

Published in final edited form as:

*Psychol Sport Exerc.* 2013 September ; 14(5): 709–718. doi:10.1016/j.psychsport.2013.04.008.

## Coaching Competency and (Exploratory) Structural Equation Modeling: A Substantive-Methodological Synergy

Nicholas D. Myers

Department of Educational and Psychological Studies, University of Miami.

### Abstract

**Objectives**—The purpose of this manuscript was to provide a substantive-methodological synergy of potential importance to future research in the psychology of sport and exercise.

**Design**—The substantive focus was the emerging role for, and particularly the measurement of, athletes' evaluations of their coach's competency within conceptual models of effective coaching. The methodological focus was exploratory structural equation modeling (ESEM), a methodology that integrates the advantages of exploratory factor analysis and confirmatory factor analysis (CFA) within the general structural equation model.

**Method**—The synergy was a demonstration of when a new and flexible methodological framework, ESEM, may be preferred as compared to a more familiar and restrictive methodological framework, CFA, by reanalyzing existing data.

**Results**—ESEM analysis on extant datasets suggested that for responses to the Athletes' Perceptions of Coaching Competency Scale II – High School Teams (APCCS II-HST), a CFA model based on the relevant literature plus one post hoc modification, offered a viable alternative to a more complex ESEM model. For responses to the Coaching Competency Scale (CCS), a CFA model based on the relevant literature did not offer a viable alternative to a more complex ESEM model.

**Conclusions**—The ESEM framework should be strongly considered in subsequent studies validity studies - for new and/or existing instruments in the psychology of sport and exercise. A key consideration for deciding between ESEM and the accompanying rotation criterion and CFA in future validity studies should be level of a priori measurement theory.

### Keywords

athlete's evaluations; coaching effectiveness; geomin; target; categorical data

---

Factor analysis has been closely linked with investigations of construct validity in psychology for several decades (Nunnally, 1978). Investigations of construct validity (e.g., scale development) in exercise and sport have frequently occurred in studies where only factor analytic measurement model(s), exploratory and/or confirmatory, were specified – typically guided by important, yet incomplete, a priori substantive measurement theory

---

© 2013 Elsevier Ltd. All rights reserved.

Correspondence concerning this article should be addressed to Nicholas D. Myers, Department of Educational and Psychological Studies, University of Miami, Merrick Building 311E, Coral Gables, FL, USA, 33124-2040. nmyers@miami.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

(Myers, Ahn, & Jin, 2011a). Incomplete substantive measurement theory often manifests as model error and offers a possible explanation as to why the majority of construct validity studies in psychology fail the test of exact fit under a strictly confirmatory approach (Jackson, Gillaspay, & Purc-Stephenson, 2009).<sup>1</sup> Construct validity studies that incorporate an exploratory methodological approach typically reduce model error, as compared to a confirmatory approach, but may fail to adequately incorporate a priori substantive measurement theory (Asparouhov & Muthén, 2009).

What constitutes a sufficient degree of a priori substantive measurement theory, and therefore should lead to a preference for a more restrictive confirmatory approach versus a more flexible exploratory approach, is an emerging area of methodological research of potential importance to substantive research in the psychology of sport and exercise. Simulation research suggests that a key consideration for deciding between exploratory structural equation modeling (ESEM; Asparouhov & Muthén, 2009) and confirmatory factor analysis (CFA; Jöreskog, 1969) in future validity studies should be level of a priori substantive measurement theory (Myers, Ahn, & Jin, 2011b). In the Myers et al. simulation study model-data fit was manipulated (as a proxy for level of a priori substantive measurement theory) with small but non-zero values in both the pattern coefficient matrix and in off-diagonals of the residual covariance matrix. Under exact (with ESEM having errors of inclusion) and close (with both ESEM and CFA having errors of exclusion) model-data fit, CFA generally performed at least as well as ESEM in terms of parameter bias, and was cautiously recommended when a priori substantive measurement theory can be viewed as at least close. Under only approximate model-data fit, on balance, ESEM performed at least as well as CFA and was cautiously recommended when a priori substantive measurement theory can be viewed as only approximate.

This invited manuscript demonstrates by way of an extant conceptual model (i.e., coaching competency) when a new and flexible methodological framework, ESEM, may be preferred as compared to a more familiar and restrictive methodological framework, CFA, within a substantive-methodological synergy format (Marsh & Hau, 2007). For spatial reasons this manuscript focuses primarily on the synergy by way of relevant empirical examples. Fuller reviews of both the relevant substance (Myers & Jin, 2013) and technical aspects of the broader methodology (e.g., Asparouhov & Muthén, 2009) are available elsewhere. From this point forward the acronym, (E)SEM, is used when referring to both ESEM and CFA simultaneously.

## Coaching Competency: The Substance

The Coaching Competency Scale (CCS; Myers, Feltz, Maier, Wolfe, & Reckase, 2006) was derived via minor changes to the Coaching Efficacy Scale (CES; Feltz, Chase, Moritz, & Sullivan, 1999). The CES 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 key expected competency domains articulated in the *National Standards for Athletic Coaches* (National Association for Sport and Physical Education, 2006) and are congruent with self-efficacy theory (Bandura, 1997). According to Horn's (2002) model of coaching effectiveness, however, *why* a coach's beliefs (e.g., coaching efficacy) are related to athletes' self-perceptions and performance is due to the influence that these beliefs exert on a coach's behavior. But, the influence that a coach's behavior exerts on athletes' self-

<sup>1</sup>Other possible explanations exist for failing the likelihood ratio test of exact fit (e.g., Type I error). The degree to which failing the exact fit test is regarded as a meaningful occurrence (e.g., an imperfect model may be useful) is an ongoing debate (e.g., Vernon & Eysenck, 2007).

perceptions, motivation, and performances is mediated, at least in part, by athletes' evaluations of their coach's behavior. The purpose of the CCS, therefore, was to measure *athletes' evaluations* of their head coach's ability to affect the learning and performance of athletes. There is evidence that measures derived from both the CCS relate to theoretically relevant variables (e.g., Bosselut, Heuzé, Eys, Fontayne, & Sarrazin, 2012).

The measurement model for the CCS posited that four latent variables covary and influence responses to 24 items. From this point forward the terms competency and efficacy were generally not included alongside any particular dimension of coaching competency. *Motivation* was defined as athletes' evaluations of their head coach's ability to affect the psychological mood and skills of athletes. *Game strategy* was defined as athletes' evaluations of their head coach's ability to lead during competition. *Technique* was defined as athletes' evaluations of their head coach's instructional and diagnostic abilities. *Character building* was defined as athletes' evaluations of their head coach's ability to influence the personal development and positive attitude toward sport in their athletes. Myers, Feltz et al. (2006) referred to this multidimensional construct as *coaching competency*. High school and lower division collegiate athletes of team sports comprise populations for which the CCS was intended.

Model-data fit within the CFA framework for the CCS (as well as the CES - see Myers & Jin, 2013) has generally not met heuristic values for close fit (e.g., Hu & Bentler, 1999). Myers, Feltz, et al. (2006) reported the following model-data fit for non-division I collegiate athletes ( $N = 585$ ) clustered within men's ( $g = 8$ ) and women's ( $g = 24$ ) teams:  $\chi^2(246) = 1266$ , CFI = .91, Tucker-Lewis index (TLI) = .90, and RMSEA = .09. Post hoc modifications by Myers, Feltz, et al. included allowing a motivation item to also indicate game strategy and freeing the covariance between two pairs of residual variances (see Figure 1). Interfactor correlations in the Myers, Feltz et al. study ranged from  $r_{\text{technique, character building}} = .80$  to  $r_{\text{game strategy, technique}} = .92$ . Given the high interfactor correlations, Myers, Feltz, et al. explored models with fewer latent variables, all of which yielded statistically significant worse fit.

Responses to the CCS have a history of dependence based on athletes nested within teams. Myers, Feltz, et al. (2006) reported item-level intraclass correlation coefficients (ICC) ranging from .22 to .44,  $M = .32$ ,  $SD = .05$ . Multilevel CFA is an appropriate methodology when data violate the assumption of independence (Muthén, 1994). Simulation research has indicated, however, that a relatively large number of groups (~100) may be necessary for optimal estimation, particularly at the between-groups level (Hox & Maas, 2001). When group-level sample size is not large, imposing a single-level model on the within-groups covariance matrix,  $\mathbf{S}_W$ , as opposed to the total covariance matrix,  $\mathbf{S}_T$ , controls for probable biases in fitting single-level models to multilevel data (Julian, 2001). Myers, Feltz, et al. modeled the  $\mathbf{S}_W$ .

The Coaching Efficacy Scale II-HST (CES II-HST) was derived via major changes to the CES as detailed in Myers, Feltz, Chase, Reckase, and Hancock (2008) and Myers, Feltz, and Wolfe (2008). The Athletes' Perceptions of Coaching Competency Scale II – High School Teams (APCCS II-HST) was derived via minor changes to the CES II-HST as detailed in Myers, Chase, Beauchamp, and Jackson (2010). Thus, moving from either CES to the CES II-HST, or the CCS to the APCCS II-HST, resulted in non-equivalent measurement models. There is evidence that measures derived from the APCCS II-HST relate to theoretically relevant variables (Myers, Beauchamp, & Chase, 2011) in a doubly latent multilevel model (Marsh et al., 2012).

The measurement model for the APCCS II-HST occurred at two-levels in Myers et al. (2010) because there was strong evidence for dependence due to the clustering of athletes ( $N = 748$ ) within teams ( $G = 74$ ). Specifically, ICCs ranged from .18 to .35,  $M = .29$ ,  $SD = .05$ .

There was evidence for close fit for the multilevel measurement model:  $\chi^2_{(201)}$  ranged from 375,  $p < .001$ , to 405,  $p < .001$ ,  $RMSEA = .04$ ,  $CFI = .99$ ,  $TLI = .99$ ,  $SRMR_{within} = .03$ , and  $SRMR_{between} = .04$ .<sup>2</sup> Evidence also was provided for both the stability of all latent variables and for the reliability of composite scores. There is evidence, at least in regard to a crude comparison of model-data fit, that the measurement model underlying responses to the APCCS II-HST may be better understood (by researchers) than the measurement model underlying responses to the CCS. While Myers et al. took a multilevel approach, the focus of this manuscript was on the athlete-level (Level 1) model only.

The measurement model for the APCCS II-HST specified that five dimensions of coaching competency covary and influence responses to 17 indicators. *Motivation* (M) was measured by four items and was defined as athletes' perceptions of their head coach's ability to affect the psychological mood and psychological skills of her/his athletes. *Game strategy* (GS) was measured by five items and was defined as athletes' perceptions of their head coach's ability to lead his/her athletes during competition. *Technique* (T) was measured by four items and was defined as athletes' perceptions of their head coach's ability to utilize her/his instructional and diagnostic skills during practices. *Character building* (CB) was measured by three items and was defined as athletes' perceptions of their head coach's ability to positively influence the character development of his/her athletes through sport. *Physical conditioning* (PC) was measured with two items and was defined as athletes' perceptions of their head coach's ability to prepare her/his athletes physically for participation in her/his sport. Figure 2 depicts the athlete-level measurement model put forth by Myers et al. (2010) - less the bolded arrow.<sup>3</sup>

There is growing evidence for the importance of coaching competency in coaching effectiveness models. An important foundational question for this emerging line of research going forward is: how best to derive measures from responses to the CCS and APCCS II-HST? In cases where the population of interest is high school team sports the APCCS II-HST, and its proposed measurement model, seems appropriate. In cases where the population of interest is not high school team sports the CCS, and its less well understood measurement model, may be appropriate. While the measurement model for CCS appears to be understood at a level below typical close-fitting CFA standards, it also seems to be understood better than imposition of a mechanical EFA model would imply. Perhaps a framework situated between traditional EFA and traditional CFA would be appropriate for deriving measures from responses to the CCS.

## (Exploratory) Structural Equation Modeling: The Methodologies

Asparouhov and Muthén (2009) introduced ESEM as a methodology that integrates the relative advantages of both EFA and CFA within the general structural equation model (SEM). A technical paper (Asparouhov & Muthén) and a substantive-methodological synergy (Marsh et al., 2009) introduced ESEM to the methodological community. The general ESEM allows for parameterizations beyond EFA, but given the measurement focus

<sup>2</sup>Many of the model-data fit indices reported in Myers et al. (2010) summarized model-data fit across levels. Future research in this area should consider providing level-specific indices put forth by Ryu and West (2009):  $\chi^2_{PS-W}$ ,  $\chi^2_{PS-B}$ ,  $RMSEA_{PS-W}$ ,  $RMSEA_{PS-B}$ ,  $CFI_{PS-W}$ , and  $CFI_{PS-B}$ .

<sup>3</sup>The non-bolded cross-loading, and a rationale for it, was provided in Myers et al. (2010). The bolded cross-loading was not specified in Myers et al. A rationale for its inclusion in the current manuscript is provided shortly in the deriving measures from the APCCS II-HST section.

of this manuscript, only reduced versions of ESEM (i.e., EFA and Multiple-Group EFA in ESEM) are reviewed.<sup>4</sup> Prior to introducing the reduced equation below, however, a brief history is provided because this history may inform how researchers have decided between EFA and CFA over the last few decades.

Over a century ago Spearman (1904) articulated what has become to be known as EFA. Several decades after this articulation, EFA became a widely used multivariate analytic framework in many disciplines. Over the past few decades, however, some limitations in the way EFA is typically implemented in software, in addition to rotational indeterminacy more generally, has likely impeded the use of the technique in favor of CFA - even when the level of a priori measurement theory was insufficient to warrant a confirmatory approach.<sup>5</sup> Examples of limitations in the way that EFA has typically been implemented in software include the absence of standard errors for parameter estimates, restrictions on the ability to incorporate a priori content knowledge into the measurement model, an inability to fully test factorial invariance, and an inability to simultaneously estimate the measurement model within a fuller structural model.

The most commonly used measurement model within SEM likely is Jöreskog's (1969) CFA (Asparouhov & Muthén, 2009). CFA provides standard errors for parameter estimates, allows a priori content knowledge to guide model specification, provides a rigorous framework to test factorial invariance, and allows the measurement model to be a part of a fuller structural equation model. An often overlooked but necessary condition for appropriate use of CFA, however, is a sufficient level of a priori measurement theory. Absence of such theory often results in an extensive post hoc exploratory approach guided by modification indexes. Such an approach is susceptible to producing an accepted model that is inconsistent with the true model – despite possible consistency with a particular dataset (MacCallum, Roznowski, & Necowitz, 1992). EFA, not CFA, may often be the better framework for post hoc explorations in search of a well-fitting measurement model (Browne, 2001).

A common misspecification in CFA is when nonzero paths from latent variables to measurement indicators are fixed to zero. This type of misspecification often results in upwardly biased covariances between the latent variables and biased estimates in the non-measurement part of the structural equation (Asparouhov & Muthén, 2009). Both of these problems can, and probably have, impeded research in sport and exercise psychology. An example of upwardly biased covariances between latent variables may have occurred in Myers, Feltz, et al. (2006). It is impossible, however, to know for certain that biased (upwardly or downwardly) parameter estimates are observed in any case where the population parameter values are unknown.

In *Mplus* (Muthén & Muthén, 1998-2010), the ESEM approach integrates the relative advantages of EFA and CFA within SEM. Like EFA, ESEM imposes fewer restrictions on the measurement model than common implementations of CFA and integrates advances in possible direct rotations of the pattern matrix. Like CFA, ESEM can be part of a general SEM which affords much greater modeling flexibility than observed in the traditional EFA framework. In fact, within a single SEM specification, both CFA and ESEM can be imposed simultaneously. This latter point is of special interest because when sufficient a priori measurement theory exists, the more parsimonious CFA is preferred (Asparouhov &

---

<sup>4</sup>Readers are referred to Marsh et al. (2009) and Morin, Marsh, and Nagengast (2013) for ESEM examples that are not EFA or Multiple-Group EFA in ESEM.

<sup>5</sup>Under some factor structures a rotation criterion may yield multiple local minima (see Browne, 2001). Encountering multiple “optimal” rotated solutions may be regarded as an opportunity for to find the “best” factor structure from a conceptual perspective (e.g., Rozeboom, 1992).

Muthén, 2009). An important methodological question within the psychology of sport and exercise research going forward is: What constitutes a sufficient degree of a priori substantive measurement theory, and therefore should lead to a preference for a more restrictive (i.e., parametrically parsimonious) CFA approach versus a more flexible (i.e., parametrically non-parsimonious) EFA approach?

### EFA in ESEM

EFA in ESEM can be written with two general equations: one for the measurement model and one for the latent variable model (Bollen, 1989). Given the focus of this manuscript neither equation will be provided here but a review of each equation is available in Appendix A of the supplemental electronic-only materials. It is important to note, however, that as compared to ESEM, the form of the measurement equation is identical for CFA but more restrictions typically are placed on it by the researcher (e.g., many elements within the pattern matrix,  $\Lambda$ , are fixed to 0). Thus, a primary difference between EFA in ESEM and CFA is the restrictions placed on  $\Lambda$ .

**Rotation**—Direct analytic rotation of the pattern matrix is based on several decades of research within the EFA framework as detailed in Browne (2001) and summarized in Appendix B\_1 of the supplemental electronic-only materials. Geomin rotation (Yates, 1987) is the default rotation criterion in *Mplus* and, therefore, may be used frequently in practice. Geomin, however, “...fails for more complicated loading matrix structures involving three or more factors and variables with complexity 3 and more;...For more complicated examples the Target rotation criterion will lead to better results” (Asparouhov & Muthén, p. 407). Target rotation has been developed over several decades and can be thought of as “a non-mechanical exploratory process, guided by human judgment” (Browne, p.125). It is important to note that a solution will be mathematically equivalent under either target or geomin rotation. Simulation research suggests, however, that in some cases the factors may be defined more consistent with a well-developed a priori theory under target rotation as compared to geomin rotation (Myers, Jin, & Ahn, 2012).

### Multiple-Group EFA in ESEM

Measurement invariance is necessary to establish that measures are comparable across groups for which inferences are sought (Millsap, 2011). Providing evidence for measurement invariance across groups for which an instrument is intended is an important facet of validity evidence (AERA, APA, NCME, 1999). Until very recently, however, fully testing for measurement invariance across groups forced researchers to use the CFA framework (Asparouhov & Muthén, 2009). Multiple-group EFA in ESEM allows researchers to fully test for measurement invariance across groups within an exploratory framework.

## Measurement of Coaching Competency within (E)SEM: The Synergy

### EFA in ESEM

An important question for the emerging line of research involving coaching competency within coaching effectiveness models going forward is: how best to derive measures from responses to the CCS and APCCS II-HST? An important methodological question within effective coaching research going forward is: how can researchers apply (E)SEM to evaluate models of effective coaching? The purpose of this section was to attempt to address both questions by demonstrating how (E)SEM may be used guided by content knowledge to develop better, or confirm existing, measurement models for coaching competency.

Two primary research questions drove subsequent analyses on extant datasets. Extant datasets were used for both demonstration purposes and to see if the new ESEM approach offered benefits not available when each dataset was originally analyzed. The first question was: how many factors were warranted to explain responses to the APCC II-HST/CCS? The first question was answered in two steps. Step 1, *number of factors* ( $m$ ), considered the fit of a particular ESEM model. Step 2,  $m$ , considered the relative fit of a ESEM model as compared to a more complex alternative.<sup>6</sup> Thus,  $m$  denoted number of factors and not necessarily the number of factors consistent with substantive a priori measurement theory.

Nested models were compared with the change in the likelihood ratio  $\chi^2$  (robust) test,  $\Delta\chi_R^2$ , with the DIFFTEST command (Asparouhov & Muthén, 2010).<sup>7</sup> The approach taken in Step 2 was susceptible to over-factoring and inflated Type I error (Hayashi, Bentler, & Yuan, 2007). Therefore  $\alpha$  was set to .01 for these comparisons (.05 otherwise) and the interpretability of the estimated rotated pattern matrix  $\widehat{\Lambda}^*$  also was considered (i.e., incorporating content knowledge) when deciding which model to accept. The second question was: could a more parsimonious CFA model, based on the relevant literature, offer a viable alternative to a more complex ESEM model? If the answer to Question 2 was no, a follow-up question was: which rotation method was preferred? Consistent with weaknesses of the  $\Delta\chi_R^2$  with real data and imperfect theories (Yuan & Bentler, 2004) and the utility of strict adherence to the null hypothesis testing framework with regard to the assessment of model-data fit (Marsh et al., 2004), a collection of fit indices were considered.

**Methods**—APCCS II-HST data from Myers et al. (2010) were re-analyzed. Athlete-level sample size was 748 and team-level sample size was 74. Seven relevant sports were represented. Girls' ( $n = 321$ ,  $g = 35$ ) and boys' ( $n = 427$ ,  $g = 39$ ) teams were included. Age data ranged from 12 to 18 years ( $M = 15.46$ ,  $SD = 1.51$ ).

CCS data from Myers, Feltz, et al. (2006) were reanalyzed. Athlete-level sample size was 585 and team-level sample size was 32. Data were collected from lower division intercollegiate men's ( $g = 8$ ) and women's ( $g = 13$ ) soccer and women's ice hockey teams ( $g = 11$ ). Age data ranged from 19 to 23 years ( $M = 19.53$ ,  $SD = 1.34$ ).

At the time of the writing of this manuscript ESEM could not be fitted within a multilevel framework in Mplus (L. Muthén, personal communication, January 11, 2013). The multilevel nature of the data was handled via the TYPE = COMPLEX command. This resulted in modeling  $\mathbf{S}_T$  with non-independence accounted for in the computation of standard errors and in relevant model-data fit indices and parallels an approach taken in Asparouhov and Muthén (2009, p. 412). Models were not fit to  $\mathbf{S}_W$  because summary data are not allowed in Mplus for ESEM (T. Nguyen, personal communication, February 22, 2013).

APCCS II-HST data were treated as ordinal due to the four-category rating scale structure (Finney & Distefano, 2006). Weighted least squares mean- and variance-adjusted (WLSMV; Muthén, 1984) estimation was used.<sup>8</sup> CCS data were treated as continuous and normal, with a correction for non-normality, due to the original ten-category rating scale structure. Geomin rotation ( $\epsilon$  value varied based on  $m$ : .0001 when  $m = 2$ , .001 when  $m = 3$ , .01 when  $m = 4$ ) was used for Question 1. Target rotation was used for Question 2. Key input files for

<sup>6</sup>This approach, “comparison\_”*ms*”, to assisting in the determination of the number of factors to retain is consistent with an approach taken in Asparouhov and Muthén (2009). Other reasonable approaches include parallel analysis (Horn, 1965; Garrido, Abad, & Ponsoda, 2012; Morin et al., 2013) or extracting a fixed number of factors based on theory, “theory\_”*m*”, (Marsh et al., 2009).

<sup>7</sup>Details regarding the approximations of the test statistic to the central chi-square distribution are provided in Asparouhov and Muthén (2010).

subsequent examples are available in Appendix E of the supplemental electronic-only materials.

**Deriving Measures from the APCCS II-HST: Question 1**—The null hypothesis for exact fit was rejected until six factors were extracted. For fuller results see Table 1s in Appendix C of the supplemental electronic-only materials. From this point forward a Table/Figure number denoted with an “s” suggests that the Table/Figure is located in the supplemental electronic-only materials. The sixth factor of  $\widehat{\Lambda}^*$ , however, was judged as not interpretable and likely the result of over-factoring as only three elements (m1, t4, c3) in this column were statistically significant and the interfactor correlations for this factor ranged from .07 to .17 – each of which was statistically non-significant. The five factor solution (i.e., Model 5) was accepted despite exhibiting statistically significant worse fit than the sixth factor solution,  $\Delta\chi_R^2(12) = 31, p = .002$ . There was evidence for close fit of Model 5:  $\chi_R^2(61) = 88, p = .013$ . RMSEA = .02, CFI = 1.00, and TLI = .99. Elements within  $\widehat{\Lambda}^*$  were generally consistent with expectations (compare Table 1 to Figure 2). Interafactor correlations ranged from  $r_{PC,CB} = .54$  to  $r_{GS,T} = .76$ . Only 4 of the 66 cross-loadings that were estimated in the ESEM approach, and inconsistent with Figure 2 (e.g., m1 on GS), were statistically significant (standardized loadings ranged from .24 to .27). A reasonable answer to Question 1 was: five factors explained responses to the APCC II-HST.

**Deriving Measures from the APCCS II-HST: Question 2**—Target rotation was then applied to Model 5 (i.e., Model 7). Selection of targets for the APCCS II-HST will be briefly and non-technically summarized here (but a fuller review of related issues, including rotation identification more generally, is available in Appendix B\_1 of the supplemental electronic-only materials). Targets within **B** were selected based on both content knowledge and previous empirical information. From a content perspective, gs2, make effective strategic decisions in pressure situations during competition, was expected to have trivially sized pattern coefficients on all factors except game strategy within the measurement model. Each of the other items that were involved in the targets (pc3, *accurately assess players' physical conditioning*; m4, motivate players for competition against a weak opponent; t1, teach players the complex technical skills of your sport during practice; cb3, *effectively promote good sportsmanship in players*) could be interpreted in a similar form. Note that unlike in CFA, the pattern targets were not *fixed* to zero. Rather, **B** simply informed the rotation criterion as to the structure that **A** should be rotated towards. It should be noted that although a sufficient number and configuration of targets were specified, simulation research suggests that the factors may be defined more consistent with a well-developed a priori theory with an increasing number of targets (Myers, Ahn, & Jin, 2013).

The measurement model depicted in Figure 2 (i.e., Model 8) exhibited statistically significant worse fit than Model 7,  $\Delta\chi_R^2(47) = 75, p = .006$ . Model 8 estimated 47 fewer parameters than Model 7. For this reason, in addition to the relatively similar fit of the two models, limited and conceptually defensible post hoc alterations to Model 8 were considered. One defensible post hoc alteration allowed g3, make effective player substitutions during competition, to load on M in addition to GS (Model 9). Simply, making a player substitution during competition can serve as a motivational intervention in addition to a tactical strategy. Model 8 exhibited statistically significant worse fit than Model 9,  $\Delta\chi_R^2(1) = 8, p = .005$ . Model 9 exhibited only marginally statistically significant worse fit than Model 7,  $\Delta\chi_R^2(46) = 67, p = .024$ . None of the modification indices for Model 9 suggested statistical significance if freely estimated and the model exhibited close fit to the data:  $\chi_R^2(107) = 146, p = .008$ . RMSEA = .02, CFI = .99, and TLI = .99. For these reasons, in



addition to estimating 46 fewer parameters than Model 7, Model 9 was accepted (see Table 2). Given the narrow scope of this approach to post hoc modification of the a priori measurement model, a reasonable answer to Question 2 was: a more parsimonious CFA model, based on the relevant literature plus one post hoc modification, offered a viable alternative to a much more complex ESEM model. For didactic considerations, however, the ESEM approach from Model 5 will be expanded upon in the subsequent multiple-group EFA section. Table 2s in Appendix C provides interfactor correlation matrices for Model 5, Model 7, and Model 9. Interfactor correlations were somewhat similar across the three models.

**Deriving Measures from the CCS: Question 1**—The null hypothesis for exact fit was rejected through seven factor extraction (for fuller results see Table 3s in Appendix C). A model with eight factors produced inadmissible estimates. Model 5 (five factors) through Model 7 (seven factors), however, were judged as not interpretable and likely the result of over-factoring due to reasons similar to those in the previous section. The four factor solution (i.e., Model 4) was accepted despite exhibiting statistically significant worse fit than the five factor solution,  $\Delta\chi^2_R(20) = 213, p < .001$ . There was evidence for approximate fit of Model 4:  $\chi^2_R(184) = 617, p < .001$ . RMSEA = .06, SRMR = .02, CFI = .99, and TLI = .99. Elements within  $\widehat{\Lambda}^*$  were somewhat inconsistent with expectations (compare Table 3 to Figure 1). Note that 15 of the 71 cross-loadings that were estimated in the ESEM approach, and inconsistent with Figure 1 (e.g., m1 loads on GS), were statistically significant (standardized loadings ranged from -.08 to .42). A reasonable answer to Question 1 was: at least four factors explained responses to the CCS.

**Deriving Measures from the CCS: Question 2**—Target rotation was the applied to Model 4 (i.e., Model 8). The measurement model depicted in Figure 1 (i.e., Model 9 – see Table 5s) exhibited statistically significant worse fit than Model 8,  $\Delta\chi^2_R(59) = 273, p < .001$ . Neither model, however, exhibited especially close fit to the data. Many of the modification indices for Model 9 suggested statistical significance if freely estimated, which was consistent with the many statistically significant cross-loadings observed in Table 3. Given the extent of this potential post hoc exploration for a well-fitting measurement model, a reasonable answer to Question 2 was: a more parsimonious CFA model, based on the relevant literature (see Figure 1), did not offer a viable alternative to a more complex ESEM model (see Figure 4s in Appendix D).

Model 8 (target rotation) had nearly as many statistically significant cross-loadings (14 of 71) as did Model 4 (geomin rotation – compare Table 3 to Table 4s in Appendix C). Interfactor correlations under Model 8 ranged from  $r_{GS,M} = .62$  to  $r_{GS,T} = .90$  and from  $r_{GS,M} = .57$  to  $r_{GS,T} = .80$  under Model 4. Problems attributed to multicollinearity have caused particular CCS measures to be dropped in substantive coaching research (e.g., Myers, Wolfe, Maier, Feltz, & Reckase, 2006). Given these two pieces of non-exhaustive pieces of information, plus the mathematical equivalence of the solution under both rotation techniques, a reasonable (though certainly not the only and perhaps not the best) answer to the follow-up question to Question 2 was: geomin rotation was preferred. Table 6s in Appendix C provides interfactor correlation matrices for Model 4, Model 8, and Model 9. Interfactor correlations were somewhat higher in Model 9.

### Multiple-Group EFA in ESEM

An important question for the emerging line of research involving coaching competency within coaching effectiveness models going forward is: are measures comparable across groups for which the CCS and APCCS II-HST are intended? Borrowing from the multiple-

group CFA approach to measurement invariance, multiple-group EFA in ESEM allows researchers to fully test for measurement invariance across groups of substantive interest while generally imposing fewer restrictions than in the CFA framework. For textual parsimony as well as sample size and model-data fit considerations (Yuan & Bentler, 2004), an illustration will be provided based on data from the APCCS II-HST only. The study design in Myers et al. (2010) assumed factorial invariance by country and this variable served as the grouping variable (i.e., United States, U.S. = reference group, and the United Kingdom, U.K. = focal group). Model 5 from Table 1s in Appendix C (i.e., 5 ESEM factors with Geomin rotation) served as the starting point.

Under the theta parameterization, data were fit to four increasingly restrictive multiple-group models for ordinal data (Millsap & Yun-Tien, 2004). Before testing for factorial invariance, the model was imposed separately in each group: Model 1a (U.S. only) and Model 1b (U.K. only). Model 2 through Model 5 tested for factorial invariance. Model 2 (baseline) imposed constraints (detailed in Appendix B\_2 of the supplemental electronic-only materials) for identification purposes. Model 3 added the constraint of invariant pattern coefficients (invariant  $\Lambda$ ) to Model 2. Model 4 added the constraint of invariant thresholds (invariant  $\Lambda$  and  $\tau$ ) to Model 3. Model 5 added the constraint of an invariant residual covariance matrix (invariant  $\Lambda$ ,  $\tau$ , and  $\Theta$ ) to Model 4. Nested models were compared as described in the EFA in ESEM section.

**Countries separately**—There was evidence for close fit in the U.S. sample, Model 1a:

$\chi^2_R(61) = 97, p = .002, RMSEA = .03, CFI = .99, \text{ and } TLI = .97$ . There was evidence for exact fit in the U.K. sample, Model 1b:  $\chi^2_R(61) = 53, p = .761, RMSEA = .00, CFI = 1.00, \text{ and } TLI = 1.01$ . At least part of the difference between the model-data fit by country was likely due to the difference in sample size:  $n_{U.S.} = 491$  and  $n_{U.K.} = 257$ .

**Model 2 through Model 5**—Model 2 (baseline) provided evidence for exact fit:

$\chi^2_R(122) = 140, p = .132, RMSEA = .02, CFI = 1.00, \text{ and } TLI = .99$ . Model 3 (invariant  $\Lambda$ ) fit the data as well as Model 2,  $\Delta\chi^2_R(60) = 72, p = .142$ , and provided evidence for exact fit:  $\chi^2_R(182) = 203, p = .139, RMSEA = .02, CFI = 1.00, \text{ and } TLI = 1.00$ . Model 4 (invariant  $\Lambda$  and  $\tau$ ) fit the data as well as Model 3 (invariant  $\Lambda$ ),  $\Delta\chi^2_R(29) = 41, p = .066$ , and provided evidence for exact fit:  $\chi^2_R(211) = 237, p = .105, RMSEA = .02, CFI = 1.00, \text{ and } TLI = .99$ . Model 5 (invariant  $\Lambda$ ,  $\tau$ , and  $\Theta$ ) fit the data as well as Model 4 (invariant  $\Lambda$  and  $\tau$ ),  $\Delta\chi^2_R(17) = 23, p = .166$ , and provided evidence for exact fit:  $\chi^2_R(228) = 254, p = .117, RMSEA = .02, CFI = 1.00, \text{ and } TLI = 1.00$ . Thus, a reasonable conclusion was that there was evidence for factorial invariance by country (i.e., U.S. and U.K.) with regard to responses to the APCCS II-HST.

From a mathematical perspective, “measurement invariance does not require invariance in the factor covariance matrices or factor means” (Millsap & Yun-Tien, 2004, p. 484). From a conceptual perspective, however, (non-)invariance of these parameters ( $\psi$  and  $\alpha$ , respectively) can be informative with regard to substantive theory (Marsh et al., 2009), and therefore, both of these input files are available in Appendix E of the supplemental electronic-only materials. Briefly, Model 6 (invariant  $\Lambda$ ,  $\tau$ ,  $\Theta$ , and  $\psi$ ) fit the data as well as Model 5,  $\Delta\chi^2_R(15) = 21, p = .128$ , while Model 7 (invariant  $\Lambda$ ,  $\tau$ ,  $\Theta$ ,  $\psi$ , and  $\alpha$ ) fit the data significantly worse than Model 6,  $\Delta\chi^2_R(5) = 38, p < .001$ , with evidence for significantly lower latent means in the UK sample as compared to the US sample with regard to game strategy, technique, and character building.

## Summary and Conclusions

This substantive-methodological synergy attempted to merge an extant conceptual model, coaching competency, and a new methodological framework, ESEM, as compared to a more familiar methodological framework, CFA, for the next generation of research in the psychology of sport and exercise to consider. The synergy was a demonstration of how (E)SEM may be used guided by content knowledge to develop better, or confirm existing, measurement models for coaching competency by re-analyzing existing data. A key consideration for deciding between ESEM and the accompanying rotation criterion and CFA in future validity studies within the psychology of sport and exercise should be level of a priori measurement theory.

It should be reiterated that there is no right or wrong rotation criterion from a mathematical perspective in regard to model-data fit. That said, it has been known for some time that mathematically equivalent solutions do not necessarily do an equivalent job of recovering particular parameter values of interest under complex structures (e.g., Thurstone, 1947). Oblique versions of two rotation criterion, geomin and target, were used in this manuscript and a brief case for their inclusion was put forth in the rotation section. Others may prefer different rotation criteria or even one of the selected rotation criterion but with the  $f(\Lambda)$  specified slightly differently (e.g., non-default values for  $\epsilon$  under geomin rotation – see Morin & Mañano, 2011, for an example) for the CCS and, perhaps, more generally. Selecting a particular rotation criterion is a multi-faceted post estimation decision that cannot be proven to be “correct” when population values are unknown. For example, in the previous section geomin rotation was preferred for the CCS, in part, because interfactor correlations were somewhat smaller than under target rotation. Clearly, lowering the interfactor correlations was not a sufficient condition for selecting the “right” rotated solution for the CCS. If it were a sufficient condition then the “right” rotation criterion presumably would be one that forced an orthogonal solution, doing so however, would have been inconsistent with a priori substantive theory depicted in both Figure 1 and Figure 2.

It is important to note that statistical models often can be used for exploratory or confirmatory purposes regardless of the semantics used (e.g., *exploratory/confirmatory* factor analysis). For example, if an a priori theory justified that each indicator was influenced by each factor then confirmatory evidence may be provided for such a theory (within theoretical identification limits) under either CFA or ESEM approach. A similar issue arises in the determination of the number of factors,  $m$ , to specify/retain in ESEM. The approach taken in this manuscript: a series of nested model comparisons manipulating the number of factors,  $m$ , while considering the interpretability of the solution with regard to a priori theory, and, possible empirical limitations of the nested model comparisons is consistent with a less than strictly confirmatory approach taken in Myers, Chase, Pierce, and Martin, 2011: “comparing\_  $ms$ ”. A fixed number of factors based on theory alone, “theory\_  $m$ ”, can be specified in ESEM consistent with a more strictly confirmatory approach. The author of this manuscript generally prefers some version of the comparing\_  $ms$  approach when ESEM is used because it empirically subsumes the theory\_  $m$  approach (i.e., the theory\_  $m$  solution is fitted). That said, as in the current manuscript, even under the comparing\_  $ms$  approach the analyst may not strictly adhere to model-data fit information alone – although a reasonable expectation in such cases may be to provide this information in a manuscript. Clearly there may be instances (e.g., very strong a priori theory and previous empirical information consistent with this theory) where theory\_  $m$  may be justifiable. A “golden rule” mandating one approach over the other in all instances seems unreasonable.

Coaching competency occupies a central role in models of coaching effectiveness, which makes measurement of this construct an important area of research. The CCS and the APCCS II-HST were designed to measure coaching competency. The initial athlete-level measurement model for responses to the CCS and the APCCS II-HST both were put forth prior to the emergence of ESEM (i.e., fitted to CFA). ESEM analysis on extant datasets suggests that for responses to the APCCS II-HST, a CFA model based on the relevant literature plus one post hoc modification (see Figure 2), offered a viable alternative to a more complex ESEM model. For responses to the CCS, a CFA model based on the relevant literature (see Figure 1) did not offer a viable alternative to a more complex ESEM model (see Figure 3s). The author of this manuscript recommends that the measurement models advocated (Figure 2 and Figure 3s) serve as a starting place for subsequent research. The author of this manuscript also recommends that the ESEM framework should be considered in subsequent studies validity studies - for new and/or existing instruments in the psychology of sport and exercise. That said, there is preliminary evidence that ESEM, as compared to CFA, may in some circumstances be more susceptible to non-convergence or convergence to an improper solution when  $N \geq 250$  (Myers, Ahn, et al., 2011b). In cases where ESEM encounters such problems (and when they are intractable), and when more data are not possible, other practical solutions (e.g., summed scores) may be reasonable.

While the substantive part and the methodological part of this manuscript both were focused primarily on measurement, the fuller ESEM provides a broader methodological framework (that includes but is not limited to the EFA and Multiple-Group EFA examples within this manuscript) within which the next generation of effective coaching research may evaluate fuller models of effective coaching. Many latent variable models that could be imposed within the SEM framework could also be imposed within ESEM framework (Asparouhov & Muthén, 2009). An advantage of the ESEM framework, in instances of insufficient a priori measurement model(s), is reduced likelihood of biased estimates wrought by misspecification of the measurement model(s). Simply, studies not focused on measurement can still be adversely affected by misspecification of the measurement model(s) because misspecification of the measurement model(s) can produce biased path coefficients (Kaplan, 1988).

There is much work yet to be done to investigate the utility of coaching competency within broader models of coaching effectiveness. While a few studies have explored the ability of coaching competency to predict satisfaction with the head coach (e.g., Myers, Wolfe, et al., 2006) and role ambiguity (Bosselut et al., 2012), many other key relations implied within Horn's (2002) model await investigation. For example, coaching competency is specified to at least partially mediate the effect of coaching behavior on three sets of variables: (a) athletes' self-perceptions, beliefs, and attitudes, (b) athletes' level and type of motivation, and (c) athletes' performance and behavior. Future research that explores these (or other) important implied relations is encouraged. More broadly, the author of this manuscript believes that it may be time for research on coaching competency to evolve beyond such a strong focus on measurement alone because it has long been suspected that "the ultimate effects that coaching behavior exerts are mediated by the meaning that players attribute to them" (Smoll & Smith, 1989, p. 1527) – effects measurement models do not test. Finally, a theory-based investigation for why, on average, UK high school athletes may rate their head coach's game strategy, technique, and character building competency lower than US high school athletes rate their head coach's game strategy, technique, and character building competency could advance the literature.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Research reported in this publication was supported, in part, by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR000460. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

## References

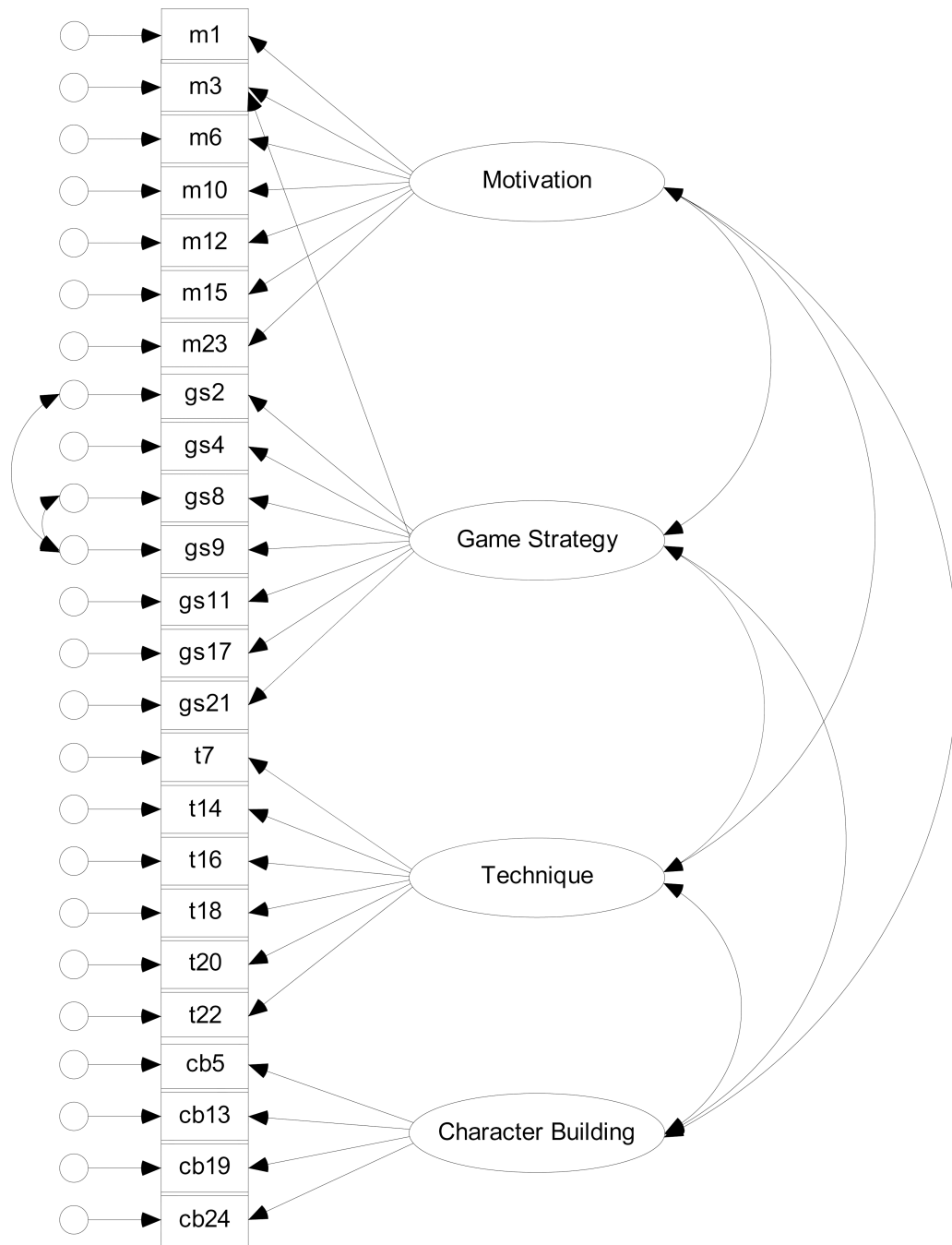
- American Educational Research Association. American Psychological Association. National Council on Measurement in Education. Standards for educational and psychological testing. American Educational Research Association; Washington, DC: 1999.
- Asparouhov T, Muthén BO. Exploratory structural equation modeling. *Structural Equation Modeling*. 2009; 16:397–438. doi:10.1080/10705510903008204.
- Asparouhov, T.; Muthén, BO. Simple second order chi-square correction. 2010. Retrieved from [http://www.statmodel.com/download/WLSMV\\_new\\_chi21.pdf](http://www.statmodel.com/download/WLSMV_new_chi21.pdf)
- Bandura, A. *Self-efficacy: The exercise of control*. Freeman; New York: 1997.
- Bollen, KA. *Structural equations with latent variables*. Wiley; New York: 1989.
- Bosselut G, Heuzé J-P, Eys MA, Fontayne P, Sarrazin P. Athletes' perceptions of role ambiguity and coaching competency in sport teams: A multilevel analysis. *Journal of Sport & Exercise Psychology*. 2012; 34:345–364. [PubMed: 22691398]
- PubMed Browne MW. An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*. 2001; 36:111–150. doi:10.1207/S15327906MBR3601\_05.
- Feltz DL, Chase MA, Moritz SE, Sullivan PJ. A conceptual model of coaching efficacy: Preliminary investigation and instrument development. *Journal of Educational Psychology*. 1999; 91:765–776. doi:10.1037/0022-0663.91.4.765.
- Finney, SJ.; DiStefano, C. Nonnormal and categorical data in structural equation modeling.. In: Serlin, RC.; Hancock, GR.; Mueller, RO., editors. *Structural equation modeling: A second course*. Information Age; Greenwich, CT: 2006. p. 269-313.
- Garrido LE, Abad FJ, Ponsoda V. A New Look at Horn's Parallel Analysis With Ordinal Variables. *Psychological Methods*. 2012 Advance online publication. doi:10.1037/a0030005.
- Hayashi K, Bentler PM, Yuan K. On the likelihood ratio test for the number of factors in exploratory factor analysis. *Structural Equation Modeling*. 2007; 14:505–526. doi:10.1080/10705510701301891.
- Horn, TS. Coaching effectiveness in the sports domain.. In: Horn, TS., editor. *Advances in sport psychology*. 2nd ed.. Human Kinetics; Champaign, IL: 2002. p. 309-354.
- Hox JJ, Maas CJM. The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural Equation Modeling*. 2001; 8:157–174. doi:10.1207/S15328007SEM0802\_1.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55. doi:10.1080/10705519909540118.
- Jackson DL, Gillaspay JA, Purc-Stephenson R. Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*. 2009; 14:6–23. doi:10.1037/a0014694. [PubMed: 19271845]
- Jöreskog KG. A general approach to confirmatory maximum-likelihood factor analysis. *Psychometrika*. 1969; 34:183–202. doi:10.1007/BF02289343.
- Julian MW. The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*. 2001; 8:325–352. doi:10.1207/S15328007SEM0803\_1.
- Kaplan D. The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research*. 1988; 23:467–482. doi:10.1207/s15327906mbr2301\_4.

- MacCallum RC, Roznowski M, Necowitz LB. Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*. 1992; 111:490–504. doi: 10.1037/0033-2909.111.3.490. [PubMed: 16250105]
- Marsh HW, Hau K-T. Application of latent variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*. 2007; 32:151–171. doi:10.1016/j.cedpsych.2006.10.008.
- Marsh HW, Hau KT, Wen Z. In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modeling*. 2004; 11:320–341. doi:10.1207/s15328007sem1103\_2.
- Marsh HW, Lüdtke O, Nagengast B, Trautwein U, Morin AJS, Abduljabbar AS, Köller O. Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*. 2012; 47:106–124. doi:10.1080/00461520.2012.670488.
- Marsh HW, Muthén BO, Asparouhov T, Lüdtke O, Robitzsch A, Morin AJ, Trautwein U. Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*. 2009; 16:439–476. doi: 10.1080/10705510903008220.
- Millsap, RE. *Statistical Approaches to Measurement Invariance*. Routledge; New York: 2011.
- Millsap RE, Yun-Tien J. Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*. 2004; 39:479–515. doi:10.1207/S15327906MBR3903\_4.
- Morin AJS, Maïano C. Cross-Validation of the Short Form of the Physical Self-Inventory (PSI-18) using Exploratory Structural Equation Modeling (ESEM). *Psychology of Sport and Exercise*. 2011; 12:540–554. doi:10.1016/j.psychsport.2011.04.003.
- Morin, AJS.; Marsh, HW.; Nagengast, B. Exploratory Structural Equation Modeling.. In: Hancock, GR.; Mueller, RO., editors. *Structural equation modeling: A second course*. 2nd ed.. Information Age Publishing, Inc.; Charlotte, NC: 2013. p. 395-436.
- Muthén BO. A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*. 1984; 49:115–132. doi:10.1007/BF02294210.
- Muthén BO. Multilevel covariance structure analysis. *Sociological Methods & Research*. 1994; 22:376–398. doi:10.1177/0049124194022003006.
- Muthén, LK.; Muthén, BO. *Mplus User's Guide*. 6th ed.. Muthén & Muthén; Los Angeles, CA: 1998. 2010. Retrieved from <http://www.statmodel.com/ugexcerpts.shtml>
- Myers ND, Ahn S, Jin Y. Sample size and power estimates for a confirmatory factor analytic model in exercise and sport: A Monte Carlo approach. *Research Quarterly for Exercise and Sport*. 2011a; 82:412–423. doi:10.5641/027013611X13275191443621. [PubMed: 21957699]
- Myers, ND.; Ahn, S.; Jin, Y. Exploratory structural equation modeling or confirmatory factor analysis for (Im)perfect validity studies?. Paper presented at the Modern Modeling Methods conference; Storrs, Connecticut, USA. May. 2011b
- Myers ND, Ahn S, Jin Y. Rotation to a partially specified target matrix in exploratory factor analysis: How many targets? *Structural Equation Modeling: A Multidisciplinary Journal*. 2013; 20:131–147. doi: 10.1080/10705511.2013.742399.
- Myers ND, Beauchamp MR, Chase MA. Coaching competency and satisfaction with the coach: A multilevel structural equation model. *Journal of Sports Sciences*. 2011; 29:411–422. doi: 10.1080/02640414.2010.538710. [PubMed: 21184342]
- Myers ND, Chase MA, Beauchamp MR, Jackson B. The Coaching Competency Scale II – High School Teams. *Educational and Psychological Measurement*. 2010; 70:477–494. doi: 10.1177/0013164409344520.
- Myers ND, Chase MA, Pierce SW, Martin E. Coaching efficacy and exploratory structural equation modeling: A substantive-methodological synergy. *Journal of Sport & Exercise Psychology*. 2011; 33:779–806. [PubMed: 22262705]
- PubMed Myers ND, Feltz DL, Chase MA, Reckase MD, Hancock GR. The Coaching Efficacy Scale II - High School Teams. *Educational and Psychological Measurement*. 2008; 68:1059–1076. doi: 10.1177/0013164408318773.

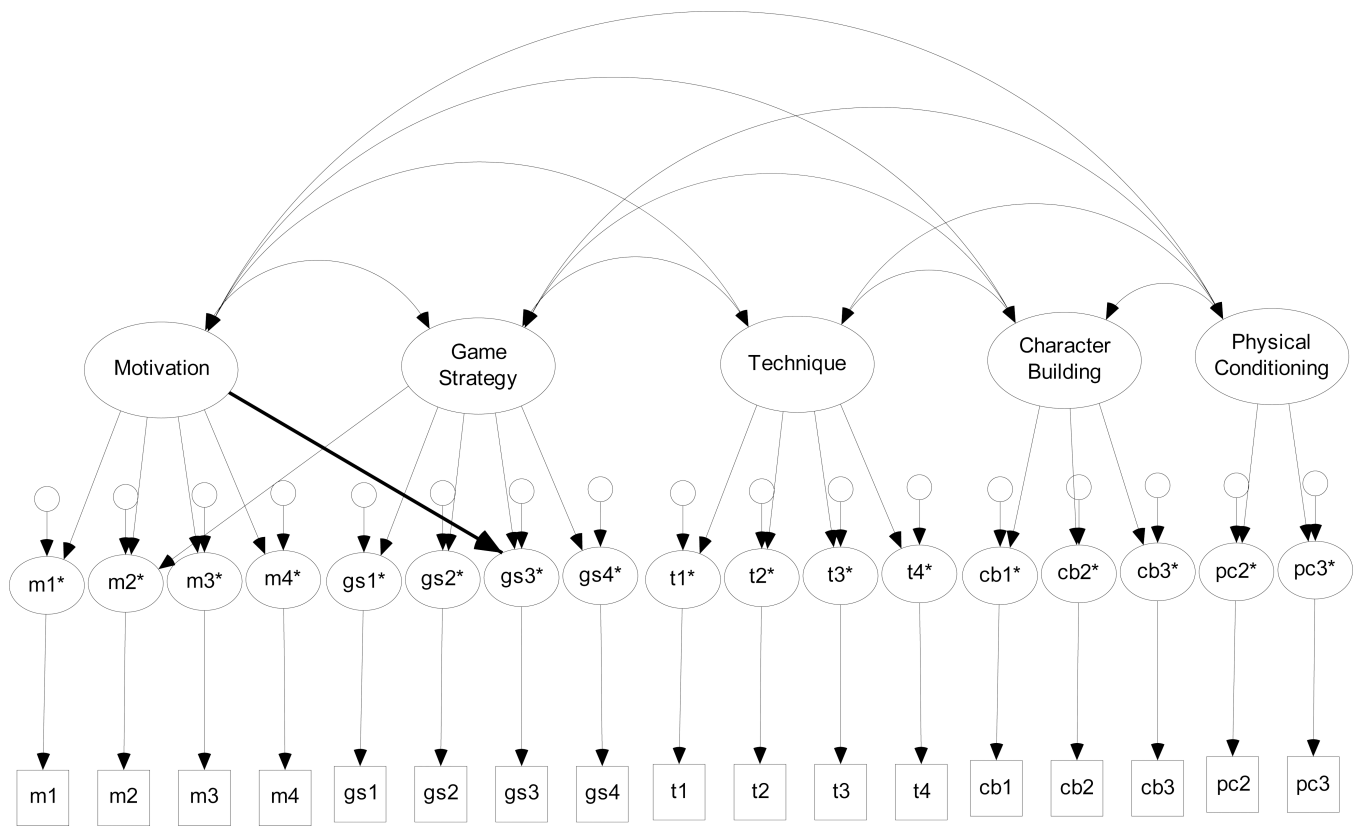
- Myers ND, Feltz DL, Maier KS, Wolfe EW, Reckase MD. Athletes' evaluations of their head coach's coaching competency. *Research Quarterly for Exercise and Sport*. 2006; 77:111–121. doi: 10.5641/027013606X13080769704082. [PubMed: 16646358]
- Myers ND, Feltz DL, Wolfe EW. A confirmatory study of rating scale category effectiveness for the coaching efficacy scale. *Research Quarterly for Exercise and Sport*. 2008; 79:300–311. doi: 10.5641/193250308X13086832905752. [PubMed: 18816941]
- Myers, ND.; Jin, Y. Exploratory structural equation modeling and coaching competency.. In: Potrac, P.; Gilbert, W.; Dennison, J., editors. *Routledge Handbook of Sports Coaching*. Routledge; London: 2013. p. 81-95.
- Myers, ND.; Jin, Y.; Ahn, S. Rotation to a Partially Specified Target Matrix in Exploratory Factor Analysis in Practice. 2012. Manuscript submitted for publication
- Myers ND, Wolfe EW, Maier KS, Feltz DL, Reckase MD. Extending validity evidence for multidimensional measures of coaching competency. *Research Quarterly for Exercise and Sport*. 2006; 77:451–463. doi:10.5641/027013606X13080770015247. [PubMed: 17243220]
- National Association for Sport and Physical Education. *Quality coaches, quality sports: National standards for athletic coaches*. 2nd ed.. Author; Reston, VA: 2006.
- Nunnally, JC. *Psychometric theory*. 2nd ed.. McGraw-Hill; New York: 1978.
- Rozeboom WW. The glory of suboptimal factor rotation: Why local minima in analytic optimization of simple structure are more blessing than curse. *Multivariate Behavioral Research*. 1992; 27:585–599. doi:10.1207/s15327906mbr2704\_5.
- Smoll FL, Smith RE. Leadership behaviors in sport: A theoretical model and research paradigm. *Journal of Applied Social Psychology*. 1989; 19:1522–1551. doi:10.1111/j.1559-1816.1989.tb01462.x.
- Spearman C. "General intelligence," objectively determined and measured. *American Journal of Psychology*. 1904; 15:201–293. Retrieved from <http://www.jstor.org/stable/1412107>.
- Thurstone, LL. *Multiple factor analysis*. University of Chicago Press; Chicago: 1947.
- Vernon T, Eysenck S. Special issue on structural equation modeling [Special issue]. *Personality and Individual Differences*. 2007; 42(5)
- Yates, A. *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. State University of New York Press; Albany: 1987.
- Yuan KH, Bentler PM. On chi-square difference and z tests in mean and covariance structure analysis when the base model is misspecified. *Educational and Psychological Measurement*. 2004; 64(5): 737–757. doi:10.1177/0013164404264853.

- ESEM should be strongly considered in subsequent validity studies.
- Level of a priori measurement theory is a key for deciding between ESEM and CFA.
- Figure 2 and Figure 3s provide measurement models for use in future research.





**Figure 1.**  
Athlete-level a priori measurement model for the CCS



**Figure 2.** Athlete-level a priori measurement model for the APCCS II-HST. Thick (bolded) arrow denotes a post hoc modification proposed in this manuscript.

Table 1

Geomin-Rotated Pattern Coefficients ( $\Lambda^*$ ), Standard Errors (SE), Geomin-Rotated Standardized Pattern Coefficients ( $\Lambda^{*0}$ ), and Percentage of Variance Accounted For ( $R^2$ )

Item	Factor 1 = GS		Factor 2 = PC		Factor 3 = T		Factor 4 = M		Factor 5 = CB		$R^2$			
	$\Lambda^*$	SE	$\Lambda^*$	SE	$\Lambda^{*0}$	SE	$\Lambda^*$	SE	$\Lambda^*$	SE		$\Lambda^{*0}$	SE	
gs1	<b>1.07</b>	.24	<b>0.63</b>	.09	0.05	0.22	.14	0.13	0.09	-0.07	.09	-0.04	.65	
gs2	<b>1.32</b>	.19	<b>0.74</b>	0.04	0.02	0.04	.09	0.02	-0.08	.09	-0.05	.24	.13	.68
gs3	<b>0.67</b>	.18	<b>0.45</b>	-0.07	.08	0.23	.14	0.15	<b>0.35</b>	.14	<b>0.24</b>	0.03	.07	.55
gs4	<b>0.87</b>	.20	<b>0.50</b>	<b>0.46</b>	.13	<b>0.27</b>	-0.02	.08	-0.01	.26	.16	0.01	.06	.67
pc2	0.10	.12	0.05	<b>1.56</b>	.34	<b>0.79</b>	0.08	.11	0.04	-0.01	.06	0.00	0.04	.74
pc3	0.00	.07	0.00	<b>1.44</b>	.30	<b>0.78</b>	0.06	.09	0.03	0.05	.07	0.03	0.04	.71
m1	0.05	.10	0.03	-0.10	.07	-0.06	.18	.18	0.11	<b>0.79</b>	.17	<b>0.50</b>	<b>0.43</b>	.60
m2	0.24	.14	0.17	0.18	.09	0.13	0.05	.08	0.03	<b>0.48</b>	.13	<b>0.34</b>	0.19	.49
m3	-0.06	.10	-0.03	0.03	.04	0.02	.16	.16	0.10	<b>1.08</b>	.22	<b>0.65</b>	0.22	.64
m4	0.08	.11	0.05	0.13	.11	0.07	-0.18	.15	-0.09	<b>1.59</b>	.23	<b>0.85</b>	-0.05	.71
t1	-0.02	.08	-0.01	0.05	.08	0.03	<b>1.59</b>	.18	<b>0.84</b>	0.08	.08	0.04	-0.07	.73
t2	0.15	.12	0.10	0.20	.11	0.13	<b>0.93</b>	.13	<b>0.61</b>	-0.09	.10	-0.06	0.03	.58
t3	0.05	.10	0.03	-0.05	.08	-0.02	<b>1.60</b>	.22	<b>0.82</b>	0.02	.08	0.01	0.10	.74
t4	0.16	.11	0.10	0.09	.12	0.05	<b>1.00</b>	.16	<b>0.60</b>	0.10	.07	0.06	0.08	.64
cb1	<b>0.46</b>	.21	<b>0.25</b>	-0.02	.05	-0.01	-0.13	.11	-0.07	-0.02	.06	-0.01	<b>1.36</b>	.70
cb2	-0.01	.07	-0.01	0.18	.12	0.11	0.34	.19	0.20	0.16	.14	0.10	<b>0.89</b>	.65
cb3	-0.01	.08	-0.01	0.07	.08	0.05	0.18	.19	0.12	0.06	.11	0.04	<b>0.94</b>	.56

Note. Statistically significant coefficients were bolded. GS = game strategy; PC = physical conditioning; T = technique; M = motivation; CB = character building.

Table 2

Confirmatory Factor Analytic Pattern Coefficients ( $\lambda$ ), Standard Errors (SE), Standardized Pattern Coefficients ( $\lambda^0$ ), and Percentage of Variance Accounted For ( $R^2$ )

Item	Factor 1 = GS		Factor 2 = PC		Factor 3 = T		Factor 4 = M		Factor 5 = CB		$R^2$
	$\lambda_{p1}$	SE	$\lambda_{p2}$	SE	$\lambda_{p3}$	SE	$\lambda_{p4}$	SE	$\lambda_{p5}$	SE	
gs1	1.32	.11	<b>0.80</b>	---	---	---	---	---	---	---	.64
gs2	1.31	.11	<b>0.80</b>	---	---	---	---	---	---	---	.63
gs3	0.79	.13	<b>0.54</b>	---	---	---	<b>0.33</b>	.11	<b>0.23</b>	---	.53
gs4	1.48	.12	<b>0.83</b>	---	---	---	---	---	---	---	.69
pc2	---	---	<b>1.72</b>	.16	<b>0.86</b>	---	---	---	---	---	.75
pc3	---	---	<b>1.54</b>	.15	<b>0.84</b>	---	---	---	---	---	.70
t1	---	---	---	---	<b>1.49</b>	.10	<b>0.83</b>	---	---	---	.69
t2	---	---	---	---	<b>1.16</b>	.10	<b>0.76</b>	---	---	---	.58
t3	---	---	---	---	<b>1.61</b>	.13	<b>0.85</b>	---	---	---	.72
t4	---	---	---	---	<b>1.43</b>	.10	<b>0.82</b>	---	---	---	.67
m1	---	---	---	---	---	---	<b>1.28</b>	.12	<b>0.79</b>	---	.62
m2	<b>0.42</b>	.12	<b>0.29</b>	---	---	---	<b>0.65</b>	.11	<b>0.46</b>	---	.51
m3	---	---	---	---	---	---	<b>1.34</b>	.12	<b>0.80</b>	---	.64
m4	---	---	---	---	---	---	<b>1.25</b>	.12	<b>0.78</b>	---	.61
cb1	---	---	---	---	---	---	---	---	<b>1.29</b>	.12	.62
cb2	---	---	---	---	---	---	---	---	<b>1.58</b>	.12	.71
cb3	---	---	---	---	---	---	---	---	<b>1.06</b>	.10	.53

Note. Statistically significant coefficients were bolded. GS = game strategy; PC = physical conditioning; T = technique; M = motivation; CB = character building.

**Table 3**

Geomin-Rotated Pattern Coefficients ( $\Lambda^*$ ), Standard Errors (SE), Geomin-Rotated Standardized Pattern Coefficients ( $\Lambda^{*0}$ ), and Percentage of Variance Accounted For ( $R^2$ )

Item	Factor 1 = GS		Factor 2 = M		Factor 3 = T		Factor 4 = CB		$R^2$				
	$\Lambda^*_{p1}$	SE	$\Lambda^*_{p2}$	SE	$\Lambda^*_{p3}$	SE	$\Lambda^*_{p4}$	SE					
gs2	<b>1.28</b>	.22	<b>0.83</b>	0.02	0.01	-0.02	.07	-0.01	0.07	0.05	.73		
gs4	<b>1.33</b>	.22	<b>0.81</b>	0.04	0.03	0.14	.15	0.08	-0.01	0.06	-0.01	.79	
gs8	<b>1.39</b>	.24	<b>0.83</b>	0.07	0.04	0.06	.13	0.04	0.00	0.07	0.00	.79	
gs9	<b>1.21</b>	.21	<b>0.77</b>	-0.11	0.06	-0.07	.09	0.09	0.09	0.06	0.06	.72	
gs11	<b>1.03</b>	.26	<b>0.58</b>	<b>0.24</b>	<b>0.09</b>	<b>0.13</b>	0.27	.19	0.15	0.13	0.09	0.07	.73
gs17	<b>0.70</b>	.21	<b>0.41</b>	0.15	.12	0.08	<b>0.73</b>	<b>.21</b>	<b>0.42</b>	0.12	0.09	0.07	.79
gs21	<b>0.82</b>	.24	<b>0.47</b>	-0.12	0.08	-0.07	<b>0.69</b>	<b>.23</b>	<b>0.40</b>	0.20	.12	0.12	.76
ml	<b>0.57</b>	.15	<b>0.31</b>	<b>1.26</b>	<b>.17</b>	<b>0.69</b>	-0.04	.19	-0.02	-0.05	.08	-0.03	.76
m3	<b>1.07</b>	.25	<b>0.61</b>	<b>0.71</b>	<b>.11</b>	<b>0.40</b>	-0.04	.10	-0.03	-0.14	.11	-0.08	.68
m6	-0.01	.07	0.00	<b>1.66</b>	<b>.21</b>	<b>0.82</b>	0.15	.21	0.08	0.18	.11	0.09	.85
m10	<b>0.60</b>	.30	<b>0.31</b>	<b>1.14</b>	<b>.16</b>	<b>0.59</b>	0.01	.18	0.01	0.18	.12	0.09	.77
m12	0.43	.24	0.23	<b>0.83</b>	<b>.19</b>	<b>0.44</b>	0.08	.16	0.04	<b>0.46</b>	<b>.18</b>	<b>0.24</b>	.67
m15	-0.26	.14	-0.13	<b>1.48</b>	<b>.20</b>	<b>0.73</b>	<b>0.71</b>	<b>.19</b>	<b>0.35</b>	0.07	.05	0.04	.87
m23	0.20	.20	0.10	<b>1.17</b>	<b>.15</b>	<b>0.59</b>	0.23	.15	0.12	<b>0.43</b>	<b>.12</b>	<b>0.22</b>	.81
t7	<b>0.62</b>	.28	<b>0.34</b>	0.28	.19	0.15	<b>0.74</b>	<b>.37</b>	<b>0.41</b>	-0.25	.14	-0.14	.52
tl4	0.01	.08	0.01	0.32	.18	0.18	<b>1.42</b>	<b>.26</b>	<b>0.80</b>	<b>-0.15</b>	<b>.06</b>	<b>-0.08</b>	.76
tl6	0.05	.09	0.03	<b>0.31</b>	<b>.15</b>	<b>0.18</b>	<b>1.32</b>	<b>.20</b>	<b>0.76</b>	-0.01	.07	-0.01	.81
tl8	<b>0.54</b>	.22	<b>0.28</b>	0.23	.18	0.12	<b>0.71</b>	<b>.26</b>	<b>0.37</b>	<b>0.30</b>	<b>.15</b>	<b>0.16</b>	.67
t20	0.42	.22	0.24	-0.12	0.08	-0.07	<b>1.01</b>	<b>.24</b>	<b>0.58</b>	0.24	.13	0.14	.71
t22	0.30	.20	0.18	-0.02	.12	-0.01	<b>1.20</b>	<b>.28</b>	<b>0.71</b>	0.07	.07	0.04	.77
cb5	0.17	.13	0.09	<b>0.64</b>	<b>.16</b>	<b>0.32</b>	-0.10	.14	-0.05	<b>1.23</b>	<b>.22</b>	<b>0.62</b>	.78
cb13	0.16	.16	0.09	0.34	.19	0.18	0.04	.10	0.02	<b>1.18</b>	<b>.21</b>	<b>0.63</b>	.71
cb19	-0.07	.08	-0.04	0.03	.10	0.02	0.13	.15	0.07	<b>1.63</b>	<b>.25</b>	<b>0.88</b>	.84
cb24	0.01	.09	0.00	<b>0.53</b>	<b>.18</b>	<b>0.27</b>	0.10	.12	0.05	<b>1.27</b>	<b>.19</b>	<b>0.65</b>	.78

Note. Statistically significant coefficients were bolded. GS = game strategy; M = motivation; T = technique; CB = character building.