for decades in education (Cronbach, 1976; Harnqvist, 1978). Empirically, nonhierarchical analysis of the total covariance matrix of meaningfully nested data violates the assumption of independence. Conceptually and practically, nonhierarchical analyses often confound within-group and between-group relations and hamper theory development at both levels. Whether nesting is due to students clustered within classrooms, or athletes within teams, if both individual attributes and group characteristics are relevant then multilevel modeling, also known as hierarchical modeling, should be considered.

Most multilevel modeling applications in social science research are multilevel extensions of the conventional multiple regression model (Hox, 2002).<sup>2</sup> But, there are models of substantive interest, MCFA in this study, that contain multiple levels which cannot be analyzed within the linear multilevel framework. Muthén (1989) and Muthén and Satorra (1989) have provided an approach to latent variable modeling when data are

meaningfully nested. This section provides an example of applying this methodology when athletes are nested within teams and evaluation of the factor structure of an instrument is desired.<sup>3</sup> However, before introducing the complex methodology, it is worthwhile to explain typical and practical consequences of ignoring multilevel data structures when evaluating the factor structure of latent variables.

### *Consequences of Ignoring Multilevel Data Structures*

In a simulation study, Julian (2001) generated multilevel data, manipulating three design factors: (a) intraclass correlation (ICC): .05, .15, and .45; (b) group and member configuration: 100 groups with five members each (100/5), 50/10, 25/20, and 10/50; and (c) the internal model of the between group variance components: four factor oblique within and four factor oblique between (4FOBW/4FOBB), 4FOBW/2FOBB, and 4FOBW/5FOBB. Data were fitted to CFA models that ignored the multilevel structure. Biases in the chi-square statistic ( $\chi^2$ ), model parameters (i.e., factor loadings, variances, covariances, and error variances), and standard errors were examined. The design factor that exerted the most influence was the ICC. When the ICC was  $> .05$  the  $\chi^2$  was inflated, model parameters were inflated, and standard errors were deflated.

Practically, CFA on the total covariance matrix ( $S_T$ ) when the ICCs are non-trivial ( $\geq .10$ , as a guideline; Muthen, 1997) will likely result in the model exhibiting more overall misfit (i.e., elevated  $\chi^2$  test), hypothesis tests that are overly optimistic (i.e., deflated standard errors leading to increased Type I error), and inflation of the absolute value of parameter estimates (e.g., factor loadings). A correction by Satorra and Bentler (1990, 1994) which is often used to correct for such bias in the  $\chi^2$  statistic and its standard

errors does not adjust for bias in the absolute value of the factor loadings (Julian, 2001), and does not allow modeling at both levels.

### *Steps in a MCFA*

Muthén (1994) advised that MCFA should generally follow four steps. First, factor analyses of  $S_T$  are conducted, exploratory (EFA) and/or confirmatory depending on the adequacy of an a priori internal model. Although this solution will likely be biased if the data truly exhibit a multilevel structure (Julian, 2001; Hox & Maas, 2001), it can provide a rough estimate of model fit. Second, estimate the degree of between group variation in the variables of interest (i.e., the ICC). The ICC for each item is estimated,

where  $ICC = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2}$ ,  $\sigma_B^2$  is the between group variance, and  $\sigma_W^2$  is the within

group variance. If the ICCs are trivial a multilevel model may be unnecessary. Third, estimate the within factor structure of the within group covariance matrix ( $S_{PW}$ ). Because  $S_{PW}$  is an unbiased estimator of the within group population covariance matrix ( $\Sigma_W$ ) and is not confounded by the scaled between group covariance ( $S_B^*$ ), model fit should be at least as good as in Step 1. Fourth, estimate the between factor structure. Because  $S_B^*$  is not an unbiased estimator of the between group population covariance matrix ( $\Sigma_B$ ) and includes  $\Sigma_W$ , this is the most difficult part of the analysis and requires a technical synopsis.

### *A Technical Synopsis*

Assume that athletes are nested within teams, and that the number of athletes within teams is unbalanced. To perform a MCFA athletes' competency scores need to be broken down into an athlete-level component (i.e., athlete's deviation from the team

mean:  $Y_{gi} - \bar{Y}_g$ ) and a team-level component (the disaggregated team mean:  $\bar{Y}_g$ ). An unbiased estimate of the population within team covariance matrix ( $\Sigma_w$ ) is provided by

$$S_{PW}, \text{ where } S_{PW} = \frac{\sum_g \sum_i^{N_g} (Y_{gi} - \bar{Y}_g)(Y_{gi} - \bar{Y}_g)'}{N - G} \text{ and subscript } g = 1 \dots G \text{ teams, and}$$

subscript  $i = 1 \dots N_g$  athletes per group with a total of  $N$  athletes (Muthén, 1989). Because  $S_{PW}$  is the maximum likelihood estimator of  $\Sigma_w$  with sample size  $N - G$ , Step 3 is reasonable (Muthén, 1990).

$$\text{In the unbalanced case, } S_B^* \text{ is calculated where } S_B^* = \frac{\sum_g n (\bar{Y} - \bar{Y}_g)(\bar{Y} - \bar{Y}_g)'}{G - 1}.$$

$S_B^*$  estimates the composite  $\Sigma_w + c^* \Sigma_B$  where  $c^*$  is a scaling parameter. Full Maximum Likelihood (FIML) estimation of  $\Sigma_B$  is problematic because a  $S_B^*$  would need to be computed for each distinct group size (Muthén, 1990). Muthén suggests computing a

$$\text{single } S_B^* \text{ by including } c^*, \text{ where } c^* = \frac{N^2 - \sum_g n_g^2}{N(G - 1)}.$$

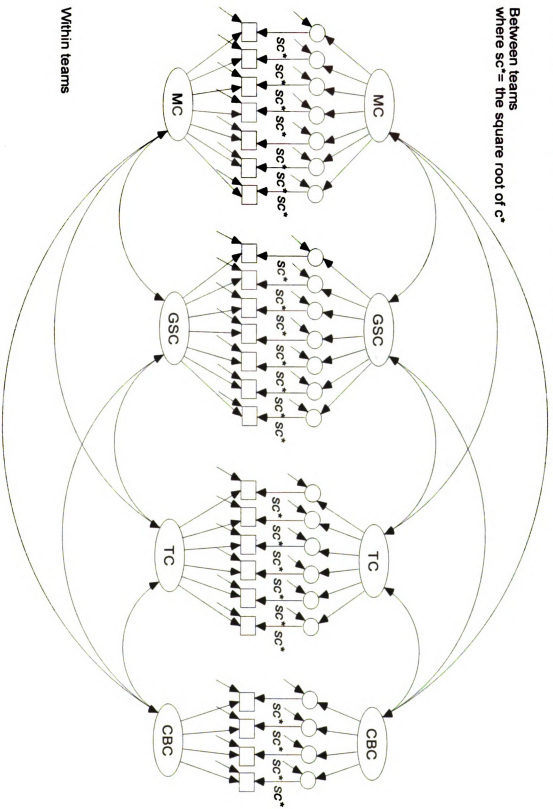
The value of  $c^*$  is approximately

equal to the average team size. This estimation procedure is referred to as Muthén's maximum likelihood (MUML). Simulation studies and comparisons of FIML and MUML estimates suggests that Muthén's methodology provides a reasonable approximation of FIML estimates in many applications (Hox, 1993; Hox & Maas, 2001; McDonald, 1994, Muthén, 1990).

Once the appropriate covariance matrices are estimated, the within and between factor structures can be evaluated simultaneously via the multi-group option found in most structural equation modeling software (Hox, 2002). However, because  $S^*_B$  estimates the composite  $\Sigma_W + c^*\Sigma_B$ , two models need be specified for  $S^*_B$ : one for the within team structure and one for the between team structure. Thus, the within team structure is specified at both levels with equality restrictions between both “groups”. The between team structure is specified only in the second “group” with the square root of the scaling factor,  $sc^*$ , built into the model. See Figure 3 for an example where the model specified in Figure 2 is specified at both the within and between team levels.

Figure 3. Multidimensional MUML model of CCS

Between teams  
where  $sc^* =$  the square root of  $c^*$



Within teams

*Sample size.* Hox and Maas (2001) generated multilevel data to explore the accuracy of Muthén's methodology with pseudobalanced groups and small samples at both levels. They concluded that the within teams part of the model performs well even with small samples at both levels (i.e.,  $n = 5$  to 15 within "small" groups, and  $G \leq 50$ , as guidelines). They also concluded that problems can arise when estimating the between teams part of the model when the number of teams is small. In such instances inadmissible estimates, such as negative error variances, may be observed at the between teams level. In such cases the error variance is fixed to zero. But, even when an admissible solution is reached, residual error variances and standard errors may still be underestimated. However, factor loadings are generally accurate. Due to the said biases, Type I error rates may be approximately 8%, when an alpha of .05 is specified. Thus, ideally, within teams sample sizes should  $\geq 5$ , and team-level sample size should be  $\geq 50$ . In cases where resources limit the number of teams but MCFA is still deemed appropriate (i.e., modeling the between teams variance is important), as in this study, it is necessary to qualify inferences made from the between teams portion of the model.

*Judging the fit of MCFA.* As in CFA, there is no shortage of indices to judge the fit of a MCFA to the observed data. Given common practice in sport and exercise science and guidelines by Hu and Bentler (1999) and Kline (1998), the following fit indices are suggested: the  $\chi^2$  test, Bentler's comparative fit index (CFI; 1990), the Tucker-Lewis index (TLI; Tucker & Lewis, 1973), the standardized root mean squared residual (SRMR), and the root mean square error of approximation (RMSEA; Browne & Cudek, 1992). In instances where comparing the fit of competing models is desired, the

likelihood ratio chi-square statistic ( $\chi^2_{LR}$ ) and the consistent Akaike information criterion (CAIC; Bozdogan, 1987) are also suggested.

Introductions to and criterion guidelines for the suggested fit indices are offered here guided by recommendations from Hu & Bentler (1999) and Kline (1998). The  $\chi^2$  test represents a test of the significance of the difference in fit between the specified model and a just-identified version of it. Because this test is sensitive to sample size a common guideline is to divide the  $\chi^2$  by the degrees of freedom ( $df$ ), where a ratio  $\leq 3$  suggests reasonable fit. The CFI and TLI are incremental fit indices which indicate the proportion in the improvement of the overall fit of the specified model to a null model. Well-fitting models have values  $\geq .95$ , while marginally acceptable models have values  $\geq .90$ . The SRMR denotes the average residual from fitting the correlation matrix for the specified model to the correlation matrix of the observed data. Well-fitting models have a value  $\leq .08$ , while marginally acceptable models have a value  $\leq .10$ . The RMSEA represents the amount of error per  $df$  in the specified model. Well-fitting models have a value  $\leq .06$ , while marginally acceptable models have a value  $\leq .10$ .

The  $\chi^2_{LR}$  statistic can evaluate the relative fit of nested models ( $\chi^2_{LR} = G^2_{simple} - G^2_{complex}$ ) where  $G^2$  is the deviance value for the model in question. The  $\chi^2_{LR}$  statistic is distributed with degrees of freedom equal to the difference between the number of parameters in the nested models (McCullagh & Nelder, 1990). Because the  $\chi^2_{LR}$  statistic is sensitive to sample size, the CAIC is also suggested. The CAIC depicts the fit of the model in question relative to the number of parameters estimated, where lower values indicate better fit and competing model do not need to be nested (Wicherts & Dolan, 2004).

### *An Application of Muthén's Steps*

Assume that the internal model in Figure 3 is posited. It bears mentioning that the internal models need not be the same at both levels. Step 1 and Step 2 can be performed on the raw data matrix. Assume that a CFA on  $S_T$  in Step 1 provided some support for the posited internal model in Figure 2 but with more misfit than is acceptable. Assume further that Step 2 revealed ICCs for all of the 24 items that were considerably greater than .10 and that MCFA was deemed appropriate. Practically, this decision implies that athletes' perceptions of their head coach's coaching competency were influenced by athlete-level attributes as well as shared team-level characteristics. Given the construct of interest, athletes' evaluations of their head coach's coaching competencies, strong within teams and between teams effects were expected.

Step 3 and Step 4 are multi-staged steps. Step 3 begins with estimating both  $S_{PW}$  and  $S_B^*$  from the raw data matrix,  $S_T$ . All MCFA analyses that will be reviewed can, and later will be, performed in *EQS 6.1* (Bentler, 2004). The covariances in  $S_B^*$  should be larger than the values in  $S_{PW}$  because this matrix equals the within team matrix plus the between team matrix multiplied by  $c^*$ .  $S_{PW}$  is fitted to the hypothesized model, Figure 2 in this case, ignoring  $S_B^*$ . Justifiable modifications to the internal model should then be made, if necessary.

The final step, evaluating the fit of the within teams and between teams models simultaneously, also can be considered multi-staged (Hox, 2002). Using the multi-group procedure, the final within teams model is specified in both "groups" with equality constraints across groups for all parameter estimates. This part of the model is constant throughout the proceeding analyses. A series of between teams models should then be

specified. A null between teams model (i.e., the lack of any between team model) should be specified. If this model fits, there is no between teams model. Assuming poor fit, an independent model (i.e., only variances are estimated in the between model) should be specified. If this model fits, team-level variance exists but a team-level structural model does not. Assuming improved yet still unacceptable fit, the hypothesized between model should be specified. Justifiable modifications to the internal model should then be made, if necessary.

## CHAPTER 3

### METHOD

#### *Sample*

To maximize group-level sample size, teams were recruited from both lower division intercollegiate soccer and ice hockey programs. Soccer and ice hockey were chosen because both can be considered open team sports, because head coaches tend to be involved in the coaching of most, if not all, positions on the team, and because both have a fairly large number of athletes within teams. Men's ( $g = 14$ ) and women's ( $g = 29$ ) soccer teams ( $g = 43$ ) were recruited from the midwest of the United States, and 31 teams (12 men's and 19 women's) agreed to participate. Women's ice hockey teams ( $g = 28$ ) were recruited from the northeast and midwest regions of the United States, and 16 teams agreed to participate. Despite numerous reminders from a variety of sources, only 21 soccer teams (8 men's, 13 women's) and 11 hockey teams submitted data (response rate = 68%;  $G = 32$ ). Within the soccer sample, three coaches of the women's teams were women; men coached the men's teams. Within the women's ice hockey sample, five of the coaches were women.

At the athlete-level, participants were 407 soccer players (165 male, 242 female) and 183 ice hockey players ( $M = 18.44$  athletes per team). Within teams, number of participants varied from 13 to 25 ( $N = 590$ ). Across teams, most participants were Caucasian (94%) and between the ages of 18 – 23 years (99%;  $M = 19.53$ ,  $SD = 1.34$ ). Distribution of year on team was 43% first-year, 23% second-year, 20% third-year, and 13% fourth-year.

#### *Procedure*

All necessary permissions were obtained from the institutional review board and the 32 head coaches prior to data collection. An explanation of the study was presented to each team by the head coach. Informed consent was obtained from all athletes. Athletes were guaranteed confidentiality for their responses. Questionnaires were completed at approximately the one-half mark of the season to ensure that athletes had enough experience to make informed judgments regarding their head coach's coaching competency and their own satisfaction with the coach. On each team, an identified trainer or team manager administered the questionnaires. Completed questionnaires were returned to the trainer or team manager, who mailed the returns to the researchers. Trainers or team managers who successfully followed through were given a \$60 honorarium.

### *Measures*

*Coaching competence.* Coaching competence was measured at the athlete-level by the CCS as described in the initial validity framework subsection of the Introduction. The CCS items are listed in Appendix A.

*Satisfaction with the coach.* Satisfaction with the coach was measured at the athlete-level and consisted of selected items from a scale that was intended to measure, in part, attitudes toward the head coach (Smith, Smoll, & Curtis, 1978).<sup>4</sup> Indicators used in this study were the same as those used by Feltz et al. (1999) and Myers, Vargas-Tonsing et al. (in press) and included (a) how much do you like playing for your coach, (b) if you were able to play next year how much would you like to have the same coach again, (c) how much does your coach like you and, (d) how much does your coach know about soccer/hockey. Ratings were made on a 7-point Likert scale ranging from 1 (*very little*) to

7 (*a lot*). Myers, Vargas-Tonsing et al. provided psychometric evidence for the unidimensionality of the scale for a similar purpose. Specifically, in their study, the principal component accounted for 70% of the total variance, all items had a high loading ( $\geq .70$ ) on this component, and the set of items displayed good internal consistency,  $\alpha = .85$ .

### *Treatment of Data*

*Missing data.* Five cases were missing data (i.e., < 1% of all cases). In each case data were missing for no more than two responses. Thus, incomplete cases had at least 93% of the responses of primary interest. These cases were retained and data were judged to be randomly missing. Scores for missing responses were imputed using observed responses based on case-wise maximum likelihood estimation using the Jamshidian-Bentler EM algorithm (1999).

*Outliers.* Multivariate outliers were identified using Mahalanobis distance estimates. Five cases, or less than 1% of the data, were identified ( $p < .001$ ). As suggested by Tabachnick and Fidell (2001) visual inspection of the cases and stepwise regression ensued to inform judgments regarding what caused the cases to be outliers prior to deciding their fate. In the identified cases, data were entered correctly, subjects were within expected age ranges, represented both genders, and were nested within different teams. Empirically, six to eight competency items distinguished each outlier from the other cases. The specific items that generated outlying responses were not consistent across the outliers. The five outlying cases were determined to be random and were dropped. The final athlete-level sample size was,  $N = 585$ .

*Normality.* Normality was assessed with univariate skew and univariate kurtosis estimates and Mardia's (1970) coefficient. Due to the size of the sample, absolute values of univariate estimates, not significance tests, were examined. Guidelines used were values  $\geq |3|$  for extreme skewness and values  $\geq |10|$  for extreme kurtosis (Kline, 1998). None of the univariate estimates suggested extreme non-normality (see Table 2). Mardia's coefficient suggested multivariate departures from the kurtosis of a normal distribution, 120.94,  $p < .001$ . However, large samples can inflate values for this coefficient (Bollen, 1989). A common adjustment for multivariate kurtosis, Satorra and Bentler's (1994) correction, was not applied because it applies the correction on the raw data matrix and most analyses in MUML methodology use covariance matrices. However, because nonnormal distributions can inflate test statistics (Muthén & Kaplan, 1985), subsequent model fit may have been artificially inflated (i.e., worse fit).

### *Analyses*

*Rating scale.* Competency data were calibrated to the Rasch Rating Scale Model (RSM; Andrich, 1978) using *Winsteps* (Wright & Linacre, 1998). Rasch models are a family of 1-parameter item response theory (IRT) measurement models. IRT is an alternative to true score test (TST) theory and is well suited to analyze rating scale data (Wright & Masters, 1982). In this case, the RSM described the probability that a specific athlete ( $n$ ) would rate a particular item ( $i$ ) using a specific rating scale category ( $k$ ), conditioned on the athlete's competency judgment ( $\theta_n$ ) and the item's difficulty ( $\delta_i$ ). The log-odds equation for this probability,  $\log(P_k / P_{k-1}) = \theta_n - \delta_i - \tau_k$ , contains three parameters:  $\theta_n$ ,  $\delta_i$ , and category threshold ( $\tau_k$ ), the threshold between two adjacent rating scale categories. Parameters of this model are estimated from the observed data via a

Joint Maximum Likelihood Estimation method, which does not impose distributional assumptions on the parameter estimates (Wright & Masters, 1982). In addition to the parameter estimates, *Winsteps* produces standard errors of these estimates and model-to-data fit indices. In accordance with posited internal models, both an omnibus unidimensional construct (TCC) and consecutive unidimensional dimensions (MC, GSC, TC, CBC) were explored.<sup>5</sup> The consecutive approach was just a unidimensional approach repeated for each dimension.

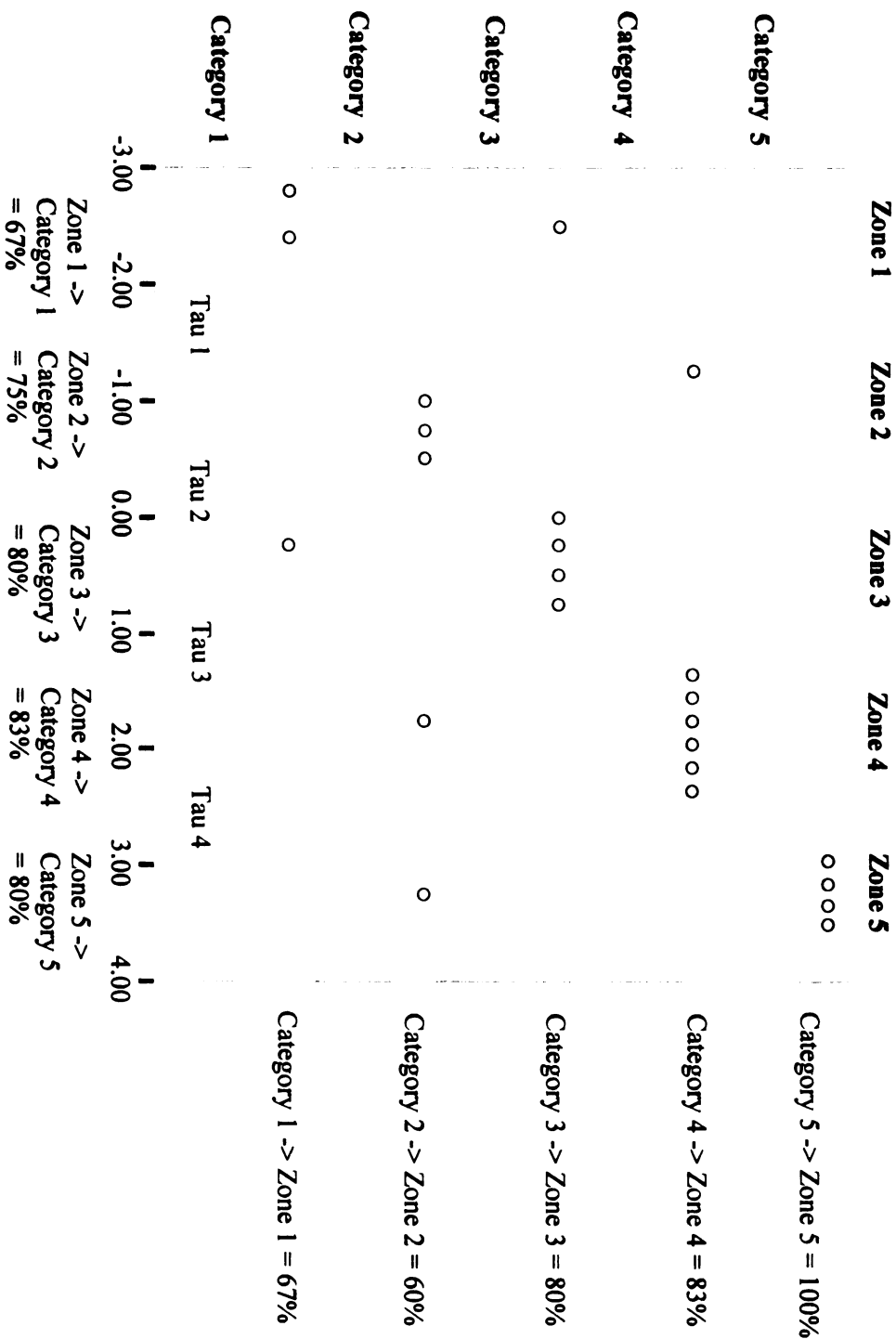
Following calibration of the data to the RSM, the degree to which athletes employed the rating scale structure in the manner that the authors intended was evaluated according to guidelines suggested by Linacre (2002). These guidelines can be summarized as: (a) all categories should have at least 10 observations, (b) distributions of ratings for each category should be unimodal, (c) average measures should increase with the categories, (d) unweighted mean square (UMS) fit statistics should be less than 2.0 for each threshold, (e) category thresholds should increase with the categories, (f) ratings imply measures (coherence > 39%), (g) measures imply ratings (coherence > 39%), (h) category thresholds should increase by at least 1.2 logits, and, (i) category thresholds should increase by no more than 5 logits.

Because criteria (d), (f) and (g) are not well-defined in the sport and exercise science literature, an elaboration of these guidelines is provided. The UMS fit statistic depicts the degree to which the observed ratings are consistent with the expected values, and is sensitive to large residuals from pairings of item difficulty and ability estimates that are far apart on the underlying scale. The UMS fit statistic is reported as a chi-square divided by its degrees of freedom, resulting in an expected value of 1.00 and a range from

0.00 to  $\infty$ . In rating scale analyses, thresholds with UMS fit statistics  $< 2.00$  are considered to demonstrate adequate fit (Linacre, 2002).

*Coherence* refers to the degree to which the observed ratings match the modeled expectations for a particular rating scale category, and vice versa. Each rating can be depicted as belonging to a rating scale category. The logit scale upon which athletes are located can be demarked into measurement zones that are defined by the locations of the category thresholds. As illustrated in Figure 4, the location of each athlete can be associated with one measurement zone (X axis—one zone per rating scale category with a zone representing the expected rating for that coach-by-item combination), and the observed ratings assigned by that athlete (X) can be placed in one of the rating scale categories (located in the Y axis). Each point indicates a rating assigned by an athlete to a single item. Note that two of the three ratings in Zone 1 are in the observed rating category of 1 and one of these ratings is in the rating category of 3. Hence, 67% of the measures in Zone 1 are coherent with Category 1 observations. Similarly, three of the five Category 2 observations fall in Zone 2. Hence, 60% of Category 2 ratings are coherent with Zone 2.

Figure 4. Coherence example for a well-fitting five category structure.



Because a number of Linacre's guidelines (2002) were not realized in the original rating scale structure (to be discussed in the Results section), a post-hoc approach was applied to arrive at an improved rating scale structure. To determine a post-hoc structure, categories were collapsed based on general principles (Linacre, 1995; Wright & Linacre, 1992) and statistical indicators (Zhu et al., 1997). General principles for collapsing categories state that collapsed categories (a) should be explainable, and, (b) should balance observed frequencies as much as possible. Statistical indicators of improved fit for a post-hoc structure should include (a) improved model-data fit statistics, (b) category and parameter estimates that come closer to satisfying Linacre's guidelines, and, (c) separation indices should not decrease drastically as compared to the original rating scale structure.<sup>6</sup> Of interest in Rasch analysis are the person and item separation indices. In non-technical terms (see Linacre, 1994, for a technical introduction), both of these indices indicate how well the scale distinguishes individual people and items, respectively, where larger values indicate greater separation.

*Internal models.* A MCFA approach to model fit was used to evaluate the utility of a unidimensional model (TCC) and a multidimensional model both within teams and between teams (see Figure 3). Post hoc respecification was considered only for models that approached acceptable fit (MacCallum, Roznowski, & Necowitz, 1992). Empirically-based possibilities for model respecification were guided by inspection of Lagrange multiplier tests [LM] (Silvey, 1959), Wald tests (Wald, 1943), and standardized residuals. LM values approximate the amount by which the model's overall  $\chi^2$  would decrease if the identified parameter were estimated. Wald test values estimate the amount by which the model's overall  $\chi^2$  would increase if the identified parameter were fixed to

zero. Standardized residuals indicate the degree of misfit between the observed correlation matrix and the predicted correlation matrix, where absolute values greater than .10 can indicate misfit (Kline, 1999). Empirically-based possibilities were evaluated, in part, based on relevant coaching effectiveness literature. Disattenuated correlations among latent factors were also examined to depict the degree of redundancy in the multidimensional models.

Once an internal model was determined, variability of the factor structure between sub-samples was evaluated to determine the degree to which the assumption that all of the athletes were a random sample of a single population was reasonable. Sub-samples of note were males ( $n = 165$ ) and females ( $n = 420$ ), soccer athletes ( $n = 403$ ) and women's hockey athletes ( $n = 182$ ), and year on team, where 1<sup>st</sup> = 150, 2<sup>nd</sup> = 135, and 3<sup>rd</sup> and 4<sup>th</sup> = 190. Athletes in their third and fourth year were collapsed so that each sub-sample had at least 100 observations. Because determining the degree of factorial invariance between sub-samples was not the purpose of this study, and because of the relatively large size of the sub-samples, only the invariance of factor loadings and factor covariances were explored, and the alpha level selected for these comparisons was equal to .001.

*Internal consistency reliability.* The consistency of rank orderings of competence estimates across measurement contexts was examined with reliability of separation coefficients ( $\alpha$ ). The reliability of separation coefficient is analogous to Cronbach's (1951) alpha, but it is based on estimates of true and error variance derived from IRT models. Specifically, the reliability of separation for competency estimates =  $[V(\hat{\theta}) - MSE(\hat{\theta})] / V(\hat{\theta})$ , where  $V(\hat{\theta})$  is the variance of the competency estimates and  $MSE(\hat{\theta})$  is the mean error variance of the competency estimates. This equation is comparable to

the TST theory definition of reliability as the ratio of true variance to observed variance. As suggested by Kline (1999) and in relation to the given purpose,  $\alpha$ s greater than 0.90 were considered excellent,  $\alpha$ s greater than 0.80 were considered very good, and  $\alpha$ s greater than 0.70 were considered adequate.

*Forming measures to test questions 4 and 5.* Satisfaction with the coach data were calibrated to the RSM using *Winsteps*. In this case, the RSM described the probability that a specific athlete ( $n$ ) would rate a particular item ( $i$ ) using a specific rating scale category ( $k$ ), conditioned on the athlete's satisfaction with the coach ( $\theta_n$ ) and on the item's difficulty ( $\delta_i$ ). The log-odds equation for this probability,  $\log (P_k / P_{k-1}) = \theta_n - \delta_i - \tau_k$ , contains three parameters:  $\theta_n$ ,  $\delta_i$ , and category threshold ( $\tau_k$ ). Calibration of the data to the RSM resulted in a satisfaction estimate for each athlete. Estimates were on a single linear continuum in logistic ratio units (logits). A logit is the natural logarithm of the odds of an event. Because the data in this study were polytomous, odds were defined by the likelihood of assigning a rating in one category versus the odds of assigning a rating in the next lower category.

A Principal Component Analysis (PCA) of the residuals from the Rasch model was performed on the four indicators of satisfaction to judge the adequacy of the assumption of unidimensionality. The residual from the Rasch model is defined as the difference between the observed rating and the model-based expectation, where the expected score ( $E$ ) for the  $n^{\text{th}}$  athlete on the  $i^{\text{th}}$  item was given by (Linacre, 1998):

$$E_{ni} = \sum_{k=0}^m kP_{nik} , \text{ where } k \text{ is the value of the rating scale category, ranging from 0 to a}$$

maximum number,  $m$ , and  $P_{nik}$  is the probability of observing a response in category  $k$  for

athlete  $n$  on item  $i$  as defined by the Rasch model. Scaling data to the Rasch model is equivalent to extracting the first principal component, which is referred to as the Rasch component, from the data with the restriction that all items have equal loadings on that component (McDonald, 1985). A PCA of the residuals reveals whether systematicity (i.e., multidimensionality) exists once the variance accounted for by the Rasch component was taken into account. That is, the results of the PCA indicate whether the assumption of unidimensionality is tenable.

Competency data also were calibrated to the RSM using *Winsteps*, although initially these data were attempted to be fitted to a multidimensional extension of this model as implemented in *ConQuest* (Wu, Adams, & Wilson, 1998). But, due to a non-positive definite psi matrix (i.e., high correlations among the factors, which will be presented in the Results section), convergence to a stable solution for the multidimensional model was not possible. Empirically, what this means is that at least some of the factors could not be distinguished by the subjects within this particular measurement model. Instead, four unidimensional models were fit to the data, to produce logit-based scores, guided by the final internal model (i.e., MC, GSC, TC, CBC). The decision to maintain separate factors was consistent with the results of the MCFA and will be defended in the Discussion section. PCA of the residuals from these models were not performed because the dimensionality of these data was confirmed in the MCFA.

*Testing questions 4 and 5.* Satisfaction and coaching competency measures were dependent because athlete observations were nested within teams. Hierarchical linear modeling (HLM), as implemented in *HLM5* (Raudenbush, Bryk, Cheong, & Congdon, 2000), is well-suited to handle observed dependent data. Congruent with Horn's model of

coaching effectiveness (see Figure 1), satisfaction with the coach was treated as the dependant variable and the proposed positive relationships with coaching competency were tested within and between teams. To avoid problems caused by multicollinearity, bivariate correlations between coaching competencies and satisfaction with the coach, at the individual-level and team-level, were explored to determine which competency measure would be used as the independent variable.

Model-building consisted of at least three steps to test questions 4 and 5 (Raudenbush & Bryk, 2002). First, an unconditional model (i.e., Model 1) was imposed:

*Level 1:*  $Y_{ig} = \beta_{0g} + r_{ig}$ , where

$\beta_{0g}$  = the mean of athlete satisfaction for team  $g$ ,

$r_{ig}$  = the unique effect of athlete  $i$  for team  $g$ ,

*Level 2:*  $\beta_{0g} = \gamma_{00} + u_{0g}$ , where

$\gamma_{00}$  = the average team mean of athlete satisfaction,

$u_{0g}$  = the unique effect of team  $g$  on the average team satisfaction mean.

Of particular interest in Model 1 was the variance of  $r_{ig}$  or the within team variance,  $\sigma^2_w$ , and the variance of  $u_{0g}$  or the between team variance,  $\sigma^2_B$ . These variances were used to estimate the ICC of satisfaction with the coach.

Second, a random coefficient regression model (i.e., Model 2) was imposed where the coaching competency slope was estimated within teams and was free to vary between teams:

*Level 1:*  $Y_{ig} = \beta_{0g} + \beta_{1g}(\text{athlete's competency judgment})_{ig} + r_{ig}$ , where

$\beta_{0g}$  = the mean of athlete satisfaction for team  $g$ ,

$\beta_{1g}$  = the expected amount that an athlete's satisfaction score would change given a one-unit change in his/her competency judgment for team  $g$ ,

$r_{ig}$  = the residual of athlete  $i$  for team  $g$ ,

*Level 2:*  $\beta_{0g} = \gamma_{00} + u_{0g}$

$\beta_{1g} = \gamma_{10} + u_{1g}$ , where

$\gamma_{00}$  = the average team mean of athlete satisfaction,

$\gamma_{10}$  = the average satisfaction-competency slope across teams,

$u_{0g}$  = the unique effect of team  $g$  on the average team satisfaction mean,

$u_{1g}$  = the unique effect of team  $g$  on the average satisfaction-competency slope.

Of particular interest in Model 2 was  $\gamma_{10}$  and the variance of  $u_{1g}$ ,  $\sigma^2_{B1}$ .  $\gamma_{10}$  was the average satisfaction-competency slope and addressed question 4.  $\sigma^2_{B1}$  was the variance of the estimated satisfaction-competency slopes,  $\beta_{1g}$ s, around the average satisfaction-competency slope,  $\gamma_{10}$ . An alpha equal to .05 was selected for all hypotheses tests, and the magnitude of the standardized betas were interpreted according to Cohen's (1988) guidelines for effect sizes, where 0.20, 0.50, and 0.80 indicated small, medium, and large effect sizes, respectively.

Third, an intercepts as outcomes model (i.e., Model 3) was imposed where the team competency score was added to model between team variance on  $\gamma_{00}$ :

*Level 1:*  $Y_{ig} = \beta_{0g} + \beta_{1g}(\text{athlete's competency judgment})_{ig} + r_{ig}$ , where

$\beta_{0g}$  = the mean of athlete satisfaction for team  $g$ ,

$\beta_{1g}$  = the expected amount that an athlete's satisfaction score would change given a one-unit change in his/her competency judgment for team  $g$ ,

$r_{ig}$  = the residual of athlete  $i$  for team  $g$ ,

*Level 2:*  $\beta_{0g} = \gamma_{00} + \gamma_{01}(\text{team's competency judgment})_g + u_{0g}$

$\beta_{1g} = \gamma_{10} + u_{1g}$  where

$\gamma_{00}$  = the average team mean of athlete satisfaction,

$\gamma_{01}$  = the expected amount that team satisfaction would change given a one-unit change in the team competency score,

$\gamma_{10}$  = the average satisfaction-competency slope across teams,

$u_{0g}$  = the residual of team  $g$  on the average team satisfaction mean,

$u_{1g}$  = the unique effect of team  $g$  on the average satisfaction-competency slope.

Of particular interest in Model 3 was  $\gamma_{01}$  and the variance of  $u_{0g}$ ,  $\sigma^2_{B_0}$ .  $\gamma_{01}$  was the average satisfaction-competency slope across teams and addressed question 5.  $\sigma^2_{B_1}$  was the variance of the adjusted average team satisfaction scores. Sport played (0 = soccer, 1 = ice hockey) was entered as a Level-2 predictor to ensure that none of the fixed effects varied on the basis of sport played.

*Model estimation and fit.* Final parameters were estimated via restricted maximum likelihood, and differences in model fit were examined via FIML estimation (Raudenbush & Bryk, 2002). Relative fit of nested models was evaluated with a  $\chi^2_{LR}$  statistic and comparing CAIC values.

*Reliability estimates.* Point reliability estimates described how reliable, on average, the slopes were based on computing ordinary least square regressions separately for each team. A reliability estimate was provided by averaging individual team estimates. Raudenbush and Bryk (2002) suggest, as a guideline, that the point reliability estimate should be greater than .05. Slopes that do not meet this heuristic were candidates to be fixed across groups.

## CHAPTER 4

### RESULTS

#### *Did Athletes Employ the Rating Scale Structure in the Manner that the Authors Intended?*

Athletes did not employ the original rating scale structure, for either the omnibus unidimensional conceptualization or the consecutive unidimensional conceptualization, in the manner that the authors intended. Specific problems were observed with threshold estimates-criteria (e), (h), and (i), and coherences-criteria (f) and (g) (see Table 1). In the omnibus unidimensional model, all threshold estimates and most (65%) coherences failed to fully meet these criteria. In the consecutive unidimensional model, most (60%) thresholds and more than one-third (38%) of coherences failed to fully meet these criteria. In both conceptualizations, the lower portion of the scale was used less frequently than the middle and upper end of scale, and the majority of problems were observed in the lower and middle ranges with some problems associated with Category 9.

Post-hoc categorizations were evaluated in accordance with previously mentioned criteria. The best fitting, and accepted, post hoc categorization was a five-category structure where responses were collapsed in Categories 0 and 1, Categories 2 through 4, Categories 5 and 6, Category 7, and Categories 8 and 9 (see Table 1). The majority of problems, according to Linacre's (2002) guidelines, encountered in the omnibus unidimensional conceptualization were improved in the post hoc categorization. All of the problems, according to Linacre's guidelines, observed in the consecutive unidimensional conceptualization were improved in the post hoc categorization. Last, person and item and separation statistics changed little in the new categorization, which was consistent with Zhu et al.'s (1997) guidelines for post hoc categorization.

**Table 1**  
*Original and Post Hoc Rating Scale Structures*

|                               |   | Total coaching competence |          | Motivation competence |          | Game strategy competence |          | Technique competence |          | Character building competence |          |
|-------------------------------|---|---------------------------|----------|-----------------------|----------|--------------------------|----------|----------------------|----------|-------------------------------|----------|
|                               |   | Original                  | Post-hoc | Original              | Post-hoc | Original                 | Post-hoc | Original             | Post-hoc | Original                      | Post-hoc |
| (a) & (b)<br>Observations     | 0 | 176                       | 309      | 65                    | 121      | 42                       | 61       | 40                   | 69       | 29                            | 58       |
|                               | 1 | 133                       |          | 56                    |          | 19                       |          | 29                   |          | 29                            |          |
|                               | 2 | 198                       |          | 86                    |          | 37                       |          | 43                   |          | 32                            |          |
|                               | 3 | 441                       | 1299     | 189                   | 532      | 83                       | 257      | 100                  | 307      | 69                            | 203      |
|                               | 4 | 660                       |          | 257                   |          | 137                      |          | 164                  |          | 102                           |          |
|                               | 5 | 1281                      | 3760     | 444                   | 1284     | 319                      | 987      | 344                  | 994      | 174                           | 494      |
|                               | 6 | 2479                      |          | 841                   |          | 668                      |          | 650                  |          | 320                           |          |
|                               | 7 | 3665                      | 3665     | 1042                  | 1042     | 1108                     | 1108     | 929                  | 929      | 586                           | 586      |
|                               | 8 | 3118                      |          | 707                   |          | 1114                     |          | 747                  |          | 550                           |          |
|                               | 9 | 1888                      | 5006     | 409                   | 1116     | 568                      | 1682     | 464                  | 1211     | 448                           | 998      |
| (c)<br>Average measure        | 0 | -2.48                     | -3.55    | -3.59                 | -5.26    | -4.26                    | -6.01    | -3.05                | -4.21    | -3.90                         | -4.40    |
|                               | 1 | -1.55                     |          | -2.52                 |          | -3.47                    |          | -2.19                |          | -2.54                         |          |
|                               | 2 | -0.75                     |          | -1.55                 |          | -2.27                    |          | -1.15                |          | -1.62                         |          |
|                               | 3 | -0.27                     | -0.38    | -0.79                 | -1.34    | -0.83                    | -1.20    | -0.65                | -0.85    | -1.00                         | -1.66    |
|                               | 4 | 0.06                      |          | -0.32                 |          | -0.06                    |          | -0.06                |          | -0.31                         |          |
|                               | 5 | 0.42                      | 1.14     | 0.34                  | 1.20     | 0.49                     | 1.65     | 0.43                 | 1.31     | 0.32                          | 1.25     |
|                               | 6 | 0.92                      |          | 1.11                  |          | 1.49                     |          | 1.11                 |          | 1.27                          |          |
|                               | 7 | 1.41                      | 2.11     | 2.03                  | 2.80     | 2.57                     | 3.47     | 1.85                 | 2.68     | 2.33                          | 3.22     |
|                               | 8 | 1.98                      |          | 2.96                  |          | 3.83                     |          | 2.77                 |          | 3.56                          |          |
|                               | 9 | 2.91                      | 3.19     | 4.41                  | 3.91     | 5.37                     | 4.85     | 3.74                 | 3.71     | 4.85                          | 4.26     |
| (d)<br>Unweighted mean square | 0 | 0.97                      | 1.01     | 0.97                  | 0.93     | 0.97                     | 0.69     | 1.04                 | 1.12     | 1.40                          | 1.10     |
|                               | 1 | 1.24                      |          | 1.24                  |          | 0.75                     |          | 1.09                 |          | 1.47                          |          |
|                               | 2 | 1.20                      |          | 1.20                  |          | 0.72                     |          | 1.11                 |          | 0.79                          |          |
|                               | 3 | 1.19                      | 0.98     | 1.19                  | 0.92     | 1.25                     | 1.06     | 1.01                 | 0.99     | 1.38                          | 1.04     |
|                               | 4 | 1.02                      |          | 1.02                  |          | 1.48                     |          | 1.20                 |          | 1.08                          |          |
|                               | 5 | 0.93                      | 0.99     | 0.93                  | 0.92     | 0.86                     | 0.94     | 1.00                 | 0.96     | 0.78                          | 0.82     |
|                               | 6 | 1.00                      |          | 1.00                  |          | 1.00                     |          | 0.96                 |          | 0.85                          |          |
|                               | 7 | 0.96                      | 0.99     | 0.96                  | 0.94     | 0.99                     | 1.01     | 0.83                 | 0.92     | 0.86                          | 0.95     |
|                               | 8 | 0.95                      |          | 0.95                  |          | 0.93                     |          | 0.95                 |          | 0.94                          |          |
|                               | 9 | 1.01                      | 1.06     | 1.01                  | 1.21     | 0.97                     | 1.02     | 1.05                 | 1.09     | 1.11                          | 1.11     |

| (e), (h), and<br>Thresholds  |   | 0     | 1     | 2     | 3     | 4    | 5    | 6     | 7     | 8     | 9     |
|------------------------------|---|-------|-------|-------|-------|------|------|-------|-------|-------|-------|
|                              |   | ----  | ----  | ----  | ----  | ---- | ---- | ----  | ----  | ----  | ----  |
|                              |   | -1.86 |       | -3.11 |       |      |      | -3.10 |       | ----  | ----  |
|                              |   | -1.54 |       | -2.37 |       |      |      | -3.30 |       | -2.45 |       |
|                              |   | -1.35 |       | -1.92 |       |      |      | -2.35 |       | -2.08 |       |
|                              |   | -0.52 | -3.46 | -0.85 | -5.03 |      |      | -1.07 | -5.70 | -1.76 | -4.60 |
|                              |   | -0.38 |       | -0.51 |       |      |      | -0.64 |       | -0.84 | -1.05 |
|                              |   | 0.04  | -0.58 | 0.09  | -0.80 |      |      | 0.28  | -0.94 | -0.56 | -0.80 |
|                              |   | 0.76  | 1.65  | 1.32  | 2.20  |      |      | 1.50  | 2.49  | 0.14  | -0.25 |
|                              |   | 1.84  | 2.39  | 2.86  | 3.63  |      |      | 3.19  | 4.14  | 1.13  | 2.08  |
|                              |   | 3.00  |       | 4.48  |       |      |      | 5.50  |       | 2.50  | 3.32  |
|                              |   |       |       |       |       |      |      |       |       | 3.91  | 4.89  |
| (f)<br>Category →<br>Measure | 0 | 40%   | 51%   | 45%   | 66%   | 54%  | 60%  | 47%   | 48%   | 66%   | 53%   |
|                              | 1 | 30%   |       | 42%   |       | 47%  |      | 48%   |       | 31%   |       |
|                              | 2 | 17%   |       | 20%   |       | 43%  |      | 18%   |       | 18%   |       |
|                              | 3 | 14%   | 36%   | 23%   | 55%   | 22%  | 48%  | 18%   | 42%   | 39%   | 66%   |
|                              | 4 | 22%   |       | 35%   |       | 33%  |      | 22%   |       | 36%   |       |
|                              | 5 | 28%   |       | 41%   |       | 39%  |      | 34%   |       | 39%   |       |
|                              | 6 | 38%   | 51%   | 49%   | 67%   | 51%  | 64%  | 45%   | 59%   | 53%   | 63%   |
|                              | 7 | 54%   | 66%   | 55%   | 64%   | 59%  | 62%  | 55%   | 67%   | 62%   | 72%   |
|                              | 8 | 47%   |       | 49%   |       | 62%  |      | 55%   |       | 60%   |       |
|                              | 9 | 27%   | 53%   | 37%   | 50%   | 39%  | 64%  | 28%   | 49%   | 40%   | 46%   |
| (g)<br>Measure →<br>Category | 0 | 78%   | 77%   | 74%   | 81%   | 82%  | 80%  | 88%   | 73%   | 73%   | 69%   |
|                              | 1 | 32%   |       | 42%   |       | 33%  |      | 43%   |       | 52%   |       |
|                              | 2 | 26%   |       | 36%   |       | 48%  |      | 27%   |       | 20%   |       |
|                              | 3 | 27%   | 63%   | 34%   | 75%   | 46%  | 72%  | 31%   | 70%   | 39%   | 72%   |
|                              | 4 | 27%   |       | 31%   |       | 41%  |      | 28%   |       | 38%   |       |
|                              | 5 | 30%   |       | 38%   |       | 43%  |      | 33%   |       | 42%   |       |
|                              | 6 | 34%   | 56%   | 45%   | 64%   | 47%  | 66%  | 41%   | 63%   | 44%   | 67%   |
|                              | 7 | 41%   | 39%   | 48%   | 48%   | 52%  | 49%  | 47%   | 45%   | 56%   | 51%   |
|                              | 8 | 44%   |       | 50%   |       | 58%  |      | 48%   |       | 55%   |       |
|                              | 9 | 77%   | 77%   | 76%   | 72%   | 72%  | 75%  | 72%   | 72%   | 74%   | 74%   |
| Person reliability           |   | 0.95  | 0.94  | 0.92  | 0.90  | 0.91 | 0.87 | 0.88  | 0.85  | 0.87  | 0.82  |
| Person separation            |   | 4.20  | 3.81  | 3.42  | 3.01  | 3.26 | 2.64 | 2.71  | 2.36  | 2.57  | 2.17  |
| Item separation              |   | 7.06  | 6.89  | 6.19  | 5.25  | 6.21 | 5.88 | 5.08  | 4.88  | 3.53  | 3.47  |

Note. Bolded cells indicate failure to meet a guideline.

*To What Degree did the Proposed Internal Models Fit the Data?*

*Step 1.* The proposed unidimensional model and multidimensional model were imposed on  $S_T$ . The unidimensional model fit the data poorly,  $\chi^2_{(252)} = 2565.41$ ,  $p < .001$ ,  $\chi^2/df = 10.18$ , CFI = .81, TLI = .79, SRMR = .07, and RMSEA = .13. The multidimensional model fit better than the unidimensional model,  $CAIC_{diff} = 1316.56$ . However, the multidimensional model exhibited only marginally acceptable fit to the data,  $\chi^2_{(246)} = 1204.93$ ,  $p < .001$ ,  $\chi^2/df = 4.90$ , CFI = .92, TLI = .91, SRMR = .05, and RMSEA = .08. In both models the  $\chi^2$  value was likely inflated due to the fact that the multilevel structure of the data was ignored (Julian, 2001).

*Step 2.* ICC values for the 24 CCS items ranged from .16 to .36 ( $M = .25$ ,  $SD = .06$ ; see Table 2), which made it reasonable to proceed to Step 3. In order to proceed to Step 3 and Step 4,  $S_{PW}$  and  $S^*_B$  were calculated (see Table 3 and Table 4, respectively).

Table 2

*Item Characteristics for the CCS*

| Item  | <i>M</i> | <i>SD</i> | skewness | kurtosis | ICC |
|-------|----------|-----------|----------|----------|-----|
| MC1   | 3.67     | 1.07      | -0.33    | -0.69    | .22 |
| MC3   | 3.76     | 1.06      | -0.36    | -0.79    | .24 |
| MC6   | 3.42     | 1.12      | -0.09    | -0.88    | .29 |
| MC10  | 3.61     | 1.11      | -0.31    | -0.72    | .31 |
| MC12  | 3.70     | 1.10      | -0.37    | -0.71    | .29 |
| MC15  | 3.40     | 1.11      | -0.10    | -0.78    | .25 |
| MC23  | 3.71     | 1.12      | -0.47    | -0.65    | .31 |
| GSC2  | 4.16     | 0.94      | -0.90    | 0.14     | .21 |
| GSC4  | 4.13     | 0.99      | -0.86    | -0.10    | .24 |
| GSC8  | 3.97     | 1.02      | -0.63    | -0.42    | .24 |
| GSC9  | 4.15     | 0.98      | -0.92    | 0.06     | .20 |
| GSC11 | 3.82     | 1.06      | -0.51    | -0.54    | .16 |
| GSC17 | 3.84     | 1.03      | -0.53    | -0.54    | .22 |
| GSC21 | 3.92     | 1.05      | -0.56    | -0.65    | .18 |
| TC7   | 3.91     | 1.08      | -0.52    | -0.80    | .34 |
| TC14  | 3.61     | 1.03      | -0.21    | -0.68    | .20 |
| TC16  | 3.71     | 1.04      | -0.32    | -0.69    | .20 |
| TC18  | 3.87     | 1.12      | -0.62    | -0.52    | .16 |
| TC20  | 3.85     | 1.03      | -0.55    | -0.42    | .20 |
| TC22  | 4.02     | 1.02      | -0.77    | -0.16    | .26 |
| CBC5  | 3.94     | 1.12      | -0.76    | -0.40    | .35 |
| CBC13 | 3.88     | 1.09      | -0.62    | -0.51    | .25 |
| CBC19 | 4.09     | 1.06      | -0.93    | -0.04    | .36 |
| CBC24 | 3.95     | 1.13      | -0.80    | -0.29    | .29 |

*Note.* ICC = intraclass correlation.

Table 3

*Pooled Within-Teams Correlations and Covariances*

|       | MC1 | MC3 | MC6 | MC10 | MC12 | MC15 | MC23 | GSC2 | GSC4 | GSC8 | GSC9 | GSC11 | GSC17 | GSC21 | TC7 | TC14 | TC16 | TC18 | TC20 | TC22 | CBC5 | CBC13 | CBC19 | CBC24 |
|-------|-----|-----|-----|------|------|------|------|------|------|------|------|-------|-------|-------|-----|------|------|------|------|------|------|-------|-------|-------|
| MC1   | .90 | .50 | .63 | .56  | .51  | .62  | .60  | .44  | .47  | .48  | .41  | .48   | .48   | .42   | .40 | .47  | .50  | .46  | .37  | .46  | .47  | .40   | .37   | .46   |
| MC3   | .44 | .86 | .48 | .61  | .46  | .50  | .50  | .49  | .53  | .48  | .46  | .52   | .51   | .47   | .38 | .44  | .49  | .44  | .43  | .43  | .39  | .38   | .35   | .38   |
| MC6   | .57 | .43 | .91 | .59  | .53  | .72  | .65  | .34  | .40  | .44  | .33  | .46   | .50   | .39   | .43 | .51  | .54  | .48  | .43  | .40  | .55  | .43   | .39   | .47   |
| MC10  | .49 | .53 | .52 | .86  | .58  | .64  | .67  | .44  | .49  | .48  | .46  | .59   | .60   | .50   | .43 | .51  | .59  | .51  | .45  | .47  | .47  | .47   | .45   | .48   |
| MC12  | .45 | .39 | .47 | .50  | .87  | .54  | .61  | .42  | .42  | .49  | .43  | .54   | .51   | .47   | .37 | .46  | .51  | .50  | .49  | .44  | .46  | .55   | .43   | .50   |
| MC15  | .57 | .45 | .66 | .57  | .49  | .93  | .68  | .39  | .43  | .46  | .37  | .51   | .56   | .52   | .42 | .62  | .67  | .55  | .50  | .50  | .51  | .47   | .43   | .52   |
| MC23  | .54 | .43 | .58 | .58  | .53  | .61  | .89  | .43  | .46  | .48  | .43  | .53   | .57   | .53   | .39 | .50  | .57  | .51  | .49  | .53  | .48  | .47   | .48   | .59   |
| GSC2  | .35 | .38 | .27 | .34  | .33  | .32  | .34  | .71  | .59  | .60  | .70  | .53   | .52   | .54   | .44 | .43  | .43  | .44  | .47  | .45  | .41  | .43   | .45   | .39   |
| GSC4  | .39 | .42 | .33 | .39  | .34  | .36  | .38  | .43  | .75  | .64  | .58  | .55   | .54   | .58   | .50 | .47  | .45  | .51  | .52  | .53  | .49  | .41   | .46   | .44   |
| GSC8  | .41 | .40 | .38 | .40  | .41  | .40  | .41  | .45  | .49  | .80  | .69  | .59   | .56   | .61   | .52 | .48  | .49  | .53  | .51  | .53  | .42  | .45   | .45   | .45   |
| GSC9  | .34 | .38 | .28 | .37  | .36  | .31  | .36  | .52  | .44  | .54  | .77  | .57   | .52   | .56   | .46 | .43  | .44  | .48  | .47  | .47  | .39  | .46   | .45   | .41   |
| GSC11 | .44 | .47 | .43 | .53  | .49  | .47  | .48  | .44  | .47  | .51  | .48  | .94   | .61   | .59   | .45 | .52  | .53  | .54  | .52  | .52  | .45  | .47   | .44   | .45   |
| GSC17 | .42 | .43 | .44 | .50  | .44  | .49  | .49  | .40  | .43  | .46  | .42  | .54   | .84   | .62   | .46 | .55  | .66  | .60  | .57  | .59  | .49  | .45   | .49   | .48   |
| GSC21 | .38 | .42 | .36 | .44  | .42  | .48  | .47  | .43  | .48  | .52  | .47  | .54   | .54   | .91   | .42 | .56  | .55  | .58  | .62  | .63  | .40  | .41   | .48   | .43   |
| TC7   | .33 | .31 | .36 | .35  | .31  | .36  | .33  | .33  | .38  | .41  | .36  | .39   | .37   | .36   | .78 | .46  | .44  | .38  | .43  | .57  | .40  | .36   | .34   | .36   |
| TC14  | .41 | .38 | .45 | .44  | .40  | .56  | .44  | .34  | .38  | .40  | .35  | .47   | .47   | .49   | .38 | .87  | .68  | .52  | .54  | .62  | .42  | .44   | .41   | .39   |
| TC16  | .44 | .43 | .48 | .51  | .44  | .60  | .50  | .33  | .36  | .41  | .36  | .48   | .56   | .49   | .36 | .58  | .87  | .57  | .58  | .60  | .46  | .43   | .45   | .48   |
| TC18  | .44 | .43 | .48 | .48  | .47  | .55  | .49  | .38  | .45  | .49  | .43  | .54   | .56   | .57   | .35 | .50  | .55  | 1.06 | .56  | .54  | .48  | .44   | .54   | .50   |
| TC20  | .32 | .37 | .38 | .39  | .42  | .45  | .43  | .37  | .42  | .42  | .39  | .47   | .48   | .55   | .35 | .47  | .50  | .54  | .85  | .58  | .44  | .40   | .48   | .46   |
| TC22  | .39 | .35 | .33 | .38  | .37  | .43  | .44  | .33  | .41  | .42  | .36  | .45   | .47   | .53   | .44 | .51  | .49  | .49  | .47  | .78  | .42  | .40   | .47   | .45   |
| CBC5  | .40 | .33 | .48 | .40  | .39  | .45  | .41  | .31  | .39  | .35  | .31  | .40   | .41   | .35   | .32 | .36  | .39  | .45  | .37  | .34  | .83  | .56   | .61   | .61   |
| CBC13 | .36 | .33 | .39 | .41  | .49  | .43  | .42  | .34  | .33  | .38  | .39  | .44   | .39   | .37   | .30 | .39  | .38  | .43  | .35  | .34  | .49  | .90   | .60   | .54   |
| CBC19 | .30 | .28 | .32 | .36  | .34  | .36  | .39  | .32  | .34  | .35  | .34  | .37   | .38   | .40   | .26 | .32  | .36  | .47  | .38  | .35  | .47  | .49   | .73   | .64   |
| CBC24 | .41 | .34 | .43 | .43  | .45  | .48  | .53  | .32  | .36  | .38  | .35  | .42   | .42   | .40   | .30 | .35  | .43  | .50  | .40  | .38  | .53  | .49   | .53   | .91   |

*Note.* Italicized entries in the upper diagonal are the correlations; non-italicized entries are covariances.

Table 4

*Scaled Between-Teams Correlations and Covariances*

|       | MC1 | MC3 | MC6 | MC10 | MC12 | MC15 | MC23 | GSC2 | GSC4 | GSC8 | GSC9 | GSC11 | GSC17 | GSC21 | TC7 | TC14 | TC16 | TC18 | TC20 | TC22 | CBC5 | CBC13 | CBC19 | CBC24 |
|-------|-----|-----|-----|------|------|------|------|------|------|------|------|-------|-------|-------|-----|------|------|------|------|------|------|-------|-------|-------|
| MC1   | 5.6 | .84 | .91 | .92  | .78  | .91  | .91  | .71  | .69  | .70  | .63  | .72   | .78   | .70   | .64 | .80  | .80  | .74  | .68  | .75  | .80  | .67   | .62   | .80   |
| MC3   | 4.7 | 5.7 | .77 | .89  | .84  | .76  | .84  | .86  | .82  | .81  | .80  | .86   | .90   | .82   | .53 | .77  | .83  | .82  | .79  | .76  | .74  | .73   | .67   | .80   |
| MC6   | 5.9 | 5.1 | 7.6 | .87  | .81  | .95  | .91  | .64  | .59  | .59  | .54  | .68   | .68   | .62   | .56 | .74  | .75  | .72  | .63  | .68  | .89  | .77   | .75   | .88   |
| MC10  | 6.0 | 5.9 | 6.7 | 7.7  | .86  | .88  | .94  | .70  | .69  | .67  | .62  | .75   | .80   | .71   | .47 | .70  | .78  | .81  | .68  | .66  | .80  | .75   | .71   | .85   |
| MC12  | 4.9 | 5.4 | 6.0 | 6.6  | 7.2  | .80  | .88  | .74  | .66  | .67  | .62  | .78   | .80   | .76   | .47 | .68  | .79  | .79  | .74  | .69  | .84  | .86   | .79   | .90   |
| MC15  | 5.5 | 4.7 | 6.6 | 6.3  | 5.5  | 6.5  | .89  | .64  | .63  | .62  | .55  | .69   | .72   | .63   | .63 | .77  | .77  | .74  | .63  | .68  | .85  | .77   | .71   | .82   |
| MC23  | 6.0 | 5.6 | 7.0 | 7.3  | 6.6  | 6.4  | 7.9  | .69  | .65  | .67  | .62  | .73   | .78   | .72   | .52 | .71  | .76  | .76  | .67  | .69  | .87  | .81   | .77   | .92   |
| GSC2  | 3.4 | 4.2 | 3.5 | 3.9  | 4.0  | 3.3  | 3.9  | 4.1  | .91  | .92  | .96  | .93   | .92   | .89   | .68 | .79  | .85  | .86  | .85  | .81  | .68  | .67   | .62   | .72   |
| GSC4  | 3.7 | 4.5 | 3.7 | 4.3  | 4.0  | 3.7  | 4.1  | 4.1  | 5.1  | .92  | .92  | .91   | .90   | .89   | .73 | .82  | .87  | .84  | .87  | .88  | .66  | .67   | .60   | .69   |
| GSC8  | 3.8 | 4.5 | 3.7 | 4.3  | 4.1  | 3.6  | 4.3  | 4.3  | 4.8  | 5.3  | .92  | .91   | .93   | .91   | .71 | .83  | .87  | .85  | .88  | .83  | .65  | .68   | .60   | .68   |
| GSC9  | 3.1 | 4.0 | 3.1 | 3.6  | 3.5  | 2.9  | 3.6  | 4.0  | 4.3  | 4.4  | 4.4  | .89   | .89   | .85   | .66 | .78  | .80  | .78  | .83  | .79  | .63  | .62   | .59   | .66   |
| GSC11 | 3.5 | 4.3 | 3.8 | 4.3  | 4.3  | 3.6  | 4.2  | 3.9  | 4.3  | 4.3  | 3.8  | 4.3   | .93   | .93   | .63 | .84  | .89  | .88  | .90  | .85  | .74  | .77   | .73   | .78   |
| GSC17 | 4.2 | 4.9 | 4.2 | 5.0  | 4.9  | 4.2  | 5.0  | 4.2  | 4.6  | 4.9  | 4.2  | 4.4   | 5.1   | .93   | .66 | .81  | .89  | .89  | .88  | .84  | .75  | .77   | .71   | .78   |
| GSC21 | 3.5 | 4.1 | 3.6 | 4.2  | 4.3  | 3.4  | 4.2  | 3.8  | 4.2  | 4.4  | 3.7  | 4.0   | 4.4   | 4.4   | .61 | .77  | .86  | .86  | .90  | .83  | .72  | .75   | .72   | .77   |
| TC7   | 4.3 | 3.6 | 4.4 | 3.7  | 3.6  | 4.5  | 4.2  | 3.9  | 4.7  | 4.6  | 3.9  | 3.7   | 4.2   | 3.7   | 8.0 | .78  | .78  | .64  | .65  | .84  | .58  | .51   | .40   | .53   |
| TC14  | 4.1 | 4.0 | 4.4 | 4.2  | 4.0  | 4.3  | 4.4  | 3.5  | 4.0  | 4.2  | 3.6  | 3.8   | 4.0   | 3.5   | 4.8 | 4.8  | .93  | .78  | .86  | .90  | .74  | .69   | .61   | .68   |
| TC16  | 4.2 | 4.4 | 4.5 | 4.8  | 4.7  | 4.3  | 4.7  | 3.8  | 4.3  | 4.4  | 3.7  | 4.1   | 4.4   | 4.0   | 4.9 | 4.5  | 4.9  | .88  | .90  | .94  | .76  | .74   | .65   | .76   |
| TC18  | 3.8 | 4.3 | 4.3 | 4.9  | 4.6  | 4.1  | 4.6  | 3.7  | 4.1  | 4.2  | 3.5  | 4.0   | 4.4   | 3.9   | 3.9 | 3.7  | 4.2  | 4.7  | .85  | .78  | .75  | .74   | .74   | .78   |
| TC20  | 3.5 | 4.1 | 3.8 | 4.1  | 4.3  | 3.5  | 4.1  | 3.7  | 4.3  | 4.4  | 3.8  | 4.0   | 4.3   | 4.1   | 4.0 | 4.0  | 4.3  | 4.0  | 4.7  | .88  | .75  | .77   | .72   | .74   |
| TC22  | 4.2 | 4.3 | 4.4 | 4.4  | 4.4  | 4.1  | 4.6  | 3.9  | 4.7  | 4.6  | 3.9  | 4.2   | 4.5   | 4.2   | 5.7 | 4.7  | 5.0  | 4.1  | 4.5  | 5.7  | .74  | .71   | .62   | .73   |
| CBC5  | 5.6 | 5.3 | 7.3 | 6.6  | 6.7  | 6.4  | 7.3  | 4.1  | 4.4  | 4.5  | 3.9  | 4.5   | 5.0   | 4.5   | 4.9 | 4.8  | 5.0  | 4.8  | 4.8  | 5.2  | 8.9  | .91   | .92   | .94   |
| CBC13 | 3.9 | 4.3 | 5.3 | 5.1  | 5.7  | 4.9  | 5.6  | 3.3  | 3.8  | 3.9  | 3.2  | 4.0   | 4.3   | 3.9   | 3.6 | 3.7  | 4.1  | 4.0  | 4.2  | 4.2  | 6.7  | 6.2   | .93   | .91   |
| CBC19 | 4.1 | 4.5 | 5.8 | 5.6  | 6.0  | 5.1  | 6.1  | 3.5  | 3.8  | 3.9  | 3.5  | 4.3   | 4.5   | 4.3   | 3.2 | 3.8  | 4.0  | 4.5  | 4.4  | 4.1  | 7.7  | 6.5   | 8.0   | .91   |
| CBC24 | 5.2 | 5.3 | 6.7 | 6.5  | 6.6  | 5.8  | 7.1  | 4.0  | 4.3  | 4.4  | 3.8  | 4.4   | 4.9   | 4.5   | 4.1 | 4.1  | 4.6  | 4.7  | 4.4  | 4.8  | 7.8  | 6.2   | 7.1   | 7.6   |

*Step 3.* Prior to imposing the internal models on  $S_{PW}$ , item-level correlations were examined. Correlations among items within proposed subscales appeared to be greater than correlations between items outside of respective subscales: MC items with MC items ( $M = .58, SD = .07$ ), MC items with non-MC items ( $M = .47, SD = .06$ ); GSC items with GSC items ( $M = .58, SD = .05$ ), GSC items with non-GSC items ( $M = .48, SD = .06$ ); TC items with TC items ( $M = .54, SD = .08$ ), TC items with non-TC items ( $M = .48, SD = .06$ ); and, CBC items with CBC items ( $M = .59, SD = .04$ ), CBC items with non-CBC items ( $M = .45, SD = .05$ ). This provided some support for the multidimensional model.

The proposed unidimensional model and multidimensional model were imposed on  $S_{PW}$  (see Table 5). The unidimensional model fit the data poorly,  $\chi^2_{(252)} = 1732.15, p < .001, \chi^2/df = 6.87, CFI = .83, TLI = .82, SRMR = .06$ , and  $RMSEA = .10$ . The multidimensional model fit better than the unidimensional model,  $CAIC_{diff} = 715.84$ . However, the multidimensional model exhibited only marginally acceptable fit to the data,  $\chi^2_{(246)} = 972.39, p < .001, \chi^2/df = 3.95, CFI = .92, TLI = .91, SRMR = .05$ , and  $RMSEA = .07$ . As expected, both models appeared to fit the data better than the parallel models in Step 1.

Because the multidimensional model marginally fit the data, model respecification was considered. None of the Wald tests were statistically significant. LM values and standardized residuals suggested several potential modifications, three of which were considered defensible and were adopted: correlated error terms between items GSC2 (*recognize opposing team's strength during competition*) and GSC9 (*recognize opposing team's weaknesses during competition*), and GSC9 and GSC8 (*adapt to different game situations*), and MC3 loaded on MC and GSC. In regard to the

correlated errors, both pairs of items may have measured athletes' perceptions of their coach's competence in recognizing and making critical decisions about *the other team* during competition, in addition to measuring a more general competency to lead during competition. In regard to the within item multidimensionality for MC3 (*mentally prepare athletes for game strategies*), it makes sense that it may have measured both MC (*athletes' evaluations of their head coach's ability to affect the psychological mood and skills of athletes*) and GSC (*athletes' evaluations of their head coach's ability to lead during competition*).

First, error terms for GSC2 and GSC9 were allowed to correlate ( $r = .39, p = .02$ ); and this model fit better than the previous model,  $\chi^2_{LR(1)} = 75.55, p < .001$ ,  $CAIC_{diff} = 68.23$ , and negated what had been the second largest standardized residual (.15). Second, error terms for GSC9 and GSC8 were allowed to correlate ( $r = .27, p = .02$ ); and this model fit better than the previous model,  $\chi^2_{LR(1)} = 42.76, p < .001$ ,  $CAIC_{diff} = 35.44$ , and dramatically reduced (.02) what had been the fifth largest standardized residual (.12). Third, MC3 was allowed to load on both MC, .46,  $SE = .09$ , and GSC, .57,  $SE = .11$ ; and this model fit better than the previous model,  $\chi^2_{LR(1)} = 28.67, p < .001$ ,  $CAIC_{diff} = 21.35$ , and dramatically reduced (.05, .04, .03) what were three of the eight largest standardized residuals (.13, .12, .11). Overall, this final model marginally fit the data,  $\chi^2_{(243)} = 825.41, p < .001$ ,  $\chi^2/df = 3.40$ , CFI = .93, TLI = .93, SRMR = .04, and RMSEA = .07. Table 6 illustrates within-teams estimates for this model.

Factors in the retained model were moderately to highly correlated with one another, with the strongest association occurring between GSC and TC,  $r = .92$ , and the weakest associations occurring between CBC and GSC, and CBC and TC,  $r = .76$ .

Because of the high correlation between GSC and TC another model was specified where all of these items loaded on only one factor. This simpler model fit worse than the previous model,  $CAIC_{diff} = 53.74$ . Thus, the previous model was retained as the final model. Reasons for retaining, and implications of, subscales with such high correlations with one another will be explicated in the Discussion section.

Table 5

*Model-Data Fit Statistics*

| Model                                                | $\chi^2$ | $df$ | $\chi^2/df$ | $\chi^2_{LR}$ | $df_{diff}$ | CAIC    | CAIC <sub>diff</sub> | CFI  | TLI  | SRMR | RMSEA |
|------------------------------------------------------|----------|------|-------------|---------------|-------------|---------|----------------------|------|------|------|-------|
| <b>Step 3 (on S<sub>pw</sub>)</b>                    |          |      |             |               |             |         |                      |      |      |      |       |
| Unidimensional                                       | 1732.15  | 252  | 6.87        | -----         | -----       | 2083.51 | -----                | 0.83 | 0.82 | 0.06 | 0.10  |
| Multidimensional                                     | 972.39   | 246  | 3.95        | -----         | 6           | 1367.67 | 715.84               | 0.92 | 0.91 | 0.05 | 0.07  |
| 1st respecification                                  | 896.84   | 245  | 3.66        | 75.55         | 1           | 1299.44 | 68.23                | 0.93 | 0.92 | 0.04 | 0.07  |
| 2nd respecification                                  | 854.08   | 244  | 3.50        | 42.76         | 1           | 1264.00 | 35.44                | 0.93 | 0.92 | 0.04 | 0.07  |
| 3rd respecification                                  | 825.41   | 243  | 3.40        | 28.67         | 1           | 1242.65 | 21.35                | 0.93 | 0.93 | 0.04 | 0.07  |
| <b>Step 4 (on S<sub>pw</sub> and S*<sub>B</sub>)</b> |          |      |             |               |             |         |                      |      |      |      |       |
| Unidimensional<br>between                            | 1539.33  | 521  | 2.95        | -----         | -----       | 2114.19 | -----                | 0.92 | 0.91 | 0.18 | 0.06  |
| Multidimensional<br>between                          | 1401.39  | 515  | 2.72        | -----         | 6           | 2027.84 | 86.35                | 0.93 | 0.92 | 0.07 | 0.05  |
| 1st respecification                                  | 1385.89  | 514  | 2.70        | 15.50         | 6           | 2019.71 | 8.13                 | 0.93 | 0.93 | 0.06 | 0.05  |

*Note.*  $df$  = degrees of freedom,  $df_{diff}$  = degrees of freedom difference compared to the previous model, CAIC = consistent Akaike information criterion, CFI = comparative fit index, TLI = Tucker-Lewis index, SRMR = standardized root mean squared residual, and RMSEA = root mean square error of approximation.

*Step 4.* Prior to imposing the internal models, item-level correlations within  $S^*_B$  were examined. Correlations between items within proposed subscales appeared to be greater than correlations between items outside of respective subscales: MC items with MC items ( $M = .87, SD = .06$ ), MC items with non-MC items ( $M = .72, SD = .09$ ); GSC items with GSC items ( $M = .91, SD = .02$ ), GSC items with non-GSC items ( $M = .74, SD = .09$ ); TC items with TC items ( $M = .83, SD = .09$ ), TC items with non-TC items ( $M = .74, SD = .09$ ); and, CBC items with CBC items ( $M = .92, SD = .02$ ), CBC items with non-CBC items ( $M = .72, SD = .09$ ). This provided some support for the multidimensional model.

Using the multi-group procedure, the accepted within teams model was specified in both groups for all subsequent analyses. The scaling factor,  $c^*$ , equaled 4.27. Both a between teams null model,  $\chi^2_{(567)} = 2087.01, p < .001$ , and a between teams independent model,  $\chi^2_{(541)} = 1751.25, p < .001$ , were rejected. In both models, the estimated covariances of the pair of correlated error terms (GSC8 and GSC11 = .139, GSC10 and GSC11 = .096) were inserted as starting values due to convergence problems.<sup>7</sup> To maintain model comparability, these starting values were inserted from this point forward. Next, the unidimensional structure was specified between teams (see Table 5). This model fit the data poorly to marginally,  $\chi^2_{(521)} = 1539.33, p < .001, \chi^2/df = 2.95, CFI = .92, TLI = .91, SRMR = .18$ , and  $RMSEA = .06$ . Next, the original multidimensional structure (i.e., no correlated error terms or cross-loadings) was specified between teams (see Table 5). This model fit better than the previous model,  $CAIC_{diff} = 86.35$ . However, this model exhibited only marginally acceptable fit to the data,  $\chi^2_{(515)} = 1401.39, p < .001$ ,

$\chi^2/df = 2.72$ , CFI = .93, TLI = .92, SRMR = .07, and RMSEA = .05. Because the multidimensional model marginally fit the data, model respecification was considered.

Due to concerns associated with a modest sample size at Level 2 (e.g., stability of between team estimates), only one model respecification was considered defensible: MC3 loaded on both MC and GSC. MC3 was allowed to load on both MC, .50,  $SE = .13$ , and GSC, .66,  $SE = .15$ ; this model fit better than the previous model,  $\chi^2_{LR(1)} = 15.50$ ,  $p < .001$ ,  $CAIC_{diff} = 8.13$ , and dramatically reduced (range = .00 - .05) what were seven of the 11 largest standardized residuals at the between-level (range = .28 to .19). Overall, this final model marginally fit the data,  $\chi^2_{(514)} = 1385.89$ ,  $p < .001$ ,  $\chi^2/df = 2.70$ , CFI = .93, TLI = .93, SRMR = .06, and RMSEA = .05. Table 6 illustrates between-teams estimates for this model. Because a multidimensional model was retained at both the within-teams and between-teams levels, the remaining analyses focus on multidimensional estimates of coaching competency only.

Factors in the between teams portion of the model were also moderately to highly correlated with one another, with the strongest association occurring between GSC and TC,  $r = .93$ , and the weakest association occurring between MC and CSC, and GSC and CBC,  $r = .74$ . Because of the high correlation between GSC and TC another model was specified where all of these items loaded on only one factor. This simpler model fit worse than the previous model,  $CAIC_{diff} = 72.00$ . Thus, the previous model was retained as the final model. Reasons for retaining, and implications of, such high correlations among subscales will be explicated in the Discussion section.

Table 6

*Within-Teams and Between-Teams Estimates*

| Item  | Within teams                       |                                   |                                   |              | Between teams                       |                                   |                                   |              |
|-------|------------------------------------|-----------------------------------|-----------------------------------|--------------|-------------------------------------|-----------------------------------|-----------------------------------|--------------|
|       | Loading                            | SE                                | Std Loading                       | Std Residual | Loading                             | SE                                | Std Loading                       | Std Residual |
| MC1   | 1.00 <sup>a</sup>                  | 0.00                              | .74                               | .68          | 1.00 <sup>a</sup>                   | 0.00                              | 0.98                              | 0.20         |
| MC3   | .46 <sup>b</sup> .57 <sup>c*</sup> | .09 <sup>b</sup> .11 <sup>c</sup> | .35 <sup>b</sup> .36 <sup>c</sup> | .74          | .50 <sup>b*</sup> .66 <sup>c*</sup> | .13 <sup>b</sup> .15 <sup>c</sup> | .48 <sup>b</sup> .55 <sup>c</sup> | 0.26         |
| MC6   | 1.08*                              | 0.06                              | .79                               | .62          | 1.18*                               | 0.11                              | 0.97                              | 0.25         |
| MC10  | 1.05*                              | 0.06                              | .79                               | .61          | 1.18*                               | 0.11                              | 0.97                              | 0.25         |
| MC12  | 0.95*                              | 0.06                              | .71                               | .70          | 1.05*                               | 0.14                              | 0.89                              | 0.45         |
| MC15  | 1.15*                              | 0.06                              | .83                               | .55          | 1.07*                               | 0.11                              | 0.96                              | 0.29         |
| MC23  | 1.11*                              | 0.06                              | .83                               | .57          | 1.22*                               | 0.11                              | 0.99                              | 0.17         |
| GSC2  | 1.00 <sup>a</sup>                  | 0.00                              | .70                               | .71          | 1.00 <sup>a</sup>                   | 0.00                              | 1.00                              | 0.00         |
| GSC4  | 1.10*                              | 0.07                              | .75                               | .66          | 1.13*                               | 0.10                              | 0.98                              | 0.19         |
| GSC8  | 1.17*                              | 0.07                              | .77                               | .64          | 1.14*                               | 0.10                              | 0.99                              | 0.14         |
| GSC9  | 1.05*                              | 0.05                              | .71                               | .70          | 1.03*                               | 0.08                              | 0.99                              | 0.14         |
| GSC11 | 1.26*                              | 0.07                              | .77                               | .64          | .99*                                | 0.09                              | 1.00                              | 0.00         |
| GSC17 | 1.22*                              | 0.07                              | .79                               | .61          | 1.11*                               | 0.09                              | 0.99                              | 0.06         |
| GSC21 | 1.27*                              | 0.07                              | .79                               | .62          | 0.99*                               | 0.10                              | 0.98                              | 0.18         |
| TC7   | 1.00 <sup>a</sup>                  | 0.00                              | .62                               | .79          | 1.00 <sup>a</sup>                   | 0.00                              | 0.83                              | 0.56         |
| TC14  | 1.29*                              | 0.09                              | .76                               | .65          | 0.84*                               | 0.14                              | 0.97                              | 0.24         |
| TC16  | 1.35*                              | 0.09                              | .79                               | .61          | 0.88*                               | 0.14                              | 1.00                              | 0.00         |
| TC18  | 1.38*                              | 0.09                              | .73                               | .68          | 0.75*                               | 0.14                              | 0.92                              | 0.39         |
| TC20  | 1.25*                              | 0.09                              | .74                               | .68          | 0.81*                               | 0.14                              | 0.96                              | 0.27         |
| TC22  | 1.24*                              | 0.08                              | .77                               | .64          | 0.96*                               | 0.15                              | 0.99                              | 0.17         |
| CBC5  | 1.00 <sup>a</sup>                  | 0.00                              | .78                               | .63          | 1.00 <sup>a</sup>                   | 0.00                              | 0.99                              | 0.15         |
| CBC13 | 0.98*                              | 0.06                              | .73                               | .68          | 0.78*                               | 0.07                              | 0.97                              | 0.25         |
| CBC19 | 0.96*                              | 0.05                              | .79                               | .62          | 0.89*                               | 0.08                              | 0.95                              | 0.32         |
| CBC24 | 1.06*                              | 0.06                              | .79                               | .62          | 0.92*                               | 0.06                              | 0.99                              | 0.06         |

*Note.* <sup>a</sup> indicates that the parameter was fixed to set the metric; \* indicates statistical significance at the .001 level; <sup>b</sup> indicates MC; and <sup>c</sup> indicates GSC.

Comparisons of the vast majority (98%) of factor loadings ( $n = 21$  for each comparison) and factor covariances ( $n = 6$  for each comparison) did not reject the hypothesis of invariant structures across sub-samples. All of the factor loadings, and most of the factor covariances were not significantly different for males and females (see Table 7). The two factor covariances that were significantly different suggested that CBC was more strongly related to MC and TC for females,  $r = .84$  and  $.77$ , than for males,  $r = .70$  and  $.69$ , respectively. None of the said estimates varied by sport (see Table 8) or year on team (see Table 9). Because 106 of 108 of the said estimates were not significantly different across selected sub-samples, assuming that all of the athletes were a random sample of a single population was considered reasonable. It is noted that the possible violations of invariance of the identified factor covariances would likely cause greater misfit than would be observed if all of the estimates were invariant.

**Table 7**  
*Comparisons of Factor Loadings and Correlations Among Factors by Gender*

|         | Standardized Loading |      | Unstandardized Loading |      | Standard Error |      | Factor Correlation |            |
|---------|----------------------|------|------------------------|------|----------------|------|--------------------|------------|
|         | female               | male | female                 | male | female         | male | female             | male       |
| v1 mc   | *                    | *    | *                      | *    | *              | *    | ----               | ----       |
| v2 mc   | .36                  | .41  | 0.46                   | 0.51 | .08            | .16  | ----               | ----       |
| v2 gsc  | .43                  | .36  | 0.59                   | 0.53 | .08            | .18  | ----               | ----       |
| v3 mc   | .84                  | .78  | 1.12                   | 1.03 | .06            | .09  | ----               | ----       |
| v4 mc   | .84                  | .80  | 1.13                   | 1.02 | .06            | .09  | ----               | ----       |
| v5 mc   | .77                  | .69  | 1.03                   | 0.87 | .06            | .09  | ----               | ----       |
| v6 mc   | .85                  | .87  | 1.10                   | 1.17 | .06            | .09  | ----               | ----       |
| v7 mc   | .88                  | .82  | 1.22                   | 1.02 | .06            | .09  | ----               | ----       |
| v8 gsc  | *                    | *    | *                      | *    | *              | *    | ----               | ----       |
| v9 gsc  | .81                  | .80  | 1.07                   | 1.11 | .06            | .10  | ----               | ----       |
| v10 gsc | .86                  | .80  | 1.14                   | 1.20 | .06            | .11  | ----               | ----       |
| v11 gsc | .81                  | .77  | 1.05                   | 1.09 | .06            | .11  | ----               | ----       |
| v12 gsc | .81                  | .80  | 1.13                   | 1.21 | .06            | .11  | ----               | ----       |
| v13 gsc | .82                  | .82  | 1.13                   | 1.14 | .06            | .10  | ----               | ----       |
| v14 gsc | .83                  | .79  | 1.16                   | 1.14 | .06            | .11  | ----               | ----       |
| v15 tc  | *                    | *    | *                      | *    | *              | *    | ----               | ----       |
| v16 tc  | .81                  | .74  | 1.13                   | 1.15 | .07            | .16  | ----               | ----       |
| v17 tc  | .83                  | .84  | 1.15                   | 1.43 | .07            | .18  | ----               | ----       |
| v18 tc  | .79                  | .67  | 1.21                   | 1.22 | .08            | .18  | ----               | ----       |
| v19 tc  | .79                  | .76  | 1.12                   | 1.18 | .07            | .16  | ----               | ----       |
| v20 tc  | .83                  | .75  | 1.17                   | 1.10 | .07            | .15  | ----               | ----       |
| v21 cbc | *                    | *    | *                      | *    | *              | *    | ----               | ----       |
| v22 cbc | .82                  | .64  | .93                    | .88  | .05            | .10  | ----               | ----       |
| v23 cbc | .85                  | .78  | .97                    | .96  | .04            | .09  | ----               | ----       |
| v24 cbc | .86                  | .81  | 1.03                   | 1.06 | .05            | .09  | ----               | ----       |
| mc gsc  | ----                 | ---- | ----                   | ---- | ----           | ---- | .75                | .84        |
| mc tc   | ----                 | ---- | ----                   | ---- | ----           | ---- | .81                | .85        |
| mc cbc  | ----                 | ---- | ----                   | ---- | ----           | ---- | <b>.83</b>         | <b>.70</b> |
| gsc tc  | ----                 | ---- | ----                   | ---- | ----           | ---- | .92                | .91        |
| gsc cbc | ----                 | ---- | ----                   | ---- | ----           | ---- | .76                | .68        |
| tc cbc  | ----                 | ---- | ----                   | ---- | ----           | ---- | <b>.77</b>         | <b>.69</b> |

*Note.* v = variable; mc = motivation competence; gsc = game strategy competence; tc = technique competence; cbc = character building competence; \* = fixed; Bolded cells indicate significant ( $p < .001$ ) invariance between estimates.

Table 8

*Comparisons of Factor Loadings and Correlations Among Factors by Sport*

|         | Standardized<br>Loading |        | Unstandardized<br>Loading |        | Standard<br>Error |        | Factor<br>Correlation |        |
|---------|-------------------------|--------|---------------------------|--------|-------------------|--------|-----------------------|--------|
|         | soccer                  | hockey | soccer                    | hockey | soccer            | hockey | soccer                | hockey |
| v1 mc   | *                       | *      | *                         | *      | *                 | *      | ----                  | ----   |
| v2 mc   | .41                     | .31    | 0.52                      | 0.38   | .08               | .11    | ----                  | ----   |
| v2 gsc  | .35                     | .53    | 0.54                      | 0.65   | .10               | .11    | ----                  | ----   |
| v3 mc   | .83                     | .86    | 1.13                      | 1.03   | .06               | .07    | ----                  | ----   |
| v4 mc   | .82                     | .88    | 1.09                      | 1.08   | .06               | .07    | ----                  | ----   |
| v5 mc   | .72                     | .84    | 0.94                      | 1.01   | .06               | .07    | ----                  | ----   |
| v6 mc   | .87                     | .85    | 1.17                      | 1.00   | .06               | .07    | ----                  | ----   |
| v7 mc   | .87                     | .88    | 1.16                      | 1.13   | .06               | .07    | ----                  | ----   |
| v8 gsc  | *                       | *      | *                         | *      | *                 | *      | ----                  | ----   |
| v9 gsc  | .79                     | .85    | 1.14                      | 1.02   | .07               | .06    | ----                  | ----   |
| v10 gsc | .80                     | .91    | 1.19                      | 1.12   | .07               | .06    | ----                  | ----   |
| v11 gsc | .77                     | .87    | 1.11                      | 1.01   | .07               | .06    | ----                  | ----   |
| v12 gsc | .77                     | .86    | 1.16                      | 1.12   | .07               | .07    | ----                  | ----   |
| v13 gsc | .77                     | .89    | 1.12                      | 1.15   | .07               | .07    | ----                  | ----   |
| v14 gsc | .77                     | .87    | 1.17                      | 1.11   | .08               | .07    | ----                  | ----   |
| v15 tc  | *                       | *      | *                         | *      | *                 | *      | ----                  | ----   |
| v16 tc  | .78                     | .86    | 1.25                      | 1.04   | .10               | .08    | ----                  | ----   |
| v17 tc  | .82                     | .87    | 1.31                      | 1.08   | .10               | .08    | ----                  | ----   |
| v18 tc  | .69                     | .85    | 1.15                      | 1.19   | .10               | .09    | ----                  | ----   |
| v19 tc  | .76                     | .83    | 1.17                      | 1.05   | .10               | .08    | ----                  | ----   |
| v20 tc  | .78                     | .90    | 1.22                      | 1.09   | .10               | .08    | ----                  | ----   |
| v21 cbc | *                       | *      | *                         | *      | *                 | *      | ----                  | ----   |
| v22 cbc | .75                     | .85    | 0.88                      | 0.92   | .05               | .06    | ----                  | ----   |
| v23 cbc | .82                     | .88    | 0.92                      | 0.96   | .05               | .06    | ----                  | ----   |
| v24 cbc | .84                     | .87    | 1.00                      | 1.00   | .05               | .06    | ----                  | ----   |
| mc gsc  | ----                    | ----   | ----                      | ----   | ----              | ----   | .74                   | .82    |
| mc tc   | ----                    | ----   | ----                      | ----   | ----              | ----   | .81                   | .87    |
| mc cbc  | ----                    | ----   | ----                      | ----   | ----              | ----   | .79                   | .85    |
| gsc tc  | ----                    | ----   | ----                      | ----   | ----              | ----   | .88                   | .94    |
| gsc cbc | ----                    | ----   | ----                      | ----   | ----              | ----   | .67                   | .84    |
| tc cbc  | ----                    | ----   | ----                      | ----   | ----              | ----   | .71                   | .84    |

*Note.* v = variable; mc = motivation competence; gsc = game strategy competence; tc = technique competence; cbc = character building competence; \* = fixed; Bolded cells indicate significant ( $p < .001$ ) invariance between estimates.

**Table 9**  
*Comparisons of Factor Loadings and Correlations Among Factors by Year*

|         | Standardized Loading |      |      | Unstandardized Loading |      |      | Standard Error |      |      | Factor Correlation |      |      |
|---------|----------------------|------|------|------------------------|------|------|----------------|------|------|--------------------|------|------|
|         | 1st                  | 2nd  | 3rd  | 1st                    | 2nd  | 3rd  | 1st            | 2nd  | 3rd  | 1st                | 2nd  | 3rd  |
| v1 mc   | *                    | *    | *    | *                      | *    | *    | *              | *    | *    | ----               | ---- | ---- |
| v2 mc   | .51                  | .45  | .25  | 0.67                   | 0.55 | 0.30 | .13            | .11  | .10  | ----               | ---- | ---- |
| v2 gsc  | .25                  | .40  | .58  | 0.37                   | 0.55 | 0.82 | .14            | .12  | .12  | ----               | ---- | ---- |
| v3 mc   | .82                  | .87  | .84  | 1.10                   | 1.14 | 1.10 | .08            | .09  | .08  | ----               | ---- | ---- |
| v4 mc   | .85                  | .85  | .83  | 1.14                   | 1.11 | 1.06 | .08            | .09  | .08  | ----               | ---- | ---- |
| v5 mc   | .75                  | .81  | .80  | 0.99                   | 1.04 | 1.00 | .08            | .10  | .08  | ----               | ---- | ---- |
| v6 mc   | .84                  | .92  | .83  | 1.13                   | 1.19 | 1.06 | .08            | .09  | .08  | ----               | ---- | ---- |
| v7 mc   | .87                  | .91  | .85  | 1.16                   | 1.23 | 1.08 | .08            | .09  | .08  | ----               | ---- | ---- |
| v8 gsc  | *                    | *    | *    | *                      | *    | *    | *              | *    | *    | ----               | ---- | ---- |
| v9 gsc  | .79                  | .84  | .83  | 1.09                   | 1.04 | 1.10 | .08            | .09  | .09  | ----               | ---- | ---- |
| v10 gsc | .85                  | .85  | .83  | 1.24                   | 1.06 | 1.12 | .09            | .09  | .09  | ----               | ---- | ---- |
| v11 gsc | .79                  | .80  | .83  | 1.08                   | 0.98 | 1.11 | .08            | .09  | .09  | ----               | ---- | ---- |
| v12 gsc | .81                  | .84  | .77  | 1.21                   | 1.14 | 1.08 | .09            | .09  | .09  | ----               | ---- | ---- |
| v13 gsc | .85                  | .87  | .77  | 1.25                   | 1.15 | 1.02 | .09            | .09  | .09  | ----               | ---- | ---- |
| v14 gsc | .81                  | .85  | .78  | 1.19                   | 1.21 | 1.06 | .09            | .10  | .09  | ----               | ---- | ---- |
| v15 tc  | *                    | *    | *    | *                      | *    | *    | *              | *    | *    | ----               | ---- | ---- |
| v16 tc  | .82                  | .80  | .78  | 1.20                   | 1.22 | 0.98 | .11            | .14  | .10  | ----               | ---- | ---- |
| v17 tc  | .86                  | .87  | .79  | 1.28                   | 1.35 | 1.00 | .11            | .14  | .10  | ----               | ---- | ---- |
| v18 tc  | .77                  | .82  | .71  | 1.22                   | 1.34 | 0.98 | .11            | .15  | .11  | ----               | ---- | ---- |
| v19 tc  | .79                  | .85  | .79  | 1.18                   | 1.19 | 1.03 | .11            | .13  | .10  | ----               | ---- | ---- |
| v20 tc  | .83                  | .84  | .80  | 1.23                   | 1.24 | 0.98 | .11            | .13  | .10  | ----               | ---- | ---- |
| v21 cbc | *                    | *    | *    | *                      | *    | *    | *              | *    | *    | ----               | ---- | ---- |
| v22 cbc | .76                  | .87  | .77  | 0.84                   | 1.00 | 0.91 | .06            | .07  | .08  | ----               | ---- | ---- |
| v23 cbc | .88                  | .88  | .77  | 0.94                   | 1.00 | 0.89 | .05            | .07  | .07  | ----               | ---- | ---- |
| v24 cbc | .87                  | .88  | .86  | 1.03                   | 1.04 | 1.05 | .06            | .08  | .07  | ----               | ---- | ---- |
| mc gsc  | ----                 | ---- | ---- | ----                   | ---- | ---- | ----           | ---- | ---- | .82                | .75  | .73  |
| mc tc   | ----                 | ---- | ---- | ----                   | ---- | ---- | ----           | ---- | ---- | .85                | .87  | .79  |
| mc cbc  | ----                 | ---- | ---- | ----                   | ---- | ---- | ----           | ---- | ---- | .79                | .85  | .84  |
| gsc tc  | ----                 | ---- | ---- | ----                   | ---- | ---- | ----           | ---- | ---- | .93                | .91  | .87  |
| gsc cbc | ----                 | ---- | ---- | ----                   | ---- | ---- | ----           | ---- | ---- | .74                | .76  | .74  |
| tc cbc  | ----                 | ---- | ---- | ----                   | ---- | ---- | ----           | ---- | ---- | .75                | .85  | .72  |

*Note.* v = variable; mc = motivation competence; gsc = game strategy competence; tc = technique competence; cbc = character building competence; \* = fixed; Bolded cells indicate significant ( $p < .001$ ) invariance between estimates.

### *How Reliable were the Rank Orderings of Coaching Competency Estimates?*

Reliability coefficients were .90 (MC), .87 (GSC), .85 (TC) and .82 (CB). These coefficients indicated very good to excellent levels of internal consistency for multidimensional coaching competency estimates.

### *Were Coaching Competency Estimates Positively Related to Satisfaction with the Coach Within Teams?*

Prior to answering this question, the assumption of unidimensionality for the set of satisfaction items was assessed. The Rasch component extracted at least one-half of the total variance in each of the items (range of extracted communalities = .49 to .86), produced an eigenvalue equal to 2.91, was reliably measured (range of loadings = .70 to .93), and accounted for 73% of the total variance. The eigenvalue for the next unaccepted component produced was equal to .49. The reliability of separation coefficient for the set of items was .75. Hence, treating these logit-based estimates as reliable measures of satisfaction was judged to be justifiable.

Next, correlations between logit-based coaching competencies and satisfaction with the coach scores were explored at the athlete-level and team-level (see Table 8). At both levels, all coaching competency subscales had at least a moderately high correlation with satisfaction with the coach (range = .61 to .85). At the athlete-level, MC had the highest correlation, .70, with satisfaction, with TC having the next highest correlation, .66. At the team-level, MC also had the highest correlation, .85, with satisfaction, with CBC having the next highest correlation, .82. Because MC was highly correlated to the other coaching competencies at both levels (range = .79 to .91), it was selected as the sole independent variable, at both levels, to avoid problems with multicollinearity.

Table 10

*Correlations Between Competency Judgments and Satisfaction with the Coach at the Individual and Team-Level*

|     | SWC   | MC          | GSC         | TC          | CBC         |
|-----|-------|-------------|-------------|-------------|-------------|
| SWC | ----- | <i>0.70</i> | <i>0.63</i> | <i>0.66</i> | <i>0.61</i> |
| MC  | 0.85  | -----       | <i>0.82</i> | <i>0.81</i> | <i>0.79</i> |
| GSC | 0.77  | 0.88        | -----       | <i>0.87</i> | <i>0.72</i> |
| TC  | 0.78  | 0.88        | 0.94        | -----       | <i>0.72</i> |
| CBC | 0.82  | 0.91        | 0.84        | 0.84        | -----       |

*Note.* Italicized entries in the upper diagonal are the athlete-level correlations and the team-level correlations are in the lower diagonal. All of the correlations were statistically significant at the .0005 level. SWC = satisfaction with coach; MC = motivation competence; GSC = game strategy competence; TC = technique competence; CBC = character building competence.

The ICC for satisfaction with the coach was .26, which indicated that 74% of the variance in satisfaction with the coach was due to within-team differences. Fit indices for Model 1 were  $G^2 = 1552.27$  and  $CAIC = 1574.38$ . The point reliability estimate for the team-level satisfaction means was .86. The variance of the team-level means,  $\sigma^2_{B0}$ , around the average team mean,  $\gamma_{00}$ , was statistically significant,  $\chi^2_{(31)} = 226.17, p < .001$  (see Table 9). Team-level means did not vary based on sport played. Thus, I concluded that satisfaction with the coach had a significant amount of variance both within teams and between teams; and, that team-level satisfaction means were significantly different and should remain random in subsequent models.

Athlete-level MC was added as a within-team predictor in Model 2 (see Table 9). This model fit better than Model 1,  $\chi^2_{LR(3)} = 315.83, p < .001$ ,  $CAIC_{diff} = 293.72$ , and explained 39% of the within-team variance in satisfaction with the coach. The average influence of MC was moderately large and positive,  $\gamma_{10} = .70$ , and statistically significant,  $t_{(31)} = 19.77, p < .001$ . The point reliability estimate for these slopes was .06. The variance of the team-level effects,  $\sigma^2_{B1}$ , around the average satisfaction-MC slope,  $\gamma_{10}$ , was not statistically significant,  $\chi^2_{(31)} = 35.33, p = .27$ . Thus, I concluded that the average influence of MC on satisfaction with the coach was moderately large and positive, had a somewhat low reliability, was similar within teams, and should be fixed across teams in a respecified model.

The influence of MC was fixed across teams in a respecified version of Model 2 (see Table 9). Conceptually, this model assumed that the influence of MC on satisfaction with the coach was similar within teams. The respecified model fit at least as well as the previous model,  $\chi^2_{LR(2)} = 1.43, p = .49$ ,  $CAIC_{diff} = 13.31$ . The influence of MC on

satisfaction remained moderately large and positive ( $\beta_{10} = .69$ ) and statistically significant,  $t_{(583)} = 20.83, p = .001$ . Thus, I concluded that MC had a moderately large and positive relationship with satisfaction with the coach across athletes.

*Were Coaching Competency Estimates Positively Related to Satisfaction with the Coach Between Teams?*

First, it should be noted that the variance in the team satisfaction means,  $\sigma^2_{B0}$ , around the average team satisfaction,  $\gamma_{00}$ , shrank from .26 to .05 after controlling for the influence of MC within teams (see Table 9). Practically, this means that there was much less variance in team satisfaction to model once the influence of MC on satisfaction was controlled for within teams. Still, because a statistically significant,  $\chi^2_{(31)} = 99.64, p < .0005$ , amount of variance remained among team satisfaction after imposing Model 2, team MC was added as a Level-2 predictor,  $\gamma_{01}$ . This model did not fit better than the respecified version of Model 2,  $\chi^2_{LR(1)} = 0.85, p < .36, CAIC_{diff} = -6.52$ , and did not explain additional between-team variance in team satisfaction. Accordingly, the influence of team MC on team satisfaction was negligible,  $\gamma_{01} = -.05$ , and not statistically significant,  $t_{(30)} = -0.92, p = .36$ . Hence, the variance of the team satisfaction estimates,  $\sigma^2_{B0}$ , around the average team satisfaction,  $\gamma_{00}$ , remained significantly greater than zero,  $\chi^2_{(30)} = 100.12, p = .01$ . Thus, I concluded that team MC was unrelated to team satisfaction after controlling for the effect of MC on satisfaction within teams.

Table 11

*Hierarchical Linear Models where Satisfaction was the Dependent Variable*

| Estimates of fixed effects                               |          |            |          |          |          |
|----------------------------------------------------------|----------|------------|----------|----------|----------|
| Fixed effect                                             | $\gamma$ | SE         | <i>t</i> | df       | <i>p</i> |
| <b>Model 1</b>                                           |          |            |          |          |          |
| Average satisfaction mean( $\gamma_{00}$ )               | -0.02    | 0.10       | -0.17    | 31       | .87      |
| <b>Model 2</b>                                           |          |            |          |          |          |
| Average satisfaction-motivation slope( $\gamma_{10}$ )   | 0.70     | 0.03       | 21.24    | 31       | < .01    |
| Fixed satisfaction-motivation slope( $\gamma_{10}$ )     | 0.69     | 0.03       | 20.98    | 583      | < .01    |
| <b>Model 3</b>                                           |          |            |          |          |          |
| Team satisfaction-team motivation slope( $\gamma_{01}$ ) | -0.05    | 0.05       | -0.92    | 30       | 0.36     |
| Estimates of variance components                         |          |            |          |          |          |
| Random effect                                            | $\sigma$ | $\sigma^2$ | df       | $\chi^2$ | <i>p</i> |
| <b>Model 1</b>                                           |          |            |          |          |          |
| Within team residuals( $r_{igs}$ )                       | 0.87     | 0.75       |          |          |          |
| Between team residuals( $u_{0gs}$ )                      | 0.51     | 0.26       | 31       | 226.17   | < .01    |
| <b>Model 2</b>                                           |          |            |          |          |          |
| Within team residuals( $r_{igs}$ )                       | 0.68     | 0.46       |          |          |          |
| Between team residuals( $u_{0gs}$ )                      | 0.23     | 0.05       | 31       | 99.63    | < .01    |
| Between team residuals( $u_{1gs}$ )                      | 0.05     | 0.01       | 31       | 35.33    | .27      |
| <b>Model 3</b>                                           |          |            |          |          |          |
| Within team residuals( $r_{igs}$ )                       | 0.68     | 0.46       |          |          |          |
| Between team residuals( $u_{0gs}$ )                      | 0.23     | 0.05       | 30       | 100.12   | < .01    |

Note.  $\sigma$  = standard deviation;  $\sigma^2$  = variance.

## CHAPTER 5

### DISCUSSION

This study provides initial validity evidence for the CCS, and introduces MCFA as an appropriate methodology to use when data are meaningfully nested and evaluation of the factor structure of a set of indicators is desired. Results offer some support for the proposed multilevel multidimensional conceptualization of coaching competency (see Figure 3), the internal consistency of the coaching competency estimates, and a relationship between motivation competency and satisfaction with the coach *within teams*. Validity concerns are observed for the original rating scale structure and the relationship between motivation competency and satisfaction with the coach *between teams*. Results are interpreted to guide future research with the CCS, to provide recommendations for revisions to the instrument, and to assist researchers in physical education and kinesiology in understanding when and how to apply MCFA to their data.

Analysis of the original rating scale structure (i.e., 10 categories) indicated that athletes were being asked to distinguish between too many levels of coaching competency (see Table 1). This finding is similar to previous research on the CES (Myers, Wolfe, et al., in press) and congruent with long-standing recommendations for Likert scales (Likert, 1932). However, an interesting difference between the Myers, Wolfe, et al. findings and this study, is that the previous study reported that coaches employed only four categories to judge the strength of their own coaching efficacy, whereas in this study athletes appeared to utilize five categories to judge their coach's competency. This difference was probably due to the fact that only 5% of the coaches' responses were in the 0-5 range, whereas 20% of the athletes' responses were within this

range. Athletes may make finer and more critical judgments of a coach's competency than coaches make regarding their own coaching efficacy, when responding to the CCS and CES items, respectively. However, to test this hypothesis, future research should match coaches' and athletes' responses within a sample, unlike the comparison I offer across two different samples.

While post-hoc analysis identified an improved 5-category rating scale structure, it is unknown whether the modified scale would prove optimal on a cross-validation sample of intercollegiate soccer and ice hockey athletes or with high school athletes. However, there is reason for confidence in the potential utility of the modified scale, with a similar sample, as Rasch-based optimal categorizations have been confirmed in follow-up applications (Zhu, 2002). Users of the CCS are encouraged to assess the utility of the proposed 5-category structure (i.e., complete incompetence, low, moderate, high, and complete competence). Although "low" may not be selected frequently, such a category would likely attract at least a minimum number (10) of observations necessary for minimal precision of threshold estimates (Linacre, 2002).

There is reason to believe that the CCS may adequately measure athletes' perceptions of their head coach's motivation, game strategy, technique, and character building competencies in lower division intercollegiate soccer and ice hockey programs (see Table 5 and Table 6). However, both within teams and between teams, there was limited discriminant validity among subscales-particularly between GSC and TC, within item multidimensionality for MC3, and extremely high standardized factor loadings at the between teams level. Allowing MC3 (*mentally prepare athletes for game strategies*), to load on both MC (*athletes' evaluations of their head coach's ability to affect the*

*psychological mood and skills of athletes*) and GSC (*athletes' evaluations of their head coach's ability to lead during competition*) makes sense conceptually. Because I judge the content of MC3 to be sufficiently specified and relevant to both MC and GSC, revision is not suggested. Therefore, users of the CCS are encouraged to impose internal models that allow responses to MC3 to influence scores on both MC and GSC.

Distinguishing between competence in leading during competition (GSC) and instructional and diagnostic competence (TC) makes sense conceptually. Empirically, evidence exists for the differential diagnostic ability of *game strategy efficacy* and *technique efficacy* in both high school and college coaches (Feltz et al., 1999; Myers, Vargas-Tonsing, et al., in press). Thus, I propose refining the definitions of GSC and TC and modifying items to lessen the overlap among the subscales in a revised version of the CCS. For example, altering the definition of TC to focus on evaluations of one's instructional and diagnostic skills *during practice*, may help to distinguish this competency from competency to lead *during competition*. However, until a revised CCS is available, the proposed internal model should be utilized to produce multidimensional measures of coaching competency from the existing items. But, because subscale scores are likely to be highly related and cause problems associated with multicollinearity when used as a set of independent variables, theory and bivariate correlations between the dependant variable(s) and competency scores should be explored to determine which competency measure should be used.

The extremely high magnitude of most of the standardized factor loadings at the between-teams level (see Table 6) was noted and was deemed to be similar to parallel loadings in previous research (Kaplan & Kreisman, 2000; Muthén, 1994). Possible

explanations for the observation of such loadings across studies: (1) overparameterization of the model (i.e., a possible artifact of MCFA modeling), (2) nearly perfect indicators of the latent variables at the between level, were posed in personal communications to experts in MCFA methodology. J. J. Hox (December 3, 2004) noted that he had also observed this trend and that he believe that possible explanations could include: (1) overparameterization; (2) that error on the between-level tends to be averaged out by the methodology; (3) that because the group-level variance is smaller, a higher percentage is able to be explained. D. Kaplan (December 3, 2004) noted that he too had observed this trend and that he was unsatisfied with the explanations to date. He proposed, as a remedy, modeling the factor structure of only  $S_{PW}$  (i.e., stopping at Step 3) if one is not planning to model the between group variation as a function of between group variables. In this study, however, determining a group-level factor structure for coaching competency estimates was important in order to guide formation of a team-level predictor of satisfaction with the coach (i.e., test aspects of the external model). However, confidence in the proposed between factor structure should be tempered pending additional, confirmatory research.

In terms of evidence relating to the external aspect of validity, I found that athletes' evaluations of their head coach's ability to affect the psychological mood and skills of athletes has a moderately large and positive relationship with satisfaction with the coach *within teams*. But, no such relationship is observed *between teams* after modeling the said relationship within teams (see Table 9). It is interesting to note, that if this relationship was modeled between teams without modeling the said relationship within teams, team MC is a significant predictor of team satisfaction. But, because the

majority of the variance (74%) in satisfaction is within teams and because Raudenbush & Bryk (2002) advocate for settling on a Level-1 model before specifying a Level-2 model (i.e., model-building from the bottom up), I conclude that team MC is not related to team satisfaction between teams. Methodologically, this finding demonstrates the need for examining multiple levels of variance simultaneously in order to reach more accurate conclusions and improve theory development (Silverman, 2004). Methodological considerations aside, it appears that athletes' satisfaction with the coach may frequently be linked to the coach's ability to attend to the athletes' psychological needs. Unfortunately, although most coaches are well equipped in the technical aspects of their sport, they rarely receive formal training in creating a healthy psychological environment (Smoll & Smith, 2001).

Because the usefulness of the CCS is at least partially dependent on the ability of resultant measures to demonstrate relationships specified in models of coaching effectiveness (Horn, 2002), more research is needed in this area. Specifically, studies that investigate how athletes' personal characteristics influence athletes' perceptions of their coach's competency, and/or how a coach's behavior influences athletes' perceptions of their coach's competency, and/or how athletes' competency judgments of their coach affects athletes' self-perceptions and beliefs, could advance understanding in coaching effectiveness and extend validity evidence for the CCS. Also, the ability of coaching education programs to alter athletes' perceptions of their coach's competency would be a very practical way to assess an important aspect of coaching education programs. This latter area, assessing coaching education programs, is becoming increasingly important in the current evidence-based educational system (Smoll & Smith, 2001).

Because the effects that a coach's behaviors exert on athletes are likely mediated by the meaning that athletes attribute to them (Horn, 2002; Smoll & Smith, 1989), the CCS has the potential to contribute to the improvement of coaching and the further development of coaching effectiveness models. Although the CCS has this potential, it should not be viewed as a competitor to instruments that assess other aspects of coaching. Rather, the CCS should be viewed as an additional tool that measures key constructs that are not fully covered by existing instruments. For example, if one were interested in examining relationships between observed coaching behavior and athletes' perceptions of their coach's negative activation, then using the CBAS and the CBQ would be appropriate. However, if this person was also interested in how negative activation was related to athletes' perceptions of their coach's motivation competency then administering the CCS would also be appropriate. In short, a single instrument cannot fully measure the wide range of constructs involved in effective coaching. Because effective coaching is complex, a variety of well-defined instruments is necessary to the scientific investigation of sufficiently targeted inquiries.

An introduction to MCFA was detailed earlier and encouragement toward appropriate future implementations is provided here. MCFA should be considered when subjects are meaningfully nested within groups and evaluation of the factor structure of a set of indicators is desired (Muthén, 1994). It is not uncommon in physical education, sport, and exercise contexts to collect data from subjects who are meaningfully nested within groups and to be interested in determining the factor structure of a set of indicators. A few examples include when data are collected within classes, teams, or exercise groups, and indicators are intended to measure subjects' perceptions and/or

performance (e.g., academic performance, collective efficacy, exercise behavior). In such cases, researchers should determine if MCFA is appropriate as detailed in the Introduction to MCFA section. If MCFA is appropriate and is applied, improved model-data fit will likely be observed, as compared to model-data fit on the total variance-covariance matrix, because levels of variance will be separated (Julian, 2001). Separating substantive levels of variance contributes to developing more accurate practical recommendations and to improved theory development (Silverman, 2004; Silverman & Solomon, 1998). I hope that the introduction will assist researchers in physical education and kinesiology in understanding when and how to apply MCFA to their data.

## **APPENDICES**

## **APPENDIX A**

**How competent is your head coach in his or her ability to---**

- 1. help athletes maintain confidence in themselves? (MC1)**
- 2. recognize opposing team's strengths during competition? (GSC2)**
- 3. mentally prepare his/her athletes for game strategies? (MC3)**
- 4. understand competitive strategies? (GSC4)**
- 5. instill an attitude of good moral character? (CBC5)**
- 6. build the self-esteem of his/her athletes? (MC6)**
- 7. demonstrate the skills of your sport? (TC7)**
- 8. adapt to different game situations? (GSC8)**
- 9. recognize opposing team's weakness during competition? (GSC9)**
- 10. motivate his/her athletes? (MC10)**
- 11. make critical decisions during competition? (GSC11)**
- 12. build team cohesion? (MC12)**
- 13. instill an attitude of fair play among his/her athletes? (CBC13)**
- 14. coach individual athletes on technique? (TC14)**
- 15. build the self-confidence of his/her athletes? (MC15)**
- 16. develop athletes' abilities? (TC16)**
- 17. maximize his/her team's strengths during competition? (GSC17)**
- 18. recognize talent in athletes? (TC18)**
- 19. promote good sportsmanship? (CBC19)**
- 20. detect skill errors? (TC20)**
- 21. adjust his/her game strategy to fit his/her team's talent? (GSC21)**
- 22. teach the skills of his/her sport? (TC22)**
- 23. build team confidence? (MC23)**
- 24. instill an attitude of respect for others? (CBC24)**

## REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238-246.
- Bentler, P. M. (2004). *EQS* (Version 6.1) [computer program]. Encino, CA: Multivariate Software, Inc.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345-370.
- Browne, M. W., & Cudek, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21, 230-258.
- Chelladurai, P. (1978). *A contingency model of leadership in athletics*. Unpublished doctoral dissertation. University of Waterloo, Canada.
- Chelladurai, P., & Arnott, M. (1985). Decision styles in coaching: Preferences of basketball players. *Research Quarterly for Exercise and Sport*, 56, 15-24.
- Chelladurai, P., & Riemer, H.A. (1997). A classification of facets of athlete satisfaction. *Journal of Sport Management*, 11, 133-159.
- Chelladurai, P., & Saleh, S. D. (1978). Preferred leadership in sport. *Canadian Journal of Applied Sport Sciences*, 3, 85-92.
- Chelladurai, P., & Saleh, S. D. (1980). Dimensions of leader behavior in sports: Development of a leadership scale. *Journal of Sport Psychology*, 2, 34-45.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.

- Cronbach, L. J. (1976). Research in classrooms and schools: formulations of questions, designs, and analysis. Occasional paper: Stanford Evaluation Consortium.
- Feltz, D. L., Chase, M. A., Moritz, S. E., & Sullivan, P. J. (1999). A conceptual model of coaching efficacy: Preliminary investigation and instrument development. *Journal of Educational Psychology*, 91, 765-776.
- Gould, D. (1987). Your role as a youth sport coach. In V. Seefeldt (Ed.), *Handbook for youth sport coaches* (pp. 17-32). Reston, VA: American Alliance for Health, Physical Education, Recreation, and Dance.
- Harnqvist, K. (1978). Primary mental abilities of collective and individual levels. *Journal of Educational Psychology*, 70, 706-716.
- Horn, T. S. (2002). Coaching effectiveness in the sports domain. In T.S. Horn (Ed.), *Advances in Sport Psychology* (pp. 309-354). Champaign, IL: Human Kinetics.
- Hox, J. J. (1993). Factor analysis of multilevel data: Gauging the Muthén model. In J. H. L. Oud & R. A. W. van Blokland-Vogeleesang (Eds.), *Advances in Longitudinal and Multivariate Analysis in the Behavioral Sciences* (pp. 141-156). Nijmegen, Netherlands: ITS.
- Hox, J. J. (2002). Multilevel factor models. In G.A. Marcoulides (Ed.), *Multilevel Analysis* (pp. 225-250). Mahwah, NJ: Lawrence Erlbaum.
- Hox, J. J., & Maas, C. J. M. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*, 8, 157-174.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jamshidian, M. & Bentler, P. M. (1999). ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics*, 24, 21-41.
- Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, 8, 325-352.
- Kaplan, D. (2000). Multilevel structural equation modeling. *Structural Equation Modeling: Foundations and Extensions* (pp. 130-148). Thousand Oaks, CA: Sage.
- Kaplan, D. & Kreisman, M. (2000). On the validation of indicators of mathematics

- education using TIMMS: An application of multilevel covariance structure modeling. *International Journal of Educational Policy, Research and Practice*, 1, 217-242.
- Kenow, L. J. & Williams, J. M. (1992). Relationship between anxiety, self-confidence, and evaluation of coaching behaviors. *The Sport Psychologist*, 6, 344-357.
- Kline, R.B. (1998). *Principles and practice of structural equation modeling*. New York: The Guilford Press.
- Lee, K. S., Maleté, L., & Feltz, D. L. (2002). The effect of a coaching education program on coaching efficacy. *International Journal of Applied Sport Sciences*, 14, 55-67.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1-55.
- Linacre, J. M. (1995). Categorical misfit statistics. *Rasch Measurement Transactions*, 9, 450-451.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3, 85-106.
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111, 490-504.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57, 519-530.
- McCullagh, P. & Nelder J. A. (1990). *Generalized linear models* (2<sup>nd</sup> ed.). Boca Raton, FL: CRC Press.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Mahwah, NJ, Lawrence Erlbaum Associates.
- McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods & Research*, 22, 399-413.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational Measurement* (3<sup>rd</sup> ed., pp. 13-103). New York: Macmillan.
- Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations: Presidential address to the psychometric society. *Psychometrika*, 54, 557-585.
- Muthén, B. O. (1990). *Mean and covariance structure analysis of hierarchical data*. Paper presented at the annual meeting of the Psychometric Society, Princeton, NJ.

- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, 22, 376-398.
- Muthén, B. O. (1997). Latent variable modeling of longitudinal and multilevel data. In A. E. Raftery (Ed.) *Sociological methodology 1997* (pp. 453-481). Washington, DC: American Sociological Association.
- Muthén, B. O., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of nonnormal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38, 171-189.
- Muthén, B. O., & Satorra, A. (1989). Multilevel analysis of varying parameters in structural models. In R. D. Bock (Ed.), *Multilevel Analysis of Educational Data* (pp. 87-99). San Diego, CA: Academic Press.
- Myers, N. D., Vargas-Tonsing, T. M., & Feltz, D. L. (in press). Coaching efficacy in intercollegiate coaches: Sources, coaching behavior, and team variables. *Psychology of Sport & Exercise*.
- Myers, N. D., Wolfe, E. W., & Feltz, D. L. (in press). An evaluation of the psychometric properties of the coaching efficacy scale for American coaches. *Measurement in Physical Education and Exercise Science*.
- National Association for Sport and Physical Education. (1995). *Quality coaches, quality sports: National standards for athletic coaches*. Dubuque, IA: Kendall/Hunt.
- Park, J. K. (1992). *Construction of the Coaching Confidence Scale*. Unpublished doctoral dissertation. Michigan State University, East Lansing.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Raudenbush, S. W., & Bryk, A. S., Cheong, Y. F., & Congdon, R. (2000). *HLM5: Hierarchical and linear modeling*. Homewood, IL: Scientific Software International.
- Rushall, B. S., & Wiznuk, K. (1985). Athletes' assessment of the coach: The coach evaluation questionnaire. *Canadian Journal of Applied Sport Sciences*, 10, 157-161.
- Satorra, A., & Bentler, P. M. (1990). Model conditions for asymptotic robustness in the analysis of linear relations. *Computational Statistics & Data Analysis*, 10, 235-249.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors on covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent Variable Analysis* (pp. 399-419). Thousand Oaks, CA: Sage.

- Silverman, S. (2004). Analyzing data from field research: The unit of analysis issue. *Research Quarterly for Exercise and Sport*, 75, iii-iv.
- Silverman, S., & Solomon, M. (1998). The unit of analysis in field research: Issues and approaches to design and data analysis. *Journal of Teaching in Physical Education*, 17, 270-284.
- Silvey, S. D. (1959). The Lagrangian multiplier test. *Annals of Mathematical Statistics*, 30, 389-407.
- Smith, R. E., Smoll, F. L., & Curtis, B. (1978). Coaching behaviors in little league baseball. In F. L. Smoll & R. E. Smith (Eds.), *Psychological perspectives in youth sports* (pp. 173-201). Washington, DC: Hemisphere.
- Smith, R. E., Smoll, F. L., & Hunt, E. B. (1977). A system for the behavioral assessment of athletic coaches. *Research Quarterly*, 48, 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, 1522-1551.
- Smoll, F. L. & Smith, R. E. (2001). Conducting sport psychology training programs for coaches: Cognitive-behavioral principles and techniques. In J.M. Williams (Ed.), *Applied Sport Psychology* (pp. 378-400).
- Sullivan, P. J., & Kent, A. (2003). Coaching efficacy as a predictor of leadership style in intercollegiate athletics. *Journal of Applied Sport Psychology*, 15, 1-11.
- Tabachnick, B. G., & Fidell, L. S. (2001). Cleaning up your act. In *Using multivariate statistics* (4th ed., pp. 56-110). Boston: Allyn & Bacon.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1-10.
- Vargas-Tonsing, T. M., Warners, A. L., & Feltz, D. L. (2003). The predictability of coaching efficacy on team efficacy and player efficacy in volleyball. *Journal of Sport Behavior*, 26, 396-407.
- Wald, A. (1943). Tests of statistical hypothesis concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426-482.
- Wicherts, J. M., & Dolan, C. V. (2004). A cautionary note on the use of information fit indices in covariance structure modeling with means. *Structural Equation Modeling*, 11, 45-50.

- Williams, J. M., Jerome, G. J., Kenow, L. J., Rogers, T., Sartain, T. A., & Darland, G. (2003). Factor structure of the coaching behavior questionnaire and its relationship to athlete variables. *The Sport Psychologist, 17*, 16-34.
- Wright, B. D., & Linacre, J. M. (1992). Combining and splitting categories. *Rasch Measurement Transactions, 6*, 233-235.
- Wright, B. D., & Linacre, J. M. (1998). *Winsteps: A Rasch model computer program*. Chicago: MESA Press.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: Mesa Press.
- Wu, M. L., Adams, R. J., & Wilson, M. R. (1998). *ACER ConQuest: Generalized item response modeling software* (Version 1.0) [computer program]. Melbourne, Victoria, Australia: Australian Council for Educational Research.
- Zhu, W. (2002). A confirmatory study of Rasch-based optimal categorization of a rating scale. *Journal of Applied Measurement, 3*(1), 1-15.
- Zhu, W., Updyke, W. F., & Lewandowski, C. (1997). Post-hoc Rasch analysis of optimal categorization of an ordered-response scale. *Journal of Outcome Measurement, 1*, 286-304.

## FOOTNOTES

<sup>1</sup> The term *coach's behavior* is used to be consistent with Horn's (2002) model. It is noted that no instrument can fully and completely represent the wide range of behaviors involved in effective coaching.

<sup>2</sup> Analyses for questions 4 and 5 in this study provide examples of multilevel extensions of the conventional multiple regression model.

<sup>3</sup> In this study only two levels of variance, within teams and between teams, was considered.

<sup>4</sup> It is noted that multidimensional measures of athlete satisfaction have been suggested (Chelladurai & Reimer, 1997). Such measures are suggested to be used as indicators of overall organizational effectiveness. In this study, because satisfaction with the coach, not the entire organization was of interest, our measure was considered appropriately specific.

<sup>5</sup> To the authors' knowledge, IRT software programs, including *Winsteps*, are currently limited to evaluating rating scale structures for unidimensional models.

<sup>6</sup> Person and item separation indices are expected to decrease somewhat when categories are collapsed because, in general, the more categories there are the better the discrimination (Zhu et al., 1997).

<sup>7</sup> Due the complexity of MUML estimation in unbalanced groups, it is not uncommon for programs to require starting values to converge to a stable solution (Hox, 2002).

MICHIGAN STATE UNIVERSITY LIBRARIES



3 1293 02736 1132