The Mental Toughness Questionniare-48: A Re-examination of Factorial Validity

Phil D. J. Birch[1], Simon Crampton[2], Iain Greenlees[1], Ruth Lowry[1], Pete Coffee[3]

[1] Department of Sport and Exercise Sciences, University of Chichester, Bishop Otter Campus, Chichester, PO19 6PE, United Kingdom.

[2] Performance Psychology, English Institute of Sport, Bisham Abbey, Marlow, SL7, UK

[3] Faculty of Health Sciences and Sport, University of Stirling, Stirling, FK9 4LA, UK.

Corresponding author information:

Phil Birch,

P.Birch@chi.ac.uk,

Department of Sport and Exercise Sciences, University of Chichester,

Bishop Otter Campus, Chichester, PO19 6PE, United Kingdom.

Abstract

The Mental Toughness Questionnaire-48 (MTQ48; Clough, Earle, & Sewell, 2002) is the most utilised instrument to measure Mental Toughness in sport (Gucciardi, Hanton, & Mallett, 2012). To date, preliminary research (Gucciardi et al., 2012; Perry et al., 2013) examining the factorial validity of the MTQ48 in athlete samples has yielded equivocal findings. The aim of this paper was to re-examine the factorial validity of the four- and six-factor models of the MTQ48 in moderate (*n* = 480) and large (*n* = 1206) independent student athlete samples. Using confirmatory factor analyses, findings revealed little support for the hypothesised models of the MTQ48 in both samples. The results support those found by Gucciardi et al. and Perry et al. and suggest that, in its current form, the MTQ48 may not be a valid measure of the 4/6Cs model of mental toughness when using student athletes.

The Mental Toughness Questionniare-48: A Re-examination of Factorial Validity

Introduction

Mental toughness (MT) is a term frequently used by athletes, coaches and the media to differentiate "good" and "great" athletes in their pursuit of sporting excellence (Gucciardi, Gordon, & Dimmock, 2008). The concept of MT has received widespread attention in the literature, with the majority of research using qualitative designs to explore the perspectives of key stakeholders (i.e., athletes, coaches, sport psychologists) to define and conceptualise MT in sport (Sheard, 2012). Although previous qualitative explorations have provided a valuable platform to enhance our understanding of MT, the development of psychometrically sound questionnaires are crucial as they serve to validate conceptualisations of a given concept (Marsh, 2002). Despite a plethora of instruments being developed to measure the varying conceptualisations of MT (for a review, see Gucciardi, Mallet, Hanrahan, & Gordon, 2011), the measurement of the concept remains a contentious issue.

The most used instrument to measure MT in both sport and non-sport contexts is the Mental Toughness Questionnaire 48 (MTQ48: Clough, Earle, & Sewell, 2002). The MTQ48 is underpinned by the 4C's model of MT (Clough et al.) which incorporated the views and experiences of athletes and coaches with the established psychological theory of hardiness (Kobasa, 1979). Clough et al.'s 4Cs model proposes that MT resembles tenets outlined in hardiness theory (Kobasa) where a combination of dispositional attitudes (challenge, commitment, control) are thought to motivate one to respond to stressors with specific coping and social interaction efforts which facilitate resiliency by turning potential disasters into opportunities (e.g., Kobasa; Maddi, 2002; Maddi & Kobassa, 1984). However, Clough et al. suggested that hardiness alone did not fully encapsulate MT and added a fourth attitude, confidence, to account for the physical and mental demands of competitive sport. The addition of confidence was postulated to suitably transpose the health-related construct of hardiness into the more sport-specific concept of MT.

On-going development of the MTQ48 has resulted in the 6Cs model of MT. Earle (2006) posited that MT is best understood when the control and confidence constructs are subdivided into two nested components. Specifically, Earle conducted 12 interviews with a variety of sports people to explore the make-up of MT. These included three rugby coaches, one rugby chief executive, two rugby players, two golfers, two footballers, and two squash players. Although findings showed that most themes could be categorised under the construct of confidence and Kobasa's (1979) model of hardiness, Earle suggested that control and confidence have a more complex structure with control (control - emotion, control - life) and confidence (confidence - ability, confidence - interpersonal) having two nested components. This resulted in the 6Cs model of MT.

Although the MTQ48 represents a potentially promising tool for use in the assessment of MT, little evidence of the instrument's psychometric properties were published at its conception (see Clough et al., 2002). In the development of the MTQ48, Clough et al. reported an overall test-retest coefficient of 0.90, with internal consistency of the subscales reported as 0.73, 0.71, 0.71 and 0.80 for control, commitment, challenge and confidence, respectively. Construct validity of the MTQ48 was inferred by convergent and criterion validity. Convergent validity was evidenced through significant relationships between overall MT and optimism (r = .48, p < .01), self-image (r = .42, p < .05), life satisfaction (r = .56, p < .01), self-efficacy (r = .68, p < .01) and stability (r = .57, p < .01). Criterion validity was evidenced through two studies. The first study examined the relationship between MT and perceived effort and found that although there were no differences in perceived exertion at the 30% workload level, participants who were high in MT reported significantly lower levels of perceived exertion at the 70% workload level than participants low in MT. In the second study, Clough et al. investigated the effect of feedback on performance in participants who exhibited low and high levels of MT and found a significant interaction (F = 4.36, p < 0.05) between MT and feedback. Specifically, performance on the cognitive planning task of participants high in MT did not significantly change following positive and negative feedback, whereas participants low in MT performed significantly worse following negative feedback than following positive feedback.

Research using the MTQ48 has generally provided support for its convergent validity given the significant correlations between MT and a range of related psychological variables including coping style (e.g., Nicholls, Polman, Levy, & Blackhouse, 2008), coping effectiveness (e.g., Kaiseler et al., 2009), and coping self-efficacy (e.g., Nicholls, Levy, Polman, & Crust, 2011), leadership preference (Crust & Azadi, 2009), psychological skill usage (Crust & Azadi, 2010), risk taking (Crust

& Keegan, 2010), dispositional flow (Crust & Swann, 2011a), and other measures of MT (Crust & Swann, 2011b). Research by Crust and Clough (2005) has also demonstrated support for its criterion validity given the significant correlations between MT and physical endurance (for a review, see Gucciardi, Hanton, & Mallett, 2012).

Despite its widespread research utilisation and the growing body of evidence to support the MTQ48s convergent validity, some researches have raised concerns regarding its suitability and use in further research (e.g., Connaughton & Hanton, 2009; Gucciardi et al., 2011; Gucciardi et al., 2012). One of the most pronounced discussed limitations regarding the development of the MTQ48 was the lack of reported information to examine its factorial validity. According to Gignac (2009) and Marsh, Martin and Jackson (2010), it is important to ascertain factorial validity before any other forms of validity (such as convergent) are examined and established. Horsburgh, Schermer, Veselka and Vernon (2009) acknowledged that early validation research did not report factor analytic techniques on the MTQ48 and consequently conducted exploratory and confirmatory factor analyses. Their findings reported four factors present, corresponding to control, commitment, challenge and confidence and suggested that the four-factor solution provided a better fit to the data than did a single factor. However, this study could have been strengthened by providing evidence of the psychometric procedures conducted and reporting empirical data (i.e., fit indices, parameter estimates) to support conclusions.

Gucciardi et al. (2012) presented results of the first empirical evaluation of the psychometric properties (i.e. model fit, parameter estimates) of the MTQ48. Using Confirmatory Factor Analysis (CFA) and Exploratory Structural Equation Modelling (ESEM) techniques, Gucciardi et al. did not find support for the hypothesised correlated four-factor model of MT in an independent athlete (*n* = 686) and workplace sample (*n* = 639). In both samples the CFA revealed that the hypothesised correlated four-factor model of the MTQ48 was unsatisfactory, according to the multiple indices of model fit; athlete sample $\chi^2(1074) = 5511.88$, $p < .001$, CFI = .487, TLI = .462, SRMR = .104, RMSEA = .078, 90% confidence interval [CI] [.076, .080] workplace sample $\chi^2(1074) = 4928.95$, $p < .001$, CFI = .521, TLI = .497, SRMR = .093, RMSEA = .075, 90% CI [.073, .077] (see respective Statistical Analysis sections for a description of fit statistics). In addition to the poorly fitting models, the solution was improper, as indicated by factor correlations between control and confidence dimensions that exceed 1.0 in both samples (Blunch, 2008). ESEM supported the findings yielded by CFA in that model fit for the correlated four-factor solution was unsatisfactory for the athlete (x^2 (942) = 2970.25, p < 0.001, CFI = 0.766, TFI = 0.719, SRMR = 0. 045, RMSEA = 0.056, 90% CI [0.054, 0.058]) and workplace sample (x^2 (942) = 2744.20, p < 0.001, CFI = 0.776, TFI = 0.732, SRMR = 0.045, RMSEA = 0.055, 90% CI [0.052, 0.057]). Collectively, CFA and ESEM model fit indices and parameter estimates did not support the hypothesised correlated four-factor model of the MTQ48 in either the athlete or workplace samples. Gucciardi et al. also tested a variety of other MTQ48 models including the six-factor solution and found that these models were unsatisfactory (see Gucciardi et al.'s supplementary online materials). Gucciardi et al.'s analyses therefore suggest that, in its current form, the instrument may not be a valid measure of the 4/6Cs model of MT which it intends to capture.

In an attempt to provide support for the factorial validity of the MTQ48, Perry et al. (2013) present the findings of model fit analyses using CFA and ESEM. Participants (n = 8207) consisted of managers, clerical/administrative workers, athlete and student samples. Whilst stating support for the factorial validity of the MTQ48, with the six-factor model being superior to the four-factor and single-factor models, closer inspection shows that Perry et al. refrain from referring to acceptable or not acceptable model fit. This is possibly due to the fact that the CFA fit indices presented do not reach the proposed acceptable levels (see Byrne, 2013). What is most concerning is that the weakest model fit was the athlete sample (CFA; $\chi 2(1065) = 2535.4$, p < .001, CFI = .771, TLI = .758, SRMR = .063, RMSEA = .056, 90% CI = .053–.059; and ESEM; $\chi 2(855) = 1354.8$, p < .001, CFI = .922, TLI = .897, SRMR = .031, RMSEA = .036, 90% CI = .033–.040) which is the population that the measure was primarily intended for. Although, Perry et al. stated that the weakest model fit was with the athlete sample, these findings coupled with those obtained by Gucciardi et al. (2012) do appear to cast doubt regarding the factorial validity of the MTQ48 when using athlete samples.

The Present Study

Given the equivocal findings obtained by Gucciardi et al. (2012) and Perry et al. (2013), further research is required to test the factorial validity of the MTQ48 in independent athlete samples

(Marsh, 2007; MacKenzie, Podsakoff, & Podsakoff, 2011). Re-examinations of an instrument's factor structure are an important consideration in testing the robustness of theoretical models, especially when testing models across different populations to those used in the initial validation of an instrument (Gucciardi et al.). In addition, previous research by Gucciardi et al. has emphasised the need to re-examine the MTQ48s factorial validity on a larger sample of athletes to further enhance the understanding of the MTQ48s adequacy in capturing the 4/6C's model of MT.

The use of multi-study articles is becoming increasingly popular within the sport and exercise psychology literature. Upon completion of their respective first PhD studies, the first and second authors of this manuscript identified distinct similarities in their work from conference submissions (Crampton, 2010; Birch, Greenlees, Lowry, & Coffee, 2012) and convened to compile a multi-study manuscript from their existing research findings. Study 1, conducted by Simon Crampton, provides a factorial analysis of the hypothesised four- and six-factor models of the MTQ48 in a large sample of student athletes. Study 2, conducted by Phil Birch, Iain Greenlees, Ruth Lowry and Pete Coffee, also provides a factorial analysis of the hypothesised four- and six-factor models of the MTQ48 but represents an extension to Study 1 in a number of ways. First, Study 2 used a substantially larger sample of student athletes which directly addresses concerns raised by Gucciardi et al. (2012). This permitted the use of a range of estimation methods to examine the factorial validity of the MTQ48 and the use of scale refinement techniques (i.e., modification indices) to examine model re-specification protocols (Byrne, 2010; Tabachnick & Fidell, 2001). Finally, given that MT is deemed to be a desirable characteristic to possess and project (Gucciardi & Gordon, 2009) and that socially desirable responding is argued to threaten participant response validity (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003), Study 2 measured social desirability to assess the degree to which participants responded to the MTQ48 in a socially desirable manner.

Therefore, the aim of this paper was to report the findings of two, independently conducted research studies that both examined the factorial validity of the four- and six-factor models of the MTQ48 in two large student athlete samples.

Study One

Method

Participants

Participants were (*n* = 480) competitive student athletes from three Universities in England and consisted of 298 males and 178 females (*M age* = 20.06 years, *SD* = 2.52) with a mean of 9.33 years (*SD* = 4.57) competitive playing experience in their primary sport. Four student athletes did not report their gender. Participants competed in both individual sports (*n* = 137) such as tennis, athletics, and judo and team sports (*n* = 327) such as football, hockey and cricket. Participants were competing at club (*n* = 324), county (*n* = 72), national (*n* = 55) and international (*n* = 13) level at the time of the study.

Measures

Mental Toughness Questionnaire-48 (Clough et al., 2002). The MTQ48 is a 48-item inventory which requires responses to statements on a 5-point Likert scale ranging from (1) *strongly disagree*, to (5) *strongly agree*. The MTQ48 measures six subscales of challenge (8 items), commitment (11 items), control – emotion (7 items), control – life (7 items), confidence – abilities (9 items) and confidence – interpersonal (6 items). Example items include "I usually enjoy a challenge" (challenge); "I usually find something to motivate me" (commitment); I tend to worry about things well before they actually happen" (control - emotion); I generally feel that I am in control of what happens in my life (control - life); "I generally feel that I am a worthwhile person" (confidence - abilities); "I usually take charge of a situation when I feel it is appropriate" (confidence - interpersonal). Consent was obtained from the authors prior to conducting this study.

Procedure

Initial recruitment was conducted via personal communication, letter and email invitation to program convenors. Participants who described themselves as athletes currently competing in sport were provided with a consent form, athlete demographic questionnaire, and the MTQ48. Participants were recruited over a period of six months. The MTQ48 was distributed to student athletes during lectures and seminar classes, and were completed in the presence of the author (or a fully briefed assistant) so that any questions could be answered. The MTQ48 took approximately 10-15 minutes to complete. Prior to completing the MTQ48, participants were assured of confidentiality and anonymity in responses, and informed of their right to withdraw participation at any point prior to obtaining their consent. Institutional ethical approval was obtained prior to data collection.

Statistical Analyses

Data were screened for missing responses using SPSS (Version 18) and no missing values were identified. The univariate skewness values of the MTQ48 items ranged from -1.032 to .527 and the univariate kurtosis values ranged from -1.033 to 3.812 suggesting that in general, the items do not fall within the acceptable criteria (i.e., -0.179 to 0.179; Doane & Seward, 2011) and are indicative of nonnormal data (see Table 1). Mardia's normalised coefficient of multivariate kurtosis indicated that the data departed from multivariate normality (coefficient = 302.38). Some researchers have suggested using the robust maximum likelihood method in relation nonnormal data and the use of categorical variables when there are at least four or more response categories (e.g., Beauducel & Herzberg, 2006; Chou, Benter, & Satorra, 1991; Dolan, 1994; Muthén & Kaplan, 1985). All CFAs in Study 1 were conducted on EQS 6.1 for Windows (Bentler, 2006) using the Robust Maximum Likelihood (MLR) estimation procedure. Specifically, the MLR estimation method affords a robust chi-squared (χ^2) statistic called Satorra-Bentler scale statistic (S-B χ^2; Satorra & Bentler, 1994) and robust parameter standard errors (Bentler & Dijkstra, 1985).

CFA is commonly used to examine patterns of interrelationships among a variety of constructs and is widely considered a robust test of factorial validity (Raykov & Marcoulides, 2000). CFA is achieved by assessing the fit between the reproduced covariance matrix (Σ) and the observed covariance matrix (S). CFA (unlike Exploratory Factor Analysis) is underpinned by a strong theoretical foundation that enables the researcher to specify a factor model in advance and subsequently force items to load on specific factors (Jöreskog & Sorbom, 1996; Schutz & Gessaroli, 1993). Model parameters were identified in accordance with Bentler's (1995) six rules for model specification. To identify the scale of a measurement model, one of the parameter estimates for each latent construct was fixed to 1.0 as to enable model estimation to function effectively. No cross-loading of items were postulated and all factors were allowed to correlate freely.

Many researchers (e.g., Byrne, 2013; Hoyle & Panter, 1995; Kline, 1998; Tanaka, 1993) have suggested using multiple measures of fit indices to provide a more accurate model evaluation process.

Multiple fit indices were used to evaluate the overall fit of the proposed models of the MTQ48. In Study 1, the overall χ^2 statistic, the CFI, the Bentler-Bonett non-normed fit index (NNFI; Bentler & Bonett, 1980), the Standardised Root Mean Squared Residual (SRMR; Hu & Bentler, 1999; Tabachnick & Fidell, 2001), the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990) and Akaike's (1987) Information Criterion (AIC) were used to facilitate subjective model comparisons. These fit indices included measures from four different classes (descriptive fit, absolute fit, absolute fit with penalty function, alternative fit).

The choice of cut-off criteria used to evaluate model adequacy is a contentious issue. Although some researchers have advocated the use of more conservative thresholds to evaluate model fit (e.g., Hu & Bentler, 1999; Russell, 2002), others have emphasised the need to use more liberal guidelines (e.g., Marsh et al., 2004). Specifically, liberal thresholds indicate adequate model fit whereas conservative thresholds indicate good model fit. For the overall χ^2 statistic, values lower than 5.00 reflect adequate fit with values closer to 1.00 reflecting good model fit (Marsh & Hocevar, 1985). CFI values close to 0.90 reflect adequate fit (Bentler, 1992) and values close to 0.95 reflect good model fit (Hu & Bentler). Values on the NNFI that are greater than 0.90 are generally taken to reflect adequate fit to the data (Bentler) although Hu and Bentler suggest a value of 0.95 might be more desirable and reflect good model fit. For the SRMR, values close to 0.08 indicate adequate fit (Browne & Cudeck, 1993) and values of 0.05 or less indicate a good fit (Steiger), and RMSEA below 0.08 indicate adequate fit (Browne & Cudeck) and values below 0.05 indicate good fit. For completeness, the 90% confidence intervals (90% CI) are provided for the RMSEA. Confidence intervals closely surrounding the RMSEA statistic are indicative of good model fit. For the AIC, lower values are indicative of well-fitting models (Raykov & Marcoulides, 2000). The strength of an item is indicated by high parameter estimates and low standard errors. Comrey and Lee (1992) suggested that parameter estimates higher than 0.71 are excellent, 0.63 very good, 0.55 good, 0.45 fair and 0.32 poor.

Results

Descriptive Statistics

Inter-factor correlations of the MTQ48 were weak to moderate suggesting that the factors represent related yet independent components of the 4/6Cs model of MT. Correlations between the non-nested factors (i.e., control and confidence) and their subordinate nested factors (i.e., control – emotion, control – life, confidence – abilities, confidence – life) were moderate as theoretically expected. For the related four-factor solution, the MTQ48 demonstrated adequate internal reliability for the subscales of commitment and confidence, but not for challenge and control ($\alpha < .70$; Nunnally & Bernstein, 1994). For the related six-factor solution, the MTQ48 demonstrated adequate internal reliability for the subscales of commitment, confidence – abilities and confidence – interpersonal, but not for challenge, control – emotion and control – life.

INSERT TABLE 1 HERE

INSERT TABLE 2 HERE

Confirmatory Factor Analysis

Results of the CFAs revealed that the hypothesised related four- and six-factor solutions of the MTQ48 were unsatisfactory when using both the conservative and liberal model fit thresholds. Fit statistics revealed that the related four-factor solution was inadequate S-B χ^2 (1074) = 2599.046, $p < .001$, RCFI = .623, RNNFI = .604, SRMR = .070, RMSEA = .054, 90% CI [.052, .057], likewise the related six-factor model was not optimal either S-B χ^2(1065) = 2301.866, $p < .001$, RCFI = .694, RNNFI = .676, SRMR = .067, RMSEA = .049, 90% CI [.046, .052]. Despite the results indicating inadequate fit for the related four- or the related six-factor solutions, the results from the RMSEA and AIC provide additional support to suggest that the related six-factor solution is superior when compared to the related four-factor model. Inspection of the parameter estimates revealed a large degree of inconsistency between the hypothesised structure and the current data. Parameter estimates did not support the hypothesised solutions with only two items considered very good (4.16%), five items good (10.42%), 17 items fair (35.42%) and 24 items considered poor (50%) on the related four-factor solution and only four items considered very good (8.33%), 12 items good (25%), 14 items fair (29.17%) and 18 items considered poor (37.5%) on the related six-factor solution. Collectively, CFA model fit indices and parameter estimates did not support the related four- or related six-factor solutions of the MTQ48.

Study Two

Method

Participants

Participants were (*n* = 1206) competitive student athletes from six Universities in the United Kingdom. These six Universities were independent from those used in Study 1. Twenty two participants did not fully complete the MTQ48 and were subsequently removed from the data analyses. The remaining 1184 participants (*M age* = 20.06 years, *SD* = 2.52) consisted of 783 males and 400 females with a mean of 8.51 years (*SD* = 4.09) competitive playing experience in their primary sport. Five participants did not specify their age and one student athlete did not specify their gender. Participants were involved in both team sports (*n* = 842), such as football (*n* = 427), rugby (*n* = 129), netball (*n* = 90), and cricket (*n* = 59), and individual sports (*n* = 339), such as athletics (*n* = 87), swimming (*n* = 37), tennis (*n* = 35) and badminton (*n* = 22). Three student athletes did not specify their primary sport participation. The highest level of primary sport participation ranged from club (*n* = 415) through county (*n* = 320), regional (*n* = 128), national (*n* = 98), international (*n* = 73) and recreational (*n* = 42) level. Thirty-three participants did not specify their highest level of primary sport participation.

Measures

***Mental Toughness Questionnaire*.** See Study 1 for details.

Social Desirability Scale (SDS; Reynolds, 1982). Participants completed the 12-item version of the Marlowe-Crowne Social Desirability scale. Participants were required to rate whether 12 statements regarding personal attributes and traits were true or false to them personally. Items included "There have times when I was quite jealous of the good fortune of others" and "I have never deliberately said something that hurt someone's feelings." Responses were summed to give a total social desirability score with higher scores indicating more socially desirable behaviours. SDS scores can range from 0 (low) to 12 (high).

Procedure

The procedures adopted in Study 2 were similar to those used in Study 1 with the inclusion of the completion of the SDS and the use of a cover story prior to completing the MTQ48. Specifically, participants were provided with a cover story and informed that the study was investigating the psychological characteristics and thought processes of competitive student athletes which required them to complete a questionnaire. They were told that the questionnaire would assess their general psychological attributes in sport. This cover story was implemented in an effort to minimise potential social desirability effects in responding to the MTQ48. Once the questionnaire was completed, participants were thanked for their participation and received a verbal debrief explaining the true nature of the study. Due to time constraints, only 551 participants were able to complete the SDS. Consent was gained from the authors of the MTQ48 prior to testing.

Statistical Analyses

Data were screened for missing responses using SPSS (Version 18) and revealed 22 participants with missing values. Listwise deletion resulted in 1,184 participants being included for analyses. CFA was conducted to assess the factorial validity of the hypothesised related four- and six-factor models of the MTQ48. In light of the data being nonnormally distributed, three different estimation methods were utilised in AMOS statistics (Version 18.0); namely the Maximum Likelihood (ML), Generalised Least Squares (GLS) and Asymptotic Distribution-free (ADF) method. This enabled an initial comparison of the fit indices and parameter estimates of the respective methods (see supplementary online Table 1)[1]. However, the more stringent estimation methods (i.e., GLS, ADF) had difficulty in specifying a proper solution (Blunch, 2008). Despite the sample size exceeding the required threshold, researchers have suggested that unless the sample size is extremely large (1,000 to 5,000 cases: West, Fitch, & Curran, 1995; > 2,500: Tabachnick & Fidell, 2001), the ADF method can perform very poorly and yield severely distorted estimated values and standard errors (Curran, West, & Fitch, 1996; Hu, Bentler, & Kano, 1992; West et al., 1995). Consequently, the large sample size used in Study 2 may not have been sufficient to afford the accurate use of the ADF method. Furthermore, the findings from Study 2 support those of Olsson, Foss, Troye, and Howell (2000) in that the ML estimation method provided the best fitting models when compared to GLS and

[1] In order to provide a full examination of the MTQ48s factorial validity, alternative model structures were also examined; namely the second-order four factor model and the second-order six-factor model. Fit indices of these alternative models can be found in the online supplementary materials (Table 1). Findings revealed inadequate model fit for both alternative solutions.

ADF in conditions of nonnormality. The ML estimation method was therefore the primary estimation method used for further analyses.

Multiple fit indices were used to evaluate the overall fit of the hypothesised models of the MTQ48. These included the CMIN/DF (Wheaton, Muthén, Alwin, & Summers, 1977), CFI, Parsimonious Comparative Fit Index (PCFI), RMSEA and AIC. These fit indices included measures from four different classes (descriptive fit, absolute fit, absolute fit with penalty function, alternative fit). Similar to Study 1, liberal thresholds indicate adequate model fit whereas conservative thresholds indicate good model fit. The χ^2 statistic assesses the magnitude of discrepancy between the hypothesised covariance matrix (Σ) and the sample covariance matrix (*S*) and a significant test result indicates a poor fit. However, when the sample is large, the χ^2 value is a very conservative estimate of model fit (Byrne, 2013); consequently a χ^2/degrees of freedom ratio (χ^2/df) is also calculated. The criteria for adequate fit of the CMIN/DF are values below 5.00 with a non-significant ($p > 0.05$) test result (Wheaton et al., 1977) and values below 2.00 with a non-significant ($p > 0.05$) test result reflect a good fit (Byrne). PCFI values close to or above 0.60 reflect good fit (Blunch, 2008). See Study 1 for CFI, RMSEA and AIC guidelines.

In order to further examine the factorial validity of the MTQ48, model re-specification protocols were employed when there was poor model fit between the sample covariance matrix (S) and the estimated covariance matrix (∑). Poor model fit can be due to a number of characteristics including mis-specified correlations between factors, items having low parameter estimate loadings on their hypothesised factors, and when items inadvertently ask the same question (Byrne, 2013). Modification indices were assessed since they provide the only meaningful information sources regarding CFA model mis-specification (Byrne).

In the event that the hypothesised model structures were not supported by the data, analyses were conducted to generate model re-specifications. The decisions underlying these model re-specifications were directed by identifying high measurement error covariances which sat in isolation and away from all the other modification indices. Such covariances are indicative of mis-specified items (Byrne, 2013). In this study, covariances above 30 were deemed to reflect high measurement error covariances. A progressive item removal protocol was then administered in an effort to respecify the model. Measurement error covariances between items which were specified to load on the same factor were collated. The content of each item incorporated in the error covariance was compared and assessed to determine whether they were inadvertently asking the same question. In order to address the potential overlap in item content, the item with the lowest relationship (parameter estimate) with its hypothesised factor was removed from the model (Byrne). This process was progressive in that once an item had been removed, full CFA analysis was administered so that model fit could be assessed in light of the respective phases of model re-specification. Items were removed in hierarchical order in relation to their error covariances with the highest modification indices being addressed first. Analysis at the individual item level is reported in Table 3. The methods used to assess parameter estimates were identical to those used in Study 1.

The impact of social desirability on responses to the MTQ48 was assessed by examining the relationships between the SDS and the factors of MTQ48. In line with previous research (e.g., Freeman, Coffee & Rees, 2011; Gucciardi Gordon, & Dimmock, 2009), weak correlations suggest that socially desirable responding had little impact on parameter estimates observed.

Results

Descriptive Statistics

The univariate skewness values of the MTQ48 items ranged from -1.222 to 0.558 and the univariate kurtosis values ranged from -.972 to 2.058 suggesting that in general, the items do not fall within the acceptable criteria (i.e., -0.179 to 0.179; Doane & Seward, 2011) and are in indicative of nonnormal data (see Table 3). Examination of Mardia's normalised coefficient of multivariate kurtosis indicated that the data departed from multivariate normality, where the coefficient was 344.16. Bentler (2005) suggests values below 5.00 to indicate normal data. Inter-factor correlations of the MTQ48 were weak to moderate suggesting that the factors represent related yet independent components of the 4/6Cs model of MT (see Table 4). Correlations between the non-nested factors (i.e., control and confidence) and their subordinate nested factors (i.e., control – emotion, control – life, confidence – abilities, confidence – life) were moderate as theoretically expected. For the related four-factor solution, the MTQ48 demonstrated adequate internal reliability in all subscales except control ($\alpha < .70$; Nunnally & Bernstein, 1994). For the related six-factor solution, the MTQ48 demonstrated adequate internal reliability in all subscales except control – emotion and control – life (see Table 4). The correlations between the respective factors of the MTQ48 and SDS were weak to moderate which suggests that socially desirable responding had little impact on the parameter estimates observed (see Table 4).

INSERT TABLE 3 HERE

INSERT TABLE 4 HERE

Confirmatory Factor Analyses

Goodness of fit

Consistent with results in Study 1, CFAs revealed that the related four- and related six-factor solutions of the MTQ48 were unsatisfactory when using both the conservative and liberal thresholds of the multiple indices of model fit and inspection of parameter estimates. Fit statistics revealed that the related four-factor solution was inadequate CMIN/DF = 6.129, $p < .001$, CFI = .614, PCFI = .585, RMSEA = .066, 90% confidence interval [CI] [.064, .067], AIC = 6786.535. Likewise the related six-factor solution was also inadequate CMIN/DF = 5.334, $p < .001$, CFI = .677, PCFI = .639, RMSEA = .061, 90% confidence interval [CI] [.059, .062], AIC = 5902.412. Despite the RMSEA indicating adequate fit, inspection of the overall model fit indicated Although the analyses indicated inadequate fit for the related four- or the related six-factor solutions, the results from the RMSEA and AIC provide some support to suggest that the related six-factor solution is somewhat superior.

Parameter estimates

Inspection of the parameter estimates revealed a large degree of inconsistency between the hypothesised structures, according to the related four-factor and related six-factor solutions and the current data (see Table 3). Parameter estimates did not support the hypothesised solutions with only six items considered good (12.50%), 19 items fair (39.58%) and 23 (47.92%) items considered poor on the related four-factor solution and only two items considered very good (4.12%), five items good (10.42%), 17 items fair (35.42%) and 24 items poor (50%) on the related six-factor solution. Collectively, CFA model fit indices and parameter estimates did not support the related four- or related six-factor solutions of the MTQ48 with the very large student athlete sample used.

Model-specification

Due to the poor fit of the respective hypothesised models of MT and the need to fully examine the 4/6Cs model of MT, modification indices were analysed to guide model re-specification in an effort to improve model fit. Independent model re-specification protocols and their respective fit indices can be found within the online supplementary materials (Table 2-5). Fit statistics revealed that the best fitting revised four-factor solution CMIN/DF = 5.480, $p < .001$, CFI = .788, PCFI = .723, RMSEA = .0602, 90% confidence interval [CI] [.059, .064], AIC = 2318.705, and the best fitting revised six-factor solution CMIN/DF = 4.723, $p < .001$, CFI = .798, PCFI = .731, RMSEA = .056, 90% confidence interval [CI] [.054, .058], AIC = 2744.062 were inadequate. Findings revealed that although the revised four-factor model showed the highest level of change in CFI from its respective hypothesised model specification, the revised six-factor model provided the best fitting model. However, the improvements in model fit failed to provide an adequately fitting model in accordance with the conservative and liberal thresholds.

Discussion

The MTQ48 has been the focus of debate for many researchers interested in examining MT in recent years due to concerns regarding its psychometric properties. Despite recent research examining the MTQ48s factorial validity (Gucciardi et al. 2012; Perry et al. 2013), its factor structure is still being questioned when used with athlete samples. This is problematic given that data gleaned from psychometric instruments serve to validate their underpinning conceptualisation (Marsh, 2002). Therefore, the purpose of this current research was to re-examine the factorial validity of the MTQ48 using two large and independent student athlete samples.

The findings of the respective CFAs in this current paper provided little support for the hypothesised models of the MTQ48 in that fit indices of the related four- and six-factor models revealed inadequate model fit when using both conservative and liberal guidelines. The findings of this current research are therefore consistent with those obtained by Gucciardi et al. (2012) and Perry et al. (2013) whereby little support was provided for the hypothesised models of the MTQ48 when using student athlete samples. The collective findings therefore appear to suggest that in its current form, the MTQ48 may not be a valid measure of the 4/6Cs model of MT which it intends to capture.

Using Comrey and Lee's (1992) guidelines, inspection of parameter estimates provides further evidence to question the factorial validity of the MTQ48 in that associations between the items and their respective factors were relatively weak (see Table 1 and 3). Results indicated that there were very few items which could be considered to have very good to excellent relationships with their hypothesised factors. In Study 1, only two items from the confidence-interpersonal factor (MTQ38, MTQ43) for the related six-factor model reached this threshold. In Study 2, only one item from the challenge factor (MTQ48) and one item from the commitment factor (MTQ29) could be considered to have very good to excellent relationships with their hypothesised factors. These findings are comparable to those observed by Gucciardi et al. (2012) in that only 22 out of 48 items in the related four-factor model could be considered as fair or above and only 5 out of 48 items could be considered very good or above, respectively (Comrey & Lee). Further inspection of parameter estimates in this current paper identified five items which had particularly low relationships with their hypothesised factors across both samples examined. These items included three items from the control - emotion factor (MTQ26R, MTQ34, MTQ37R), one item from the control - life factor (MTQ9R), and one item from the confidence - abilities factor (MTQ24R). According to Comrey and Lee, these values could be considered too low to be interpreted, which may provide evidence to suggest that these items are major contributors to the lack of support found for the hypothesised models of the MTQ48.

Similarly, Gucciardi et al. (2012) identified the aforementioned items as having extremely low relationships with their hypothesised factors and identified a further 12 items which could be considered too low for interpretation. Specifically, Gucciardi et al. identified poor items relating to all four factors examined; challenge (MTQ6R, MTQ14R), commitment (MTQ35R), control (MTQ15R, MTQ33R, MTQ21R, MTQ27R) and confidence (MTQ10R, MTQ28R, MTQ32R, MTQ36R, MTQ46R). The findings of this present paper coupled with those found by Gucciardi et al. therefore indicate that there are a number of items relating to all four MT factors which are inadequate representations of their hypothesised factors, with the most pronounced inadequate items relating to the control and confidence factors.

In addition, item MTQ34 demonstrated a negative association with its control - emotion factor. This finding is similar to that observed in previous research (e.g., Crust & Swann, 2011b;

Gucciardi et al., 2012; Perry et al., 2013) in that item MTQ34 was found to be unrelated to control - emotion. One possible reason for this unexpected negative association might be due to the content of item MTQ34. This item “I generally hide my emotion from others” reflects the facet of the control - emotion factor which argues that mentally tough performers are “less likely to reveal their emotional state to other people” (Clough et al., 2007). Although this facet of control - emotion may appear intuitively compelling, no empirical rationale for its inclusion was provided. Nicholls and Polman’s (2007) systematic review of coping in sport would appear to partially conflict with this supposition in that findings generally appear to highlight the importance of being able to effectively control one’s emotions to maintain/enhance performance as opposed to supporting the use of covert coping strategies to suppress one’s emotions. Based upon Nicholls and Polman’s review, an athlete’s MT levels may be more accurately determined by their effectiveness in controlling emotions as opposed to whether one controls their emotions covertly or overtly. The ambiguity surrounding item MTQ34 and its apparent negative association with its hypothesised control - emotion factor suggests that this item warrants major revision. Indeed, Perry et al. stated that the removal of item MTQ34 would significantly improve the reliability of the control factor across all samples. However, Perry et al. did not perform model re-specifications to potentially remove item MTQ34 and thereby enhance the MTQ48s factorial validity.

In an effort to further our understanding of the MTQ48s factorial validity, Study 2 of this present research used model re-specification protocols to determine whether item removal could enhance the instruments psychometric properties. Examination of the best fitting revised four- and six-factor models (inclusive of item 34 being removed) revealed that model fit did improve across both models, yet did not collectively reach the required thresholds (both conservative and liberal) to adequately support the fit of the respective revised models of the MTQ48. The findings appear to suggest that despite rigorous model re-specification protocols, the best fitting revised models of the MTQ48 remain unacceptable. Despite a growing body of evidence supporting the MTQ48s convergent validity, the findings of this present research and Gucciardi et al. (2012) emphasise the importance of assessing an instrument’s factorial validity prior to assessing other forms of validity (Gignac, 2009; Marsh et al., 2010). Indeed, the utility of an instrument is underpinned by the degree at which its items accurately measure the constructs it intends to capture (e.g., MacKenzie et al. 2011; McGrath, 2005).

A possible explanation for the poor fit of the models examined could have been due to the extent to which the items represent their hypothesised factor. As reflected by the poor parameter estimates obtained, it appears that a large proportion of items are inadequately representing their hypothesised factor definitions forwarded by Clough et al. (2007). Inspection of item content provides credence to this argument. For instance, item MTQ9R – "I usually find myself just going through the motions" does not appear to accurately represent the hypothesised definition of the control - life factor which states that individuals high in control – life have the ability to control one's life, feel that their plans will not be thwarted and that they can make a difference. Similarly, item MTQ37R – "When I am feeling tired I find it difficult to get going" does not appear to accurately reflect the control – emotion factor which encapsulates an individual's ability to control their emotions, keep their anxieties in check and be less likely to reveal their emotional state to other people. One may therefore argue that there appears to be some concerning issues at item-level with the instrument. Consequently, it appears that in the development of the MTQ48, critical steps were overlooked in the scale development process whereby the ultimate objective of item generation is to develop an item pool that encapsulates the core facets of the focal construct (MacKenzie et al., 2011).

A further explanation for the poor fit indices obtained could have been due to the structure and clarity of the items. The scale development literature (e.g., Clark & Watson, 1995; MacKenzie *et* al., 2011) provides guidelines to inform item wording and structure that emphasise the importance of ensuring clarity, specificity, and brevity with each item. Specifically, Clark and Watson suggest that the exact phrasing of items can exert a profound influence on the construct being measured. However, inspection of the MTQ48 indicates that some items do not adequately fulfil these criteria. For instance, item MTQ11R – "'I just don't know where to begin' is a feeling I usually have when presented with several things to do at once" does not appear to share the same structure (i.e., the use of a quotation) as the remaining MTQ48 items. Similarly, item MTQ16 - "I generally look on the bright side of life" and item MTQ33R – "Things just usually happen to me" appear to be quite vague and thus may have confused the respondent in terms of what the question is really asking. In addition, item MTQ26R – "When I am upset or annoyed I usually let others know" contains double barrelled wording which conflicts with the recommendations of the scale development literature (e.g., Clark & Watson). One may therefore argue that the lack of clarity and specificity surrounding the items of the MTQ48 may have made it difficult for the respondents to accurately answer the question, which in turn, may have contributed towards the poor factorial validity of the models examined.

Another explanation for the poor fit of the models examined could have been due to the non-normal distribution of data. In order to comprehensively assess the hypothesised models of the MTQ48 in light of the data being nonnormal, in Study 1 the MLR (S-B $\chi 2$; Satorra & Bentler, 1994) estimation method was used, and in Study 2 the ML, GLS, and ADF estimation methods were utilised. However, given the range of estimation methods used in this paper, it is unlikely that non-normality was a large contributor towards the poor factorial validity found for the MTQ48. Researchers are encouraged to explore a range of estimation methods to further our understanding of their use and applicability when examining the psychometric properties instruments within the sport and exercise psychology literature. Moreover, researchers are encouraged to combine and publish independent research projects which are closely related as this may provide a fruitful and progressive means to further our understanding of a given concept.

Limitations

The major strength of this paper is undoubtedly the use of rigorous statistical analyses that accounted for measurement error but most importantly the inclusion of two independent samples of student athletes of significant size which attempted to address limitations raised in Gucciardi et al.'s (2012) study. Indeed, the samples used in Study 1 and 2 could be considered good (> 300) and excellent (> 1,000), respectively (Comrey & Lee, 1992). Although the sample was limited to student athletes, the range of athletic abilities captured was indeed diverse and not limited to collegiate playing levels given that the distribution of athletes' highest level of competitive experience was relatively balanced, ranging from international to recreational (see respective Participants section for demographic information). The use of a cover story and social desirability scale in Study 2 could be considered a strength of this present manuscript. Social desirability has been argued to be one of the most prominent sources of systematic error which may compromise response validity (Podsakoff et al., 2003). Recent research (e.g., Gucciardi et al., 2009; Hagger & Chatzisarantis, 2009) has called for the use of social desirability scales to assess the propensity for socially desirable responding. The weak correlations observed in this paper suggest that minimal socially desirable responding of the MTQ48 was present in Study 2. Despite a number of recent studies examining social desirability in the development of questionnaires in sport and exercise psychology (e.g., Freeman et al., 2011; Gucciardi et al.), this area has generally been overlooked and requires more attention in future research. Future scale development research should consider the impact of social desirability when constructing and validating a scale and should not overlook the use of a cover story to minimise the adverse effects of social desirability.

Conclusion

The findings of this present research revealed little support for the factor structure of both the hypothesised models and the revised models of the MTQ48. Our results provide additional support to current concerns regarding the MTQ48s factorial validity (Connaughton & Hanton, 2009; Gucciardi et al., 2011, Gucciardi et al., 2012) and indicate that in its present form, the MTQ48 may not be a valid measure of the 4/6Cs model of MT when using student athletes. The collective findings of this present paper and Gucciardi et al. are particularly concerning given the lack of support found when using both paper (this current article) and online formats (Gucciardi et al.) of the MTQ48. Although a number of explanations have been offered to explain the poor factorial validity of the MTQ48, the data suggests that there are major concerns regarding the adequacy (face validity) of the MTQ48 items to represent the factor definitions of the 4/6Cs model of MT. Given that systematic model re-specification protocols were not sufficient to combat the inadequacies of the originally hypothesised models of the MTQ48 and that psychometric development of a measure is an ongoing process (Marsh, 2002), further refinement of the MTQ48 items appears necessary. Until such time, researchers and practitioners using the MTQ48 as a measure of MT in student athlete samples should proceed with an element of caution.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52*, 317-332.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, *13*, 186–203.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, *107*, 238.

Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological Bulletin*, *112*, 400-404.

Bentler, P. M. (1995). *EQS structural equations program manual*. Encino, CA: Multivariate software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*, 588.

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. *Multivariate analysis VI*, 9-42.

Bentler, P. M., & Wu, E. J. (2005). *EQS 6.1 for Windows*. *Structural equations program manual*. Encino, CA: Multivariate Software.

Birch, P. D. J., Greenlees. I. A., Lowry, R. G., & Coffee, P. (2012). *The Mental Toughness Questionnaire-48: A factorial analysis.* Paper presented at the British Psychological Society Annual Conference, UK. April, 2012.

Blunch, N. J. (2008). *Introduction to Structural Equation Modelling: Using SPSS and AMOS.* Sage.

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.

Byrne, B. M. (2013). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. London: Routledge.

Chou, C. P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology, 44,* 347-357.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, ***7***, 309-319.

Clough, P., Earle, K., & Sewell, D. (2002). Mental toughness: The concept and its measurement. In I. Cockerill (Eds.), *Solutions in Sport Psychology* (pp. 32-45). London: Thomson.

Clough, P., Marchant, D., & Earle, K. (2007). The MTQ48 Technical Manual. Retrieved from http://www.aqr.co.uk/sites/default/files/MTQ48%20Technical%20Manual%20Jan%202007.pdf.

Comrey, A. L., & Lee, H. B. (1992). *A first course in Factor Analysis* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Connaughton, D., & Hanton, S. (2009). Mental toughness in sport: Conceptual and practical issues. In S. D. Mellalieu & S. Hanton (Eds.), *Advances in applied sport psychology: A review* (pp. 317-346). London: Routledge.

Crampton, S. G. D. (2010). *Testing the toughness of the Mental Toughness Test: Exploring the factorial validity of the MTQ48.* Paper presented at the British Psychological Society Annual Conference, UK. Dec, 2010.

Crust, L., & Azadi, K. (2009). Leadership preferences of mentally tough athletes. *Personality and Individual Differences, 47*, 326-330.

Crust, L., & Clough, P. J. (2005). Relationship between mental toughness and physical endurance. *Perceptual and Motor Skills, 100*, 192-194.

Crust, L., & Keegan, R. (2010). Mental toughness and attitudes to risk-taking. *Personality and Individual Differences, 49*, 164-168.

Crust, L., & Swann, C. (2011a). The relationship between mental toughness and dispositional Flow. *European Journal of Sport Science, 13*, 215-220.

Crust, L., & Swann, C. (2011b). Comparing two measures of mental toughness. *Personality and Individual Differences, 50*, 217-221.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*, 16–29.

Doane, D. P., & Seward, L. E. (2011). Measuring skewness: a forgotten statistic. *Journal of Statistics Education, 19*, 1-18.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, *47*, 309-326.

Earle, K. (2006). *The construct of mental toughness: A psychometric and experimental analysis.* Unpublished PhD thesis, University of Hull.

Freeman, P., Coffee, P**.,** & Rees, T. (2011). The PASS-Q: The perceived available support in sport questionnaire. *Journal of Sport & Exercise Psychology*, *33*, 54-74.

Gignac, G. E. (2009). Psychometrics and the measurement of emotional intelligence. In C. Stough, D. H. Saklofske & J. D. A. Parker (Eds.), *Assessing emotional intelligence: Theory, research, and applications* (pp. 9-42), New York: Springer.

Gucciardi, D. F., & Gordon, S. (2009). Development and preliminary validation of the Cricket Mental Toughness Inventory (CMTI). *Journal of Sports Sciences, 27*, 1293-1310.

Gucciardi, D., Gordon, S., & Dimmock, J. (2009). Development and preliminary validation of a mental toughness inventory for Australian football. *Psychology of Sport and Exercise*, *10*, 201-209.

Gucciardi, D. F., Hanton, S., Mallett, C. J. (2012). Progressing measurement in mental toughness: A case example of the Mental Toughness Questionnaire 48. *Sport, Exercise, and Performance Psychology*, *1*, 194-214.

Gucciardi, D. F., Mallett, C. J., Hanrahan, S. J., & Gordon, S. (2011). Measuring mental toughness in sport. In S. Gordon & D. F. Gucciardi (Eds.), *Mental toughness in sport: Developments in theory and research* (pp. 108-132). London: Routledge.

Hagger, M. S., & Chatzisarantis, N. L. D. (2009). Assumptions in research in sport and exercise psychology. *Psychology of Sport & Exercise*, *10*, 511-519.

Horsburgh, V. A., Schermer, J. A., Veselka, L., & Vernon, P. A. (2009). A behavioural genetic study of mental toughness and personality. *Personality and Individual Differences, 46,* 100-105.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Eds.), *Structural equation modelling: Concepts, issues, and applications* (pp. 158-176). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*, 351-362.

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

Kaiseler, M. K., Polman, R. C. J., & Nicholls, A. R. (2009). Mental toughness, stress, stress appraisal, coping, and coping effectiveness. *Personality and Individual Differences, 47*(7), 728-733.

Kline, R. B. (1998). *Principles and practice of structural equation modelling.* New York: Guilford Press.

Kobasa, S. C. (1979). Stressful life events, personality, and health: An inquiry into hardiness. *Personality and Social Psychology, 37*, 1-11.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N, P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly*, *35*, 293-334.

Maddi, S. R. (2002). The story of hardiness: Twenty years of theorizing, research, and practice. *Consulting Psychology Journal*, *54*, 175-185.

Maddi, S. R., & Kobasa, S. C. (1984). *The hardy executive: Health under stress.* Homewood, IL: Dow-Jones Irwin.

Marsh, H. W. (2002). A multidimensional physical self-concept: A construct validity approach to theory, measurement, and research. *Psychology: The Journal of the Hellenic Psychology Society*, *9*, 459-93.

Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of sport psychology* (3rd ed., pp. 774-798). New York: Wiley.

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 overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320-341.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. *Psychological bulletin*, *97*, 562.

Marsh, H. W., Martin, A. J., & Jackson, S. A. (2010). Introducing a short version of the Physical Self-Description Questionnaire: New strategies, short form evaluation evaluative criteria, and applications of factor analyses. *Journal of Sport and Exercise Psychology*, *32*, 438-482.

McGrath, R. E. (2005). Conceptual complexity and construct validity. *Journal of Personality Assessment*, *85*, 112–24.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171-189.

Nicholls, A. R., Levy, A. R., Polman, R. C. J., & Crust, L. (2011). Mental toughness, coping self-efficacy, and coping effectiveness among athletes. *International Journal of Sport Psychology, 42*, 513-524.

Nicholls, A. R., & Polman, R. C. J. (2007). Coping in sport: A systematic review. *Journal of Sports Sciences*, *25*, 11-31.

Nicholls, A. R., Polman, R. C. J., Levy, A. R., & Backhouse, S. H. (2008). Mental toughness, optimism, pessimism, and coping among athletes. *Personality and Individual Differences*, *44*, 1182-1192.

Nunnally, J. C., & Bernstein, I. H. (1994). The assessment of reliability. *Psychometric theory, 3*, 248-292.

Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural Equation Modeling*, *7,* 557-595.

Perry, J. L., Clough, P. J., Crust, L., Earle, K., & Nicholls, A. R. (2013). Factorial validity of the mental toughness questionnaire-48. *Personality and Individual Differences, 54,* 587-592.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioural research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.

Raykov, T., & Marcoulides, G. A. (2000). *A first course in structural equation modeling*, Mahwah, NJ: Lawrence Erlbaum Associates.

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology*, *38*, 119-125.

Russell, D. W. (2002). In search of underlying dimensions: the use (and abuse) of factor analysis in personality and social psychology bulletin. *Personality and Social Psychology Bulletin*, *28*, 1629-1646.

Satorra, A. B., & Bentler, P. P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. *Latent variables analysis: Applications for developmental research*, 399-419.

Schutz, R. W., & Gessaaroli, M. E. (1993). Use, misuse, and disuse of psychometrics in sport psychology research. In R. N. Singer, M. Murphy & L. K. Tennant (Eds.), *Handbook of research on sport psychology* (pp. 901-917). New York: Macmillan.

Sheard, M. (2012). *Mental toughness: The mindset behind sporting achievement*. Routledge.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate behavioral research, 25*, 173-180.

Tabachnick, B. G, & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). MA: Allyn and Bacon.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.

West, S. G., Fitch, F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In Hoyle R. H (Eds.), *Structural equation modelling: Concepts, issues, and applications* (pp. 56-75). Thousand Oaks, CA: Sage.

Wheaton, B., Muthen, B., Alwin, D. E., & Summers, G. E. (1977). Assessing reliability and stability in panel models. In D. R. Heise (Eds.), *Sociological methodology* (pp. 85-136). San Francisco: Jossey-Bass.

Table 1. Item means, standard deviations, skewness and kurtosis values, standardised parameter estimates and squared residuals for the four- and six-factor models of the MTQ48 in Study 1.

					4 factor		6 factor	
	M	*SD*	Skewness	Kurtosis	PE	R^2	PE	R^2
CHALL								
MTQ4	4.10	.77	-.599	.359	.516	.266	.517	.267
MTQ6	3.13	.98	-.217	-.592	.248	.062	.249	.062
MTQ14	3.38	.98	-.334	-.248	.365	.133	.363	.132
MTQ23	3.61	.69	-.799	.729	.499	.249	.496	.246
MTQ30	3.68	.66	-.794	1.182	.542	.293	.545	.297
MTQ40	3.32	.92	-.430	-.429	.221	.049	.219	.048
MTQ44	4.16	.65	-.670	1.613	.611	.373	.616	.380
MTQ48	3.90	.57	-1.018	3.812	.650	.423	.644	.415
MTQ1	4.06	.60	-.721	2.790	.339	.115	.354	.125
COM								
MTQ7	3.99	.85	-1.032	1.474	.579	.335	.571	.326
MTQ11	3.10	1.05	-.175	-.773	.415	.173	.416	.173
MTQ19	4.06	.66	-.840	2.424	.383	.147	.381	.145
MTQ22	3.07	1.03	-.112	-.839	.394	.155	.405	.164
MTQ25	4.25	.76	-.981	1.114	.533	.285	.540	.291
MTQ29	4.11	.76	-1.000	1.948	.660	.435	.650	.423
MTQ35	2.72	.97	.409	-.662	.457	.209	.447	.200
MTQ39	3.35	.91	-.331	-.339	.512	.262	.515	.265
MTQ42	3.54	.84	-.582	.040	.451	.203	.462	.214
MTQ47	3.52	.92	-.510	-.132	.439	.193	.435	.189
CE								
MTQ21	3.52	1.12	-.484	-.598	.318	.101	.397	.158
MTQ26	2.95	1.12	.028	-.954	.049	.002	.157	.025
MTQ27	2.70	1.13	.263	-.916	.510	.260	.646	.417
MTQ31	3.53	.82	-.663	.127	.490	.241	.601	.361
MTQ34	3.17	1.06	-.058	-.799	-.147	.022	-.022	.000
MTQ37	2.62	.98	.442	-.581	.340	.115	.340	.116
MTQ45	3.39	.97	-.585	-.187	.529	.280	.611	.373
CL								
MTQ2	3.92	.66	-.643	1.401	.554	.307	.568	.323
MTQ5	3.55	.79	-.405	.432	.407	.166	.416	.173
MTQ9	3.12	.89	-.012	-.594	.262	.069	.304	.093
MTQ12	3.85	.78	-.850	1.402	.518	.268	.544	.296
MTQ15	3.29	.91	-.397	-.131	.374	.140	.413	.170
MTQ33	3.25	.86	-.187	-.122	.306	.093	.331	.109
MTQ41	3.65	.80	-.612	.325	.478	.229	.523	.274

Table 1 *Cont.*

	M	*SD*	Skewness ratio	Kurtosis ratio	4 factor PE	4 factor R^2	6 factor PE	6 factor R^2
CA								
MTQ3	4.04	.69	-.984	3.007	.516	.266	.512	.262
MTQ8	3.88	.81	-.858	1.055	.597	.357	.593	.352
MTQ10	2.72	.94	.284	-.640	.424	.180	.440	.194
MTQ13	3.80	.83	-.683	.530	.439	.193	.464	.216
MTQ16	3.93	.75	-.684	.771	.511	.261	.521	.272
MTQ18	3.46	1.00	-.384	-.459	.543	.295	.558	.311
MTQ24	2.39	.96	.527	-.251	.140	.019	.142	.020
MTQ32	3.40	.87	-.250	-.366	.466	.217	.497	.247
MTQ36	2.97	1.17	.045	-1.033	.559	.312	.586	.343
CI								
MTQ17	3.68	.96	-.517	-.182	.420	.176	.625	.390
MTQ20	3.88	.75	-.718	1.149	.382	.146	.487	.237
MTQ28	3.53	1.05	-.328	-.650	.463	.215	.494	.244
MTQ38	3.50	.93	-.603	.126	.423	.179	.617	.381
MTQ43	3.59	.98	-.527	-.236	.350	.123	.622	.387
MTQ46	3.58	1.05	-.452	-.526	.416	.173	.572	.328

Note: *M* = Mean; *SD* = Standard deviation; PE = Standardised parameter estimates; R^2 = Squared residuals; CHALL = Challenge; COM = Commitment; CE = Control-emotion; CL = Control-life; CA = Confidence-ability, CI = Confidence-interpersonal.

Table 2. Factor means, standard deviations, correlations and internal consistencies in Study 1.

	M	*SD*	1	2	3	4	5	6	7	8
1. Challenge	29.28	3.39	(.644)	.488**	.544**	.421**	.459**	.562**	.524**	.403**
2. Commitment	39.76	5.11		(.756)	.501**	.310**	.516**	.438**	.474**	.236**
3. Control	46.50	5.66			(.668)	.843**	.762**	.642**	.682**	.361**
4. Control-emotion	21.87	3.83				(.573)	.293**	.430**	.518**	.169**
5. Control-life	24.63	3.18					(.630)	.623**	.588**	.438**
6. Confidence	52.36	6.96						(.787)	.866**	.799**
7. Confidence- abilities	30.60	4.55							(.727)	.390**
8. Confidence-interpersonal	21.75	3.78								(.737)

Note: *M* = Mean; *SD* = Standard Deviations; **Significance is at $p < 0.01$; Cronbach's alphas are displayed in parentheses.

Table 3. Item means, standard deviations, skewness and kurtosis values, standardised parameter estimates and squared residuals for the four- and six-factor models of the MTQ48 in Study 2.

					4 factor		6 factor	
	M	*SD*	Skewness	Kurtosis	SL	R^2	SL	R^2
CHALL								
MTQ4	4.10	0.79	-.576`	-.036	.578	.334	.577	.333
MTQ6	3.26	0.99	-.372	-.409	.361	.130	.360	.130
MTQ14	3.56	1.04	-.468	-.396	.389	.151	.381	.145
MTQ23	3.66	0.68	-.580	.397	.591	.349	.592	.350
MTQ30	3.63	0.73	-.670	.807	.496	.246	.500	.250
MTQ40	3.27	1.02	-.272	-.454	.363	.132	.361	.131
MTQ44	4.24	0.71	-.769	.811	.615	.378	.619	.383
MTQ48	3.95	0.65	-.577	1.324	.610	.372	.609	.370
COM								
MTQ1	4.04	0.72	-.639	.951	.440	.194	.445	.198
MTQ7	4.09	0.84	-1.058	1.381	.504	.254	.499	.249
MTQ11	3.05	1.15	-.120	-.944	.481	.231	.477	.227
MTQ19	4.11	0.68	-.659	1.244	.430	.184	.439	.192
MTQ22	3.11	1.06	-.252	-.729	.462	.214	.461	.213
MTQ25	4.35	0.74	-1.094	1.289	.517	.268	.519	.270
MTQ29	4.19	0.82	-1.222	2.058	.622	.386	.618	.382
MTQ35	2.58	1.04	.401	-.650	.434	.188	.427	.182
MTQ39	3.23	0.96	-.211	-.574	.510	.260	.508	.258
MTQ42	3.68	0.91	-.584	-.049	.526	.277	.531	.282
MTQ47	3.59	0.95	-.546	-.250	.454	.206	.458	.210
CE								
MTQ21	3.62	1.11	-.592	-.404	.346	.120	.406	.165
MTQ26	2.98	1.19	.040	-.969	.006	.000	.077	.006
MTQ27	2.72	1.18	.270	-.907	.436	.190	.549	.301
MTQ31	3.48	0.88	-.376	-.264	.497	.247	.616	.380
MTQ34	3.17	1.14	-.134	-.876	-.136	.019	-.087	.008
MTQ37	2.57	0.99	.483	-.481	.315	.099	.298	.089
MTQ45	3.40	1.03	-.513	-.324	.518	.268	.627	.393
CL								
MTQ2	3.92	0.66	-.729	1.593	.509	.260	.514	.265
MTQ5	3.65	0.77	-.158	-.280	.358	.128	.347	.121
MTQ9	3.17	0.96	-.109	-.446	.240	.057	.247	.061
MTQ12	3.84	0.79	-.714	.785	.477	.227	.498	.248
MTQ15	3.41	0.95	-.455	-.221	.424	.180	.447	.199
MTQ33	3.33	0.96	-.181	-.333	.364	.133	.400	.160
MTQ41	3.81	0.83	-.579	.294	.518	.268	.556	.309

Table 3 *Cont.*

					4 factor		6 factor	
	M	*SD*	Skewness	Kurtosis	PE	R^2	PE	R^2
CA								
MTQ3	4.09	0.73	-.689	1.074	.514	.264	.540	.292
MTQ8	3.87	0.80	-.747	.780	.537	.289	.532	.283
MTQ10	2.78	1.01	.272	-.555	.383	.146	.408	.167
MTQ13	3.74	0.91	-.506	-.141	.429	.184	.483	.233
MTQ16	4.01	0.80	-.877	1.225	.437	.191	.501	.251
MTQ18	3.61	1.03	-.474	-.444	.556	.309	.577	.333
MTQ24	2.34	0.98	.558	-.196	.232	.054	.278	.077
MTQ32	3.55	0.95	-.453	-.213	.518	.269	.559	.313
MTQ36	3.13	1.15	-.113	-.972	.464	.215	.480	.231
CI								
MTQ17	3.73	0.94	-.485	-.404	.323	.104	.580	.336
MTQ20	3.94	0.80	-.611	.418	.381	.145	.578	.334
MTQ28	3.79	1.06	-.703	-.135	.465	.217	.460	.212
MTQ38	3.55	0.95	-.527	-.118	.449	.202	.640	.409
MTQ43	3.55	1.04	-.568	-.245	.366	.134	.643	.413
MTQ46	3.79	1.02	-.643	-.206	.435	.189	.572	.327

Note: *M* = Mean; *SD* = Standard deviation; PE = Standardised parameter estimates; R^2 = Squared residuals; CHALL = Challenge; COM = Commitment; CE = Control-emotion; CL = Control-life; CA = Confidence-ability, CI = Confidence-interpersonal.

Table 4. Factor means, standard deviations, correlations between factors of the MTQ48 and SDS, inter-factor correlations and internal consistencies in Study 2.

	M	*SD*	SDS	1	2	3	4	5	6	7	8
1. Challenge	29.67	3.84	0.20**	(.704)	.518**	.568**	.431**	.500**	.591**	.543**	.404**
2. Commitment	40.03	5.49	0.33**		(.769)	.543**	.369**	.531**	.486**	.456**	.322**
3. Control	47.06	5.81	0.22**			(.655)	.847**	.777**	.628**	.643**	.348**
4. Control-emotion	21.94	3.86	0.19**				(.530)	.324**	.414**	.480**	.160**
5. Control-life	25.12	3.26	0.17**					(.610)	.628**	.576**	.431**
6. Confidence	53.48	7.00	0.04						(.776)	.852**	.765**
7. Confidence-abilities	31.13	4.75	0.18**							(.732)	.316**
8. Confidence-interpersonal	22.35	3.86	-0.13**								(.745)

Note: *M* = Mean; *SD* = Standard deviations; SDS = Social Desirability Scale; **Significance is at $p < 0.01$; Cronbach's alphas are displayed in parentheses.

Supplementary online materials

Table 1. Summary of fit indices across hypothesised model specification and estimation method of the MTQ48.

	Fit indices					
CFA method	CMIN/DF	*P*	CFI	PCFI	RMSEA (90% CI)	AIC
Criterion values	(<2.00)	(> 0.05)	(>0.95)	(>0.6)	(<0.05)	(lower = better)
Related four factor						
ML	6.129	0.000	0.614	0.585*	0.066 (0.064, 0.067)	6786.535
GLS	3.058	0.000	0.207	0.197	0.042* (0.040, 0.043)	3488.476
ADF	18.767	0.000	0.649	0.618	0.123 (0.121, 0.124)	20359.922
Related six factor						
ML	5.334	0.000	0.677	0.639*	0.061 (0.059, 0.062)	5902.412
GLS	2.949	0.000	0.255	0.241	0.041* (0.039, 0.042)	3363.087
ADF	23.564	0.000	0.564	0.532	0.137 (0.136, 0.139)	24997.935
Second-order four factor						
ML	6.165	0.000	0.611	0.583	0.066 (0.065, 0.068)	6929.326
GLS	3.071	0.000	0.201	0.191	0.042* (0.040, 0.043)	3504.073
ADF	18.696	0.000	0.650	0.620*	0.122 (0.121, 0.124)	20317.222
Second-order six factor						
ML	5.433	0.000	0.667	0.635*	0.061 (0.060, 0.063)	6134.95
GLS	2.976	0.000	0.239	0.227	0.041* (0.039, 0.043)	3400.223
ADF	60.957	0.000	0	0	0.225 (0.224, 0.227)	65671.854

* indicates good fit.

Table 2. Fit indices of the related four factor model re-specification protocols.

		Fit indices					
CFA run Criterion values	Items removed	CMIN/DF (< 2.00)	*P* (> 0.05)	CFI (> 0.95)	PCFI (> 0.6)	RMSEA (90% CI) (< 0.05)	AIC (lower = better)
1	Hypothesised model	6.129	0.000	0.614	0.585	0.066 (0.064, 0.067)	6786.535
2	34, 26R	6.114	0.000	0.635	0.603*	0.066 (0.064, 0.067)	6206.216
3	13	6.032	0.000	0.644	0.611*	0.065 (0.064, 0.067)	5856.435
4	17	5.908	0.000	0.657	0.623*	0.064 (0.063, 0.066)	5481.366
5	20	5.762	0.000	0.671	0.635*	0.063 (0.062, 0.065)	5104.643
6	43	5.617	0.000	0.686	0.647*	0.062 (0.061, 0.064)	4746.276
7	33R	5.551	0.000	0.695	0.655*	0.062 (0.060, 0.064)	4467.123
8	46R	5.590	0.000	0.700	0.658*	0.062 (0.060, 0.064)	4275.355
9	35R	5.275	0.000	0.720	0.676*	0.060 (0.058, 0.062)	3839.625
10	14R, 38, 6R, 40, 21R, 5, 37R, 24, 9R	5.480	0.000	0.788	0.723*	0.062 (0.059, 0.064)	2318.705

* indicates good fit.

Table 3. Fit indices of the related six factor model re-specification protocols.

		Fit indices					
CFA run Criterion values	Items removed	CMIN/DF (< 2.00)	*P* (> 0.05)	CFI (> 0.95)	PCFI (> 0.6)	RMSEA (90% CI) (< 0.05)	AIC (lower = better)
1	Hypothesised model	5.334	0.000	0.677	0.639*	0.061 (0.059, 0.062)	5902.412
2	34, 26R	5.238	0.000	0.700	0.659*	0.060 (0.058, 0.061)	5316.17
3	13	5.163	0.000	0.708	0.665*	0.059 (0.058, 0.061)	5011.692
4	33R	5.101	0.000	0.717	0.673*	0.059 (0.057, 0.061)	4730.473
5	35R	4.834	0.000	0.736	0.689*	0.057 (0.055, 0.059)	4286.854
6	4[a]	4.924	0.000	0.733	0.685*	0.058 (0.056, 0.059)	4157.16
7	36R	4.767	0.000	0.744	0.695*	0.056 (0.055, 0.058)	4030.514
8	22R[a]	4.815	0.000	0.747	0.696*	0.057 (0.055, 0.059)	3872.884
9	6R	4.656	0.000	0.757	0.705*	0.056 (0.054, 0.057)	3751.455
10	21R, 5, 14R, 40, 37R, 24, 9R	4.723	0.000	0.798	0.731*	0.056 (0.054, 0.058)	2744.062

* indicates good fit.

[a] denotes item being retained due to decrement in model fit.

Table 4. Fit indices of the second-order four factor model re-specification.

		Fit indices					
CFA run Criterion values	Items removed	CMIN/DF (< 2.00)	*P* (> 0.05)	CFI (> 0.95)	PCFI (> 0.6)	RMSEA (90% CI) (< 0.05)	AIC (lower = better)
1	Hypothesised model	6.165	0.000	0.611	0.583	0.066 (0.065, 0.068)	6929.326
2	34	6.070	0.000	0.625	0.596	0.065 (0.064, 0.067)	6542.103
3	26R	6.151	0.000	0.632	0.601*	0.066 (0.064, 0.068)	6342.637
4	13	6.069	0.000	0.641	0.609*	0.065 (0.064, 0.067)	5988.914
5	17	5.948	0.000	0.654	0.621*	0.065 (0.063, 0.066)	5613.691
6	20	5.812	0.000	0.667	0.632*	0.064 (0.062, 0.066)	5240.997
7	43	5.671	0.000	0.681	0.645*	0.063 (0.061, 0.065)	4882.063
8	33R	5.600	0.000	0.691	0.653*	0.062 (0.061, 0.064)	4594.223
9	14R, 6R, 40, 21R, 5, 37R, 9R, 24	5.540	0.000	0.760	0.707*	0.062 (0.060, 0.064)	2926.319

* indicates good fit.

Table 5. Fit indices of the second-order six factor model re-specification.

		Fit indices					
CFA run Criterion values	Items removed	CMIN/DF (< 2.00)	*P* (> 0.05)	CFI (> 0.95)	PCFI (> 0.6)	RMSEA (90% CI) (< 0.05)	AIC (lower = better)
1	Hypothesised model	5.433	0.000	0.667	0.635*	0.061 (0.060, 0.063)	6134.95
2	36<>26	5.179	0.000	0.686	0.653*	0.059 (0.058, 0.061)	5858.948
3	34	5.295	0.000	0.683	0.650*	0.060 (0.059, 0.062)	5737.558
4	26R	5.348	0.000	0.690	0.655*	0.061 (0.059, 0.062)	5545.051
5	13	5.280	0.000	0.697	0.661*	0.060 (0.059, 0.062)	5240.085
6	33R	5.205	0.000	0.707	0.670*	0.060 (0.058, 0.061)	4939.714
7	4[a]	5.304	0.000	0.704	0.666*	0.060 (0.059, 0.062)	4799.538
8	35R	4.954	0.000	0.725	0.686*	0.058 (0.056, 0.060)	4500.99
9	17[a]	5.026	0.000	0.724	0.683*	0.058 (0.057, 0.060)	4349.989
10	22R	5.021	0.000	0.727	0.686*	0.058 (0.057, 0.060)	4346.154
11	14R, 21R, 40, 6R, 5, 37R, 24, 9R	5.155	0.000	0.783	0.727*	0.059 (0.057, 0.061)	2901.665

* indicates good fit.

[a] denotes item being retained due to decrement in model fit.