Mesure et évaluation en éducation

Application of Structural Equation Modeling to the Social Sciences: A Brief Guide for Researchers

Vaithehy Shanmugam and John E. Marsh

See table of contents

Explore this journal

Cite this article

Shanmugam, V. & Marsh, J. E. (2015). Application of Structural Equation Modeling to the Social Sciences: A Brief Guide for Researchers. *Mesure et évaluation en éducation*, *37*(3), 99–123. https://doi.org/10.7202/1036329ar

Article abstract

Emanating from a family of statistical techniques used for the analysis of multivariate data to measure latent variables and their interrelationships, structural equation modeling (SEM) is briefly introduced. The basic tenets of SEM, the principles of model creation, identification, estimation and evaluation are outlined and a four-step procedure for applying SEM to test an evidence-based model of eating disorders (transdiagnostic cognitive-behavioural theory; Fairburn, Cooper, & Shafran, 2003) using previously obtained data on eating psychopathology within an athletic population (Shanmugam, Jowett, & Meyer, 2011) is presented and summarized. Central issues and processes underpinning SEM are discussed and it is concluded that SEM offers promise for testing complex, integrated theoretical models and advances of research within the social sciences, with the caveat that it should be restricted to situations wherein there is a pre-existing substantial base of empirical evidence and a strong conceptual understanding of the theory undergirding the research question.

Application of Structural Equation Modeling to the Social Sciences: A Brief Guide for Researchers

Vaithehy Shanmugam
John E. Marsh
University of Central Lancashire

Emanating from a family of statistical techniques used for the analysis of multivariate data to measure latent variables and their interrelationships, structural equation modeling (SEM) is briefly introduced. The basic tenets of SEM, the principles of model creation, identification, estimation and evaluation are outlined and a four-step procedure for applying SEM to test an evidence-based model of eating disorders (transdiagnostic cognitive-behavioural theory; Fairburn, Cooper, & Shafran, 2003) using previously obtained data on eating psychopathology within an athletic population (Shanmugam, Jowett, & Meyer, 2011) is presented and summarized. Central issues and processes underpinning SEM are discussed and it is concluded that SEM offers promise for testing complex, integrated theoretical models and advances of research within the social sciences, with the caveat that it should be restricted to situations wherein there is a pre-existing substantial base of empirical evidence and a strong conceptual understanding of the theory undergirding the research question.

Cet article propose une brève introduction à la modélisation par équations structurelles (MES), une technique statistique d'analyse de données multivariées qui vise à mesurer des variables latentes et leurs interrelations. Les préceptes de la MES et les principes de création, d'identification, d'estimation et d'évaluation de modèle y sont décrits. Son utilisation est illustrée par la présentation d'une procédure d'application de la MES en quatre étapes qui teste un modèle fondé sur les données probantes des troubles de l'alimentation (théorie cognitive-comportementale transdiagnostique; Fairburn, Cooper, & Shafran, 2003) en utilisant les données obtenues précédemment sur les troubles alimentaires au sein d'une popu-

lation sportive (Shanmugam, Jowett, & Meyer, 2011). Des questions centrales et les processus qui sous-tendent la MES sont discutés, et il est conclu que la MES est une technique très prometteuse pour tester les modèles théoriques intégrés complexes et les avancées de la recherche en sciences sociales, tant que son utilisation est limitée aux situations où il existe une importante base de données probantes ainsi qu'une solide compréhension conceptuelle de la théorie sur laquelle repose la question de recherche.

Palavras-chaves: modelagem de equações estruturais, análise fatorial confirmatória, modelo de medição, modelo estrutural

Proveniente de uma família de técnicas estatísticas utilizadas na análise de dados multivariados para medir variáveis latentes e suas inter-relações, apresenta-se sumariamente a modelagem de equações estruturais (MES). Neste sentido, são descritos os princípios básicos da MES, os princípios da criação, identificação, estimativa e avaliação de modelos e são apresentadas e resumidas as quatro etapas de um procedimento para a aplicação da MES para testar um modelo baseado em evidências de transtornos alimentares (teoria cognitivo-comportamental transdiagnóstica; Fairburn, Cooper, & Shafran, 2003), utilizando os dados anteriormente obtidos em transtornos alimentares dentro de uma população de atletas (Shanmugam, Jowett, & Meyer, 2011). As questões centrais e os processos subjacentes à MES são discutidos, concluindo-se que a aplicação da MES é bastante promissora para testar modelos teóricos integrados e complexos e para avanços da investigação no âmbito das ciências sociais, com a ressalva de que deve restringir-se a situações em que pré-exista uma base substancial de evidências empíricas e uma forte compreensão conceptual da teoria que sustenta a questão de investigação.

Authors' Note: For correspondence: Vaithehy Shanmugam, School of Psychology, University of Central Lancashire, Darwin Building, Preston, Lancashire, United Kingdom, PR1 2HE, Phone: (+44) 177 28 3257, E-mail: [vshanmugam@uclan.ac.uk].

Introduction

Structural Equation Modeling (SEM) stems from the family of multivariate analyses. It serves a very similar purpose to that of multiple regressions. While multiple regressions are used to examine the independent predictors of a dependent variable, out a of set of independent variables, SEM represents translations of a series of hypothesized 'cause and effect' relationships between variables into a composite hypothesis concerning patterns of statistical dependencies (Shipley, 2000) and offers a comprehensive method for the simultaneous quantification and testing of theoretical models (Pugesek, Tomer, & von Eye, 2003). Specifically, the theoretical model represents causal processes that generate observations of multiple variables (Bentler, 1980) and the relationships between such variables are described as parameters that indicate the magnitude of the effect (either direct or indirect) that the independent/exogenous variable(s) have on the dependent/endogenous variable(s). As such, if the model achieves acceptable 'goodness of fit' then it is argued that the postulated relations of the model are plausible; however if the goodness of fit indices are inadequate/unacceptable, then the tenability of such relations is rejected (Byrne, 2006). SEM is an extension of multiple regressions in that it is multivariate and, as such, can simultaneously assess several regressions at one given time, as well as allowing variables to be classified as both exogenous and endogenous within the same model (Schumacker & Lomax, 2004). Moreover, it takes a confirmatory, as opposed to exploratory, approach to data analysis by demanding that the pattern of observed relationships are specified prior to model testing (Byrne, 2006). Finally, it exerts the ability to account and correct for measurement errors, be they random (e.g., sampling error) or systematic (e.g., underlying psychometric properties related to the measure) as the analysis is conducted at the measurement level, by incorporating the error/residual error variance in the estimated model of which traditional multivariate analyses such as regressions are not capable (Kline, 2005), as well as at the structural level by incorporating disturbances.

Although initially developed for use in genetics (Wright, 1921), since its introduction, the use of SEM as a statistical tool to evaluate theoretical and conceptual models and/or to test empirical relations between psychological constructs has gained momentum and grown in popularity in several disciplines such as psychology, sociology and economics. Although there appear to be a number of books dedicated to this topic within education (e.g., Teo & Khine, 2009), the application of such analyses within this research area is considered to be limited (Karadag, 2012). This is quite surprising given that the measures, research questions, and research designs used within education have become more complex, thus calling for more sophisticated and robust methods of analysis. Therefore the purpose of the current paper is to introduce and explain the key concepts and principles of SEM, discuss the advantages of SEM over other multivariate analyses, and integrate a research example to demonstrate the various stages involved in SEM.

Key concepts and principles in SEM

It is generally accepted that a two-step approach is undertaken when conducting SEM (e.g., James, Mulaik, & Brett, 1982; Kline, 2005; Schumacker & Lomax, 2004). Specifically, this approach involves the testing of two models: the measurement model and the structural model. Before proceeding to these, it is important to note that there are two primary variables in SEM: observed (indicators; e.g., individual items pertaining to psychometric instruments) and latent (constructs; e.g., subscales of psychometric instruments). Specifically, latent variables are not measured directly; rather, they are inferred constructs measured indirectly through the observed variables. It is common that multiple observed variables underlie the latent variable and, as such, the benefit of this is that measurement errors related to the reliability or the validity of the observed variable are accounted for (Kline, 2005).

Measurement model

The measurement model is a confirmatory factor model and is often conducted first in SEM. The main objective of the measurement model is to discover the reliability and validity of the observed variables in relation to the latent variable (e.g., are the observed variables accurately measuring the construct under examination?). Traditionally, each latent

variable should be represented by multiple indicators (three as a minimum). As such, the relationship between the latent variable and the observed variables is indicated by factor loadings (Byrne, 2006). The factor loadings generate and highlight the extent to which the observed variables are able to measure the latent variable. In addition to producing factor loadings, the measurement model also generates the measurement error associated with the observed variables. Measurement error specifically highlights the extent to which the observed variables are measuring something other than the latent variable it is proposed to measure (Kline, 2005). As such, a factor loading of .40 per observed variable is deemed acceptable (Ford, MacCallum, & Tait, 1986).

Structural model

The second process in SEM involves the structural model. While the measurement is concerned with the reliability and validity of the latent variables, the structural model is primarily concerned with the interrelations between the latent variables. Specifically, the structural model tests the extent to which the hypothesized or theorized relations between the latent variables are supported within the current sample under investigation.

Prior to conducting SEM analyses, it is advised that three preliminary issues related to sample size and convergence, model specification, and model identification are addressed (Marsh, 2007; Schumacker & Lomax, 2004). Each issue will now be discussed accordingly.

Sample size and item parceling

In order for the model to converge (run), it is recommended that there be between five and ten participants per observed variable (e.g., Bentler & Chou, 1987; Byrne, 2006), with a total of 200 participants as the minimum (Bentler, 1999). However, this may not always be feasible within the research setting, especially if a large number of psychological constructs or complicated theoretical models are being tested. A common method to overcome this shortage in participant numbers is item parceling; whereby the items of the underlying latent variable are grouped together to produce parcels of two to six (Marsh & Hau, 1999; Yang, Nay, & Hoyle, 2010).

Several methods for parceling have been suggested, including, parceling all items into a single parcel (mean of each latent variable), splitting all odd and even items into two parcels, randomly selecting a certain number of items to create three or four parcels (e.g., Yang et al., 2010), parceling items that have similar factor loadings (Cattell & Burdsal, 1975), parceling items with high factor loadings with low factor loadings to equalize the loadings (Russell, Kahn, Spoth, & Altmaier, 1998) and parceling items according to their skew (Hau & Marsh, 2004; Nasser-Abu Alhija & Wisenbaker, 2006; Thompson & Melancon, 1996). Although one method of parceling is not advocated over another, the guidelines of Hau and Marsh (2004) and Nasser-Abu Alhija and Wisenbaker (2006) are seen as favourable as in this method items are parceled according to the size and direction of their skew. Specifically, the most skewed items are parceled with the least skewed items, then the next most skewed to the next least skewed and so on. In addition, this process is counterbalanced in that items that were negatively skewed are parceled with positively skewed items.

To parcel or not to parcel?

The use of parcels instead of the indicators has sparked debate among researchers (see Little, Cunningham, Shahar, & Widaman 2002). For example, there are numerous empirical justifications for using parceling, including increased reliability (Kishton & Widamann, 1994), achieving normality within the data (Bandalos, 2002; Nasser & Wisenbaker, 2003), remedying small sample sizes and unstable parameter estimates (Bandalos & Finney, 2001), as well as a greater likelihood of achieving a proper model solution (Bandalos, 2002; Marsh, Hau, Balla, & Grayson, 1998). However, such adventurous properties of item parceling are only said to be effective if the observed items of the underlying latent factor are unidimensional (Bandalos, 2002; Hall, Snell, & Foust, 1999, Yang et al., 2010).

Empirically, the effects of parceling over individual indicators of the latent factors have been documented in several simulation studies (e.g., Bandalos, 2002; Marsh et al., 1998; Yuan, Bentler, & Kano, 1997) and results have demonstrated that it is more beneficial to parcel than to use the same number of individual items, as when parcels were used, not only were the fit indices more adequate, but the results were more likely to yield a proper solution. However, parceling has been likened to 'cheating'

as it creates bias in the individual's responses by changing their original scores, which could subsequently manufacture a false structure (Little et al., 2002). Moreover, many of the measures often employed within research have already established norms for population; however, by parceling items, the meaningfulness of these norms can be lost (Little et al., 2002; Violato & Hecker, 2007). For example, if one were to compare the eating disordered symptoms of female students and male students using the Eating Disorder Examination Questionnaire (EDEQ 6.0; Fairburn & Beglin, 2008), the use of parcels will prohibit any meaningful comparisons with pre-existing norms; thus, in terms of applied implications, the use of parcels produces arbitrary data.

Model Specification

Model specification relates to the process where assertions are made about which effects are null, which are fixed, and which are freely estimated (see Figure 1). SEM operates only using a priori hypotheses. Thus any research question developed should be guided by relevant theory and empirical evidence as well as reflected in reliable and valid psychometric measures. Using theory and empirical evidence, testable model(s) are developed and subsequently specified. Specifically, the relations between variables at both the measurement (e.g., pathways, covariances, etc.) and structural level are clarified and defined.

Model Identification

Model identification refers to whether the unique set of parameters is consistent with the data: whether it is possible to attain unique values for the parameters of the model (Violato & Hecker, 2007). Specifically, identification relates to the transposition of the variance–covariance matrix of the observed variables (the data points; number of observed variables) into the structural parameters of the model under examination (Byrne, 2006).

There are three variants of model identification; *under-identified*, *just-identified* and *over-identified*. An *under-identified* model is one where the number of parameters to be estimated exceeds the number of data points. This type of model is perceived as problematic because the model is considered to contain insufficient information to attain a fixed solution of parameters estimation, meaning that there are an infinite number

of possible solutions (Byrne, 2006). Moreover, the parameter estimates are considered to be untrustworthy (Kline, 2005). A *just-identified* model is one where the number of data points equals the number of parameters to be estimated (e.g., a saturated model). Consequently, this type of model is also considered problematic as it will always achieve a perfect fit to the empirical data (Pugesek et al., 2003) and can never be rejected. The final variant is an *over-identified* model, where the number of available data points is greater than the number of parameters to be estimated, thus resulting in positive degrees of freedom, allowing for model rejection.

To calculate whether a model is identified, the following equation is often employed, where p = data points/number of observed variables:

$$p\ (p+1)/2 > \text{parameters to be estimated}$$

Model Estimation

Following model specification and identification, the hypothesized model is then estimated. Model estimation determines how the tested model fits the generated data based on the extent to which the observed covariance matrix (data generated) is equivalent to the model-implied covariance matrix (e.g., hypothetical model; Lei & Wu, 2007). This comparison between two covariance matrices can be expressed in the following equation, $\Sigma = \Sigma(\theta)$. In this equation, Σ (sigma) represents the population covariance matrix of observed variance, θ (theta) represents the vector comprised of the population parameters and $\Sigma(\theta)$ is the covariance presented as a function of θ (Violato & Hecker, 2007). Violato and Hecker proposed a "hand in glove" metaphor, which may be useful in understanding this process. In this metaphor, the glove is the model ($\Sigma(\theta)$), while the hand is the data (Σ). In the attempt to find the perfect fitting glove (i.e., model or theory) for the hand (i.e., the data), the lack of fit (i.e., too big, too small) is represented by the θ vector.

Unlike ANOVAs and regressions, which tend to use the least squares methods of estimation, SEM uses iterative estimation methods. This method consists of repeating calculations until the best-fitting estimations for the parameters are obtained. There is a number of methods of estimation, including Maximum Likelihood (ML), Generalised Least Squares (GLS) and Asymptotic Distribution Free (ADF). However, the most frequently employed method is ML, often the default estimation procedure on many SEM programmes. The ML procedure operates by

providing estimates of parameters that maximize the likelihood that the predicted model fits the observed model based on the covariance matrix (Bollen, 1989; Violato & Hecker, 2007) and functions on the assumptions that data is normally distributed and that the sample size is large.

Model Evaluation and Respecification

This process of model estimation leads to the Goodness-of-Fit (GOF) testing. The GOF is critical to conducting SEM, as it allows the adequacy of the tested model to be evaluated and permits comparison of the efficacy of multiple competing models. Specifically, GOF reflects the extent to which the model fits the data. In order to find a statistically significant theoretical model with practical and substantive meaning, multiple goodness-of-fit indices to assess model fit have been put forward. Although there are no concrete rules about which fit statistics to use to evaluate models, a combination of fit statistics are employed when comparing and contrasting models.

The first is the non-statistical significance of the chi-square (χ^2). A non-significant χ^2 suggests that the sample covariance (e.g., theoretical model) and the reproduced model implied covariance matrix (tested model) are similar. However, it should be noted that χ^2 is considered to be highly sensitive to sample size (Cheung & Rensvold, 2002). Specifically, the larger the sample size (generally over 200), the greater the tendency for χ^2 to be significant, whereas with a lower sample size (below 100), the χ^2 test has a tendency to indicate a non-significant probability level. As such, Kline (2005) recommended employing the normed χ^2, which is calculated by dividing the χ^2 value by the degrees of freedom. A normed χ^2 value of less than three (3) has been suggested to indicate a reasonable fit to the data (Bollen, 1989).

In addition to the χ^2 statistic, other incremental fit indices have also been proposed to supplement it, which are said to be designed to avoid the problems associated with sample size as related to the χ^2 test (Bentler & Bonett, 1980). These include the Root-Mean-Square Error of Approximation (RMSEA), Standardised Root-Mean-Square residual (SRMR), Comparative Fit Index (CFI), the Non-Normed Fit Index (NNFI), Tucker-Lewis Index (TLI), Goodness-of-Fit Index (GFI), and the Akaike Information Criterion (AIC). Specifically, an RMSEA value of < 0.05 indicates a good-fitting model (Browne & Cudeck, 1993). For CFI, NNFI, TLI, and GFI, a value > 0.90 is regarded as an acceptable

fit of data, while for the SRMR a value of < 0.01 is considered good fit (e.g., Kline, 2005; Marsh, 2007; Marsh, Hau, & Wen, 2004). The AIC is used to compare a number of competing models and, in these instances, the model which generates the lowest AIC values is regarded as the best fitting model. The actual AIC value is not relevant, although AIC values which are close to zero are considered to be more favourable. When selecting a model out of a number of possibilities, parsimony should be employed, with the simplest model being selected (Bollen & Long, 1993).

If the model's fit is acceptable, this suggests that the proposed relationships or the hypothesized model fits the data. If the model's fit is not adequate, then the model needs to be respecified. However, the respecification needs to grounded in theoretical relevance, as opposed to empirical relevance. Specifically, the respecification of causal relationships needs to be theoretically meaningful and supported by empirical evidence. It should not be empirically guided, as this can result in a good-fitting model in the absence of any theoretical value. Respecification can be conducted in a number of ways. Firstly, non-significant pathways can be deleted or trimmed. Secondly, parameters can be added or deleted in the model to improve the fit. SEM contains modification indices such as the Lagrange Multiplier tests and Wald tests, which provide suggestions for this; however, proceeding with such suggestions should be driven by theory and consistent with the research hypotheses. Through respecification, once a good-fitting model is achieved, ideally the newly formulated model should be tested on a new sample/data.

Is SEM always appropriate for use?

As the research questions being tested have become more complex, there has been a concomitant rise in the demand by reviewers and journal editors for authors to undertake more sophisticated modes of analyses. However, caution must be exercised here as SEM may not be suitable for all research questions. Reviewers and journal editors often want an author to use SEM but do not always understand that it is inappropriate in some cases. In this respect, it is important to clearly understand the nature of the research question being examined, as well as the answers that one would like to generate. Therefore, prior to applying SEM, it is important to consider the strengths and weaknesses of SEM over other multivariate analyses.

Advantages of SEM:

1) Results generated by SEM can provide evidence for causal relationships between variables. However, as SEM is a priori dependent on theory and previous empirical evidence, researchers must be aware and confident of a relationship between the variables (observed and measured) as well as the direction of that relationship. Moreover, such relationship should occur in isolation and not be influenced by other variables (Kline, 2005). However, it is important to note that SEM does not prove causality: rather, it only highlights whether the hypothesized relations or model are consistent with the empirical data.

2) SEM allows researchers to test and compare a number of competing/alternative models, promoting robust theory- building and validation.

3) SEM can test models with multiple dependent and independent variables, as well as mediating and interactive effects.

4) SEM is able to manage with difficult data (e.g., non-normal, incomplete, multi-level and longitudinal data). For example, SEM programs have procedures that are robust against violations of normality (e.g., Maximum Likelihood Estimation), missing data (e.g., multisample analysis; multiple imputation; expectation maximization algorithm; full information maximum likelihood; see Tomarken & Waller, 2005). It can also work with experimental and non-experimental data, as well as continuous, dichotomous, and interval data.

5) SEM uses CFA to partial out measurement error from multiple indicators underlying each latent variable, and therefore subsequently enables the relationships between "error free" latent variables to be tested (Violato & Hecker, 2007).

Disadvantages of SEM:

1) Given that SEM is dependent on theory and previous empirical literature, there is scope for investigators to misinterpret the causal relationships between variables, especially if the model being tested is exploratory, is grounded in weak theory, employs poor research designs, or is guided by ambiguous hypotheses (Violato & Hecker, 2007).

2) SEM is an approximation of reality, in that it omits variables implicated in the model or causal processes to achieve goodness-of-fit (Tomarken & Waller, 2005). In doing so, it can create a misrepresen-

tation of the measurement and/or structural processes, resulting in more biased and/or inaccurate parameter estimates and standard errors (e.g., Reichardt, 2002).

3) SEM is unable to compensate for inadequate psychometric properties of measures (Byrne, 2010; Kline, 2005), in particular measures that are underpinned by poor reliability. The employment of unreliable measures or the use of a single measure to reflect the latent variable is likely to reduce the amount of variability in the latent variable, thus increasing measurement error. Similarly, it cannot compensate for the limitations of the research design nor its methodology.

4) SEM requires a large sample size. A minimum of 200 participants are considered sufficient; however, a rule of thumb of 5–10 participants per indicator has been proposed (e.g., Byrne, 2006). Still, it should be noted that in populations where this is not always feasible, there are ways to overcome the shortage of participants (see the section on parceling above).

5) SEM rejects theory and models on the basis of the global fit statistics. It is possible for the relations between variables to be significant although the model yields a poor fit, thus indicating that the model does not fit the data. Before rejecting the model, researchers should consider checking for errors in data or violations of SEM assumptions. Another proposed method to improve fit indices is to estimate as many parameters as there are data-points (just identified model); however, this renders the data meaningless, explains nothing more about the tested model and, as such, should be avoided (Mulaik et al., 1989). In cases where poor global fit indices persist, researchers can rely on the effect size of the association, confidence intervals, and other lower-order components when evaluating a model (Tomarken & Waller, 2003, 2005). However, researchers should be aware of alternative modes of analyses such as the MACROS developed by Preacher and Hayes (2004, 2008; Hayes, 2013) available on SPSS, which can be used to test for similar relations (e.g., mediation, moderation, temporal patterns, etc.), and are not dependent on fit statistics.

Step 1: Model specification:
Outline and define the research problem

In this step researchers should develop and formulate a research question that is grounded in theory and underpinned by empirical evidence. Moreover, as SEM functions using a priori hypotheses, it is critical that the measurements used to capture and reflect the chosen construct are valid and reliable for use within the given population. Accordingly, based on theory and evidence, researchers should formulate a testable model (or a number of competing testable models). This testable model is then specified. Specifically, the relationships between variables at both the measurement and structural model should be noted.

In the current example, the research problem was aimed at examining the applicability of the components underlining the transdiagnostic cognitive-behavioural theory of eating disorders within an athletic population. (For a more comprehensive outline of the theory and literature, see Shanmugam, Jowett, & Meyer, 2011.) The transdiagnostic cognitive-behavioural theory of eating proposes the mechanisms that cause and maintain eating disorders (be it Anorexia Nervosa, Bulimia Nervosa, or Eating Disorder Not Otherwise Specified) are the same (Fairburn et al., 2003). Specifically, Fairburn et al. postulated that the four core psychopathological processes of clinical perfectionism, unconditional and pervasive low self-esteem, mood intolerance, and interpersonal difficulties all interrelate with the core psychopathology of eating disorders – over-evaluation of eating, shape, weight, and their control – to instigate both the development and the maintenance of the disorder. While their transdiagnostic cognitive-behavioural theory of eating disorders provides a grounded conceptual framework to understand how eating disorders may arise, with relevant evidence to support the associations among its main components within the general population (e.g., Collins & Read, 1990; Dunkley, Zuroff, & Blankstein, 2003; Dunkley & Grilo, 2007; Leveridge, Stoltenberg, & Beesley, 2005; Stirling & Kerr, 2006), there is an observable gap in the scientific understanding of such processes within the athletic population, as well as a poor understanding of the concomitant interrelationships among the processes involved. Thus, the purpose of the present example was to test the main components of Fairburn et al.'s transdiagnostic theory in a sample of athletes to further understand eating psychopathology.

Guided by Fairburn et al.'s (2003) theory and relevant empirical research, the first objective was to test a model that proposed linkages between interpersonal difficulties, clinical perfectionism, self-esteem, depression, and eating psychopathology (see Figure 1). Specifically, it was hypothesized that dispositional interpersonal difficulties as reflected in athletes' insecure attachment styles[1] would negatively affect their perceptions of situational interpersonal difficulties as reflected in the quality of the athletes' relationships with parents and coaches (e.g., decreased perceived support and increased perceived conflict)[2]. It was further hypothesized that poor relationship quality would lead to higher levels of clinical perfectionism (personal standards and self-criticism)[3]. Subsequently, athletes' levels of personal-standards perfectionism was expected to negatively predict their levels of self-esteem, while athletes' levels of self-critical perfectionism were predicted to negatively estimate their levels of self-esteem[4], but to positively predict depressive symptoms[5] and eating psychopathology[6]. Finally, it was hypothesized that athletes' levels of self-esteem would negatively predict their levels of depressive symptoms, which in turn were expected to be positively associated with athletes' eating psychopathology.
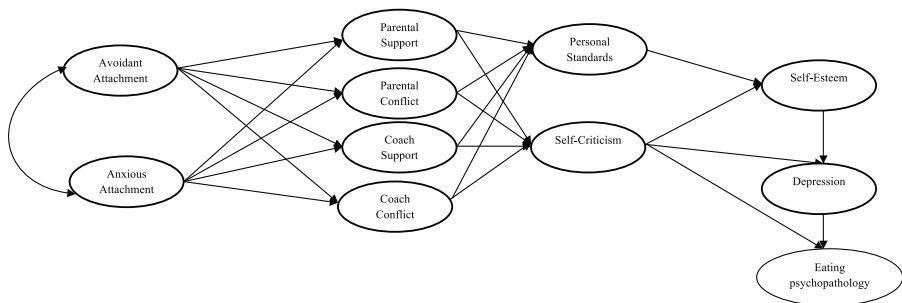

Figure 1. *The hypothesized transdiagnostic cognitive-behavioural model of athletes' eating psychopathology. (from Shanmugam et al., 2011) Copyright Human Kinetics. Reprinted with permission*

Step 2: Model identification: Review model for identification

In this step, the constructed model is reviewed for identification. The process of identification is achieved by establishing the number of observed variables and the number of parameters to be calculated. As previously mentioned, an over-identified model is recommended. Specifically, in the testable model, the number of known data points (i.e., vari-

ances, covariances) should exceed the number of data points that are unknown or being estimated (i.e., factor loadings, measurement error, disturbances, etc.). In the current example, 127 observed items were utilized. Using the recommended 10:1 ratio of participants to observed variables (Byrne, 2006); would have required a total of 1270 athletes. However, only 588 athletes participated in the study. Thus parceling was conducted, following the guidelines of Hau and Marsh (2004) and Nasser-Abu Alhija and Wisenbaker (2006), whereby the observed items were parceled according to the size and direction of their skew per latent variable, thus reducing the number of observed variables to 48. Employing the aforementioned equation, the model identification of the hypothesized model was tested prior to model estimation (see Figure 2), and revealed an over-identified model, with 1057 degrees of freedom.

= P (P+1)/2 information points > parameters to be estimated

= 48 (49)/2 information points > 37 factor loadings, 48 errors,
    31 path coefficients and disturbances, and 1 covariance

= 1176 information points > 119 parameters to be estimated
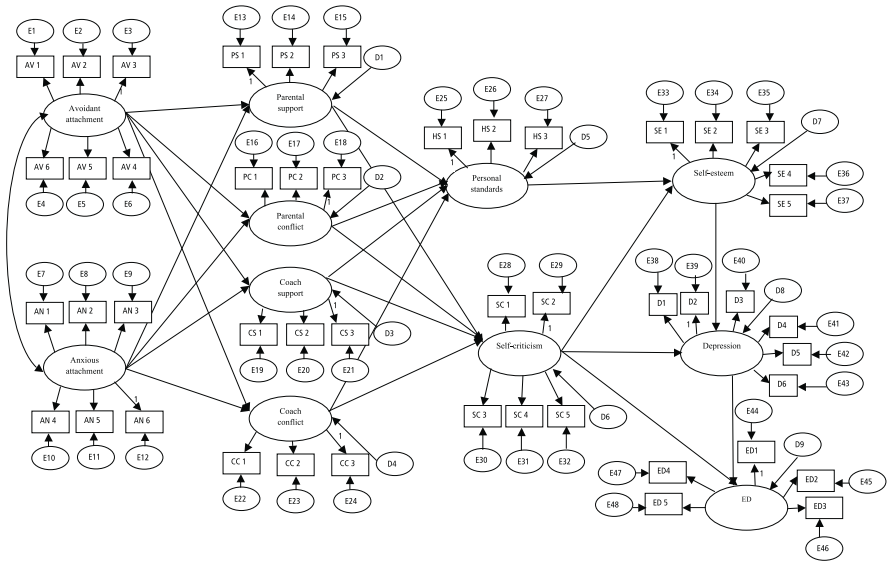
= 1057 dfs

Figure 2. *Model specification of the hypothesized model using the parcelled items. Note 1 = fixed parameter, pointed arrows parameters to be freely estimated, E= error variance, D= disturbance*

Step 3: Model estimation:
how well does the model fit the data?

In this step, the hypothesized model is estimated. One of the advantages of SEM is the number of commercial SEM softwares that are available and regularly updated.

These include and are not limited to AMOS for SPSS (Arbuckle, 2012), EQS (Bentler, 2006), LISREL (Jöreskog & Sörbom, 1996), and MPlus (Muthén & Muthén, 1998–2010). It is beyond the scope of the current paper to provide an overview of the underlying features of each program; thus, readers are directed to Lei and Wu (2007) for an overview.

Prior to model estimation, it is critical that descriptive and univariate analyses are conducted on the collected data to ensure that the data fulfill the assumption of SEM. These include multivariate normality, independence of observations and homoscedasticity (Violato & Hecker, 2007). In the current example, following model specification and identification, the hypothesized model was estimated using the Maximum Likelihood Estimation procedure within EQS 6.0. Due to the violation of multivariate normality, corrections for non-normality were employed and robust statistics were attained. Moreover, only the variables that were significantly correlated to the dependent variable were included in the SEM.

Step 4: Model evaluation and respecification:
Establish the fit of the model to the data

In this step the hypothesized model is determined and evaluated using a number of GOF indices. If the GOF indices obtained are acceptable, this indicates that the hypothesized model fits the data. However, if the GOF indices are not satisfactory, then the model needs to be respecifed. In the current example, the significance of χ^2, the normed χ^2, the Root-Mean-Square Error of Approximation (RMSEA), the Non-Normed Fit Index (NNFI), and the Comparative Fit Index (CFI) were all used to evaluate the fit of the model. GOF indices revealed that the measurement model of the hypothesized model (see Table 1) fit the data well: $\chi^2 = 2159.95$, $df = 1025$, $p < 0.0001$, RMSEA = 0.043 (90% CI = 0.041–0.046), NNFI = 0.92, and CFI = 0.93; with satisfactory factor loadings (see Table 1), and recorded above the recommended value of 0.40 (Ford et al., 1986).

Table 1
Standardised factor loadings from the measurement model

Variable	Factor loading (error)	Variable	Factor loading (error)	Variable	Factor loading (error)
ECR-AV		ECR-ANX		SCL-Depression	
Parcel 1 (3 items)	0.79 (0.62)	Parcel 1 (3 items)	0.69 (0.73)	Parcel 1 (2 items)	0.80 (0.60)
Parcel 2 (3 items)	0.70 (0.71)	Parcel 2 (3 items)	0.65 (0.76)	Parcel 2 (2 items)	0.72 (0.69)
Parcel 3 (3 items)	0.78 (0.63)	Parcel 3 (3 items)	0.70 (0.71)	Parcel 3 (2 items)	0.77 (0.64)
Parcel 4 (3 items)	0.86 (0.52)	Parcel 4 (3 items)	0.89 (0.46)	Parcel 4 (2 items)	0.86 (0.51)
Parcel 5 (3 items)	0.79 (0.61)	Parcel 5 (3 items)	0.75 (0.67)	Parcel 5 (2 items)	0.69 (0.72)
Parcel 6 (3 items)	0.82 (0.58)	Parcel 6 (3 items)	0.86 (0.51)	Parcel 6 (2 items)	0.78 (0.62)
S-SQRI-PS		S-SQRI-CS		S-SQRI-PC	
Parcel 1 (2 items)	0.83 (0.56)	Parcel 1 (2 items)	0.86 (0.50)	Parcel 1 (2 items)	0.82 (0.57)
Parcel 2 (2 items)	0.77 (0.64)	Parcel 2 (2 items)	0.75 (0.66)	Parcel 2 (2 items)	0.90 (0.45)
Parcel 3 (2 items)	0.82 (0.57)	Parcel 3 (2 items)	0.90 (0.44)	Parcel 3 (2 items)	0.77 (0.64)
S-SQRI-CC		FMPS-PS		EDEQ	
Parcel 1 (2 items)	0.78 (0.63)	Parcel 1 (2 items)	0.85 (0.54)	Parcel 1 (5 items)	0.69 (0.72)
Parcel 2 (2 items)	0.90 (0.44)	Parcel 2 (2 items)	0.65 (0.76)	Parcel 2 (5 items)	0.79 (0.62)
Parcel 3 (2 items)	0.91 (0.41)	Parcel 3 (3 items)	0.89 (0.46)	Parcel 3 (5 items)	0.95 (0.31)
				Parcel 4 (4 items)	0.90 (0.43)
				Parcel 5 (4 items)	0.90 (0.44)
RSES		DAS-SC			
Parcel 1 (2 items)	0.78 (0.63)	Parcel 1 (3 items)	0.80 (0.60)		
Parcel 2 (2 items)	0.82 (0.58)	Parcel 2 (3 items)	0.82 (0.57)		
Parcel 3 (2 items)	0.69 (0.72)	Parcel 3 (3 items)	0.88 (0.48)		
Parcel 4 (2 items)	0.79 (0.61)	Parcel 4 (3 items)	0.84 (0.54)		
Parcel 5 (2 items)	0.68 (0.74)	Parcel 5 (3 items)	0.80 (0.60)		

Note: All loadings are significant at the .05 level. ECR-AV= avoidant attachment, ECR-ANX= anxious attachment, S-SQRI-PS= parental support, S-SQRI-PC= parental conflict, S-SQRI-CS= coach support, S-SQRI-CC=coach conflict, FMPS-PS=personal standards perfectionism, DAS-SC= self-critical perfectionism, RSES=self-esteem, SCL-Depression=depression, EDEQ= eating psychopathology

However, the predicted structural model failed to achieve an acceptable goodness-of-fit: χ^2 = 2645.57, df = 1057, p < 0.0001, RMSEA = 0.051 (90% CI = 0.048–0.053), NNFI = 0.89, and CFI = 0.90. Thus the model needed to be respecifed. Guided by the Lagrange Multiplier tests' output, all empirical suggestions that were conceptually and theoretically meaningful were carried out. In particular, removing all the non-significant paths – pathways between Self-Critical perfectionism to Depression and eating psychopathology, parameters associated to Personal Standards and Anxious Attachment and creating a linear pathway

between Parental Support, Coach Support, and Parental Conflict and Coach Conflict, respectively – improved the model fit to ensure an acceptable goodness-of-fit and a parsimonious model.

The fit of the respecified model was χ^2 = 1367.94, *df* = 693, $p < 0.0001$, RMSEA = 0.041(90% CI = 0.038–0.044), NNFI = 0.94, and CFI = 0.94 (see Figure 3). The normed χ^2 value was 1.97 (1367.94/693). Thus, the normed χ^2 value and all the other incremental fit indices provide good support for the final model.

As shown in Figure 3, in the current example, avoidant attachment was associated with poor quality relationships (characterized by decreased perceived support and increased perceived conflict) with their influential parent and principal coach. Moreover, higher levels of conflict in their parent–athlete and coach–athlete relationships were related to higher levels of self-criticism. High levels of self-criticism were related to low self-esteem and feeling of worthlessness. Subsequently, low self-esteem was linked to higher depressive symptoms, which in turn were linked to elevated eating psychopathology. The findings also suggested the same processes that are likely to lead to elevated eating psychopathology are also likely to prevent it. In particular, secure attachment was associated with high quality parent-athlete and coach–athlete relationships, resulting in low levels of self-criticism, which in turn was associated with higher levels of self-esteem.
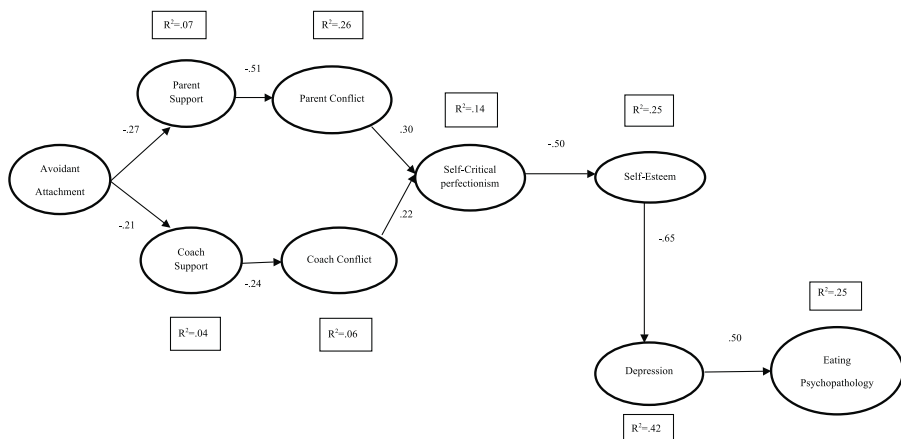

Figure 3. *A structural representation of the transdiagnostic cognitive behavioural model of athletes' eating psychopathology. The standardized coefficients presented are significant at .05 level. Taken from Shanmugam et al. (2011). Copyright Human Kinetics reprinted with permission*

Subsequently, high levels of self-esteem were associated with low levels of depression, which in turn was linked to healthy eating. Collectively, these findings are consistent with the assumptions of the transdiagnostic cognitive-behavioural theory and with previous findings that have linked avoidant attachment (e.g., Ramacciotti et al., 2001), poor quality relationships (e.g., McIntosh, Bulik, McKenzie, Luty, & Jordan, 2000), low levels of self-esteem (e.g., Shea & Pritchard, 2007), high levels of self-critical perfectionism (e.g., Dunkley, Blankstein, Masheb, & Grilo, 2006), and depression (e.g., Stice & Bearman, 2001) to disturbed eating behaviors.

Conclusion

The aim of this article was to provide an overview of SEM and to complement this with a worked empirical example. The dichotomy of models (one constituting a theorized organization of indicator variables and how they identify the latent variables, and another referring to the relationships between the latent variables) and the sequential steps involved in theory and model testing, including model specification, model identification, model estimation and model evaluation was outlined. SEM is a theory-strong approach underpinned by established research methods but must be used with caution. As a prerequisite for the proper use of SEM, a substantial base of empirical evidence must exist, combined with strong conceptual understanding of the theory relevant to the research question and access to large samples that may be difficult to access. Skills-training is also a necessity, such that researchers can begin to apprehend the advanced theoretical and statistical methods required to test complex, integrated theoretical models within the social sciences.

NOTES

1. Attachment styles were measured through the Experiences of Close Relationships (ECR; Brennan, Clark, & Shaver, 1998). This is a 36-item questionnaire which forms two subscales: anxious attachment and avoidant attachment.

2. Social support and interpersonal conflict were measured through the Sport-Specific Quality of Relationship Inventory (S-SQRI; Jowett, 2009). Both subscales contain six items and were used to capture the nature of the parent–athlete and coach–athlete relationships.

3. Personal standards perfectionism was measured through the seven-item personal standards subscale of the Multidimensional Perfectionism Scale (FMPS; Frost, Marten, Lahart, & Rosenblate, 1990) and self-critical perfectionism was measured through the 15-item self-critical perfectionism subscale of Dysfunctional Attitude Scale (DAS; Weissman & Beck, 1978).

4. Self-esteem was measured through the 10-item Rosenberg's Self Esteem Scale (RSES; Rosenberg, 1965).

5. Depressive symptoms were measured through an adapted 12-item Symptom Checklist 90R (SCL-90; Derogatis, 1983).

6. Eating psychopathology was measured through the 22-item Eating Disorder Examination Questionnaire (EDEQ; Fairburn & Beglin, 2008).

REFERENCES

Arbuckle, J. L. (1995-2011). *AMOS 20.0 User's Guide*. Crawfordville, FL: AMOS Development Corporation.

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9*, 78–102. doi: 10.1207/S15328007SEM0901_5.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumaker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269-296). Mahwah, NJ: Lawrence Erlbaum Associates.

Bentler, P. M. (1980). Multivariate analyses with latent variables: Causal modelling. *Annual Review of Psychology, 31*, 419-456. doi: 10.1146/annurev.ps.31.020180. 002223.

Bentler, P. M. (2006). *EQS: Structural equations program manual.* Encino, CA: Multivariate Software Inc.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and the goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*, 588-606. doi: 10.1037// 0033-2909.88.3.588.

Bentler, P. M., & Chou, C-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*, 78-117. doi: 10.1177/0049124187016001004.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.*), Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage. doi: 10.1177/0049124192021002005.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley. doi: 10.1002/9781118619179.

Bollen, K. A., & Long, J. S. (1993). *Testing Structural Equation Models.* Newbury Park, CA: Sage.

Brennan, K. A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement of adult romantic attachment: An integrative overview. In J. A. Simpson & W. S. Rholes (Eds.), *Attachment theory and close relationships* (pp. 46-76). New York, NY: Guildford Press.

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming*. Mahwah, NJ: Laurence Erlbaum Associates.

Byrne, B. M. (2010). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming* (2nd ed.). New York, NY: Routledge Academic.

Cattell, R. B., & Burdsal, C. A. (1975). The radial parcel double factoring design: A solution to item-vs.-parcel controversy. *Multivariate Behavioral Research, 10*, 165-179. doi: 10.1207/s15327906mbr1002_3.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. doi: 10.1207/ S15328007SEM0902_5.

Collins, N. L., & Read, S. J. (1990). Adult attachment, working models and relationship quality in dating couples. *Journal of Personality and Social Psychology, 58*, 644–663. doi: 10.1037/0022-3514.58.4.644.

Derogatis, L. R. (2000). Symptom Checklist-90-Revised. In *Handbook of Psychiatric Measures*. Washington, DC: American Psychiatric Association.

Dunkley, D. M., Blankstein, K. R., Masheb, R. M., & Grilo, C. M. (2006). Personal standards and evaluative concerns dimensions of 'clinical' perfectionism: A reply to Shafran et al. (2002, 2003) and Hewitt et al. (2003). *Behaviour Research and Therapy, 44*, 63-84. doi: 10.1016/j.brat.2004.12.004.

Dunkley, D. M., & Grilo, C. M. (2007). Self-criticism, low self-esteem, depressive symptoms and over-evaluation of shape and weight in binge eating disorder patients. *Behaviour Research and Therapy, 45*, 139-149. doi: 10.1016/j.brat.2006.01.017.

Dunkley, D. M., Zuroff, D. C., & Blankstein, K. R. (2003). Self-critical perfectionism and daily affect: Dispositional and situational influences on stress and coping. *Journal of Personality and Social Psychology*, *84*, 234–252. doi: 10.1037/0022-3514.84.1.234.

Fairburn, C. G., & Beglin, S. J. (2008). Eating Disorders Examination Questionnaire (EDE-Q 6.0). In C. G. Fairburn (Ed.), *Cognitive behaviour therapy and eating disorders* (pp. 309-313). New York, NY: The Guildford Press.

Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behaviour therapy for eating disorders: A "transdiagnostic" theory and treatment. *Behaviour Research and Therapy, 41*, 509-528.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology, 39*, 291-314. doi: 10.1111/j.1744-6570.1986.tb00583.x.

Frost, R. O., Marten, P., Lahart, C., & Rosenblate, R. (1990). The dimensions of perfectionism. *Cognitive Therapy and Research*, 14, 449-468. doi: 10.1007/BF01172967.

Hau, K. T., & Marsh, H. W. (2004). The use of item parcels in structural equation modeling: Non-normal data and small sample sizes. *British Journal of Mathematical Statistical Psychology*, *57*, 327–351. doi: 10.1111/j.2044-8317.2004.tb00142.x.

Hall, R. J., Snell, A. F., & Foust, M. S. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods*, *2*, 233-256. doi: 10.1177/109442819923002.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression based approach*. New York, NY: The Guilford Press.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal Analysis: Assumptions, Models, and Data*. Beverly Hills, CA: Sage.

Jowett, S. (2009). Validating the coach athlete relationship measures with the nomological network. *Measurement in Physical Education and Exercise Science, 13*, 34-51. doi: 10.1080/10913670802609136.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's Reference Guide*. Lincolnwood, IL: Scientific Software International.

Karadag, E. (2012). Basic features of structural equation modeling and path analysis with its place and importance in educational research methodology. *Bulgarian Journal of Science and Education Policy (BJSEP), 6*, 194-212.

Kishton, J. M., & Widaman, K. F. (1994). Unidimensional versus domain representative parcelling of questionnaire items: An empirical example. *Educational and Psychological Measurement*, *54*, 757-765. doi: 10.1177/0013164494054003022.

Kline, R. B. (2005). *Principles and practice of structural equation modeling.* New York, NY: The Guilford Press.

Lei, P. W., & Wu, Q. (2007). An NCME Instructional Module on Introduction to Structural Equation Modeling: Issues and Practical Considerations. *Educational Measurement, Issues and Practice, 26*, 33-44.

Leveridge, M., Stoltenberg, C. D., & Beesley, D. (2005). Relationship of attachment style to personality factors and family interaction patterns. *Contemporary Family Therapy, 27*, 577-597. doi: 10.1007/s10591-005-8243-9.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to Parceling and CFA Tests 9 parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151-173.

McIntosh, V., Bulik, C., McKenzie, J., Luty, S., & Jordan, J. (2000). Interpersonal psychotherapy for anorexia nervosa. *The International Journal of Eating Disorders, 27*, 125–139. doi: 10.1002/(SICI)1098-108X(200003)27:2%3C125::AID-EAT1%3E3.0. CO; 2-4.

Marsh, H. W. (2007). Application in confirmatory factor analysis and structural equation modeling in sport and exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (pp. 774-798). New York, NY: Wiley. doi: 10. 1002/9781118270011.ch35.

Marsh, H. W. & Hau, K-T. (1999). Confirmatory factor analysis: Strategies for small sample sizes. In R.H. Hoyle (Eds.), *Statistical strategies for small sample research* (pp. 251-306). Newbury, CA: Sage.

Marsh, H. W., Hau, K-T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*, 181-220. doi: 10.1207/s15327906mbr3302_1.

Marsh, H. W., Hau, K-T., & Wen, Z. (2004). 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, 11*, 320-341.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, *105*, 430–435. doi: 10.1037//0033-2909.105.3.430.

Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.

Nasser, F., & Wisenbaker, J. (2003). A Monte Carlo study investigating the impact of item parcelling on measures of fit in confirmatory factor analysis. *Educational and Psychological Measurement, 63*, 729–757. doi: 10.1177/0013164403258228.

Nasser-Abu Alhija, F., & Wisenbaker, J. (2006). A Monte Carlo study investigating the impact of item parceling strategies on parameter estimates and their standard errors in CFA. *Structural Equation Modeling, 13*, 204–228.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717-731. doi: 10.3758/BF03206553.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and re-sampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*, 879-891. doi: 10.3758/BRM.40.3.879.

Pugesek, B. H., Tomer, A., & von Eye, A. (2003). *Structural Equation Modeling: Applications in ecological and evolutionary biology*. Cambridge, UK: Cambridge University Press.

Ramacciotti, A., Sorbello, M., Pazzagli, A., Vismara, L., Mancone, A., & Pallanti, S. (2001). Attachment processes in eating disorders. *Eating and Weight Disorders, 6*, 166-170. doi: 10.1007/BF03339766.

Reichardt, C. S. (2002). The priority of just-identified, recursive models. *Psychological Methods, 7*, 307-315. doi: 10.1037/1082-989X.7.3.307.

Rosenberg, M. (1965). *Society and the adolescent self image*. Princeton, NJ: Princeton University Press.

Russell, D. W., Kahn, J. H., Spoth, R., & Altmaier, E. M. (1998). Analyzing data from experimental studies: A latent variable structural equation modeling approach. *Journal of Counseling Psychology, 45*, 18–29. doi: 10.1037/0022-0167.45.1.18.

Schumacker, R. E. & Lomax, R. G. (2004). *A beginner's guide to Structural Equation Modeling* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Shanmugam, V., Jowett, S., & Meyer, C. (2011). Application of the transdiagnostic cognitive-behavioral model of eating disorders to the athletic population. *Journal of Clinical Sport Psychology, 5*, 166-191.

Shea, M. E., & Pritchard, M. E. (2007). Is self-esteem the primary predictor of disordered eating? *Personality and Individual Differences, 42*, 1527–1537. doi: 10.1016/j.paid.2006.10.026.

Shipley, B. (2000). *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference*. Cambridge, UK: Cambridge University Press.

Stice, E., & Bearman, S. K. (2001). Body image and eating disturbances prospectively predicts increases in depressive symptoms in adolescent girls: A growth curve analysis. *Developmental Psychology, 37*, 597–607. doi: 10.1037//0012-1649.37.5.597.

Stirling, A. E., & Kerr, G. A. (2006). Perfectionism and mood states among recreational and elite Athletes. *Athletic Insight: The Online Journal of Sport Psychology, 8*, 13-27.

Teo, T., & Khine, M. S. (2009). *Structural Equation Modeling in Educational Research: Concepts and Applications*. Rotterdam, Netherlands: Sense Publishers.

Thompson, B., & Melancon, J. (1996, November). *Using item "testlets/parcels" in confirmatory factor analysis: An example using the PPDP-78*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Tuscaloosa, AL.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with «well fitting» models. *Journal of Abnormal Psychology, 112*, 578-5. doi: 10.1037/0021-843X.112.4.578.

Tomarken, A. J., & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology, 1*, 31-65. doi: 10.1146/annurev.clinpsy.1.102803.144239.

Violato, C., & Hecker, K. G. (2007). How to use structural equation modeling in medical education research: A brief guide. *Teaching and Learning in Medicine*, *19*, 362–371. doi: 10. 1080/10401330701542685.

Weissman, A. N., & Beck, A. T. (1978, August/September). *Development and validation of the Dysfunctional Attitude Scale: A preliminary investigation*. Paper presented at the 86th Annual Convention of the American Psychological Association, Toronto, Ontario, Canada.

Wright, S. S. (1921). Correlation and causation. *Journal of Agricultural Research, 20*, 557–85.

Yang, C., Nay, S., & Hoyle, R. H. (2010). Three approaches to using lengthy ordinal scales in structural equation models: Parceling, latent scoring, and shortening scales. *Applied Psychological Measurement, 34*, 122-142.

Yuan, K-H., Bentler, P. M., & Kano, Y. (1997). On averaging variables in a CFA model. *Behaviormetrika, 24*, 71-83.