*Title page with all author information

Running head: RESPONSE TO CLOUGH ET AL.

Progressing Measurement in Mental Toughness: A Response to Clough, Earle, Perry, and Crust

Daniel F. Gucciardi[1, 2]*, Sheldon Hanton[3], and Clifford J. Mallett[2]

[1]*School of Physiotherapy, Curtin University*

[2]*School of Human Movement Studies, The University of Queensland*

[3]*Cardiff School of Sport, Cardiff Metropolitan University*

Authors' Notes

Gucciardi was supported by a University of Queensland Postdoctoral Research Fellowship during the preparation of this manuscript. Appreciation is extended to Mark Andersen, Scott Fleming, David Fletcher, Sandy Gordon, Dan Gould, Martin Hagger, Stephanie Hanrahan, Lew Hardy, Ben Jackson, Masato Kawabata, Steve Mellalieu and three anonymous reviewers for providing a critical review on earlier versions of our rejoinder. Any errors or inaccuracies reported in this article rest solely with Gucciardi, Hanton, and Mallett.

*Address correspondence to Daniel Gucciardi, School of Physiotherapy, Curtin University, GPO Box U1987, Perth, Western Australia. Email: daniel.f.gucciardi@gmail.com.

Submitted for Publication: January 18th 2013

*Masked Manuscript without Author Information
Click here to download Masked Manuscript without Author Information: MANUSCRIPT_response to Clough et al_FINAL_ms.doc

Progressing Measurement in Mental Toughness: A Response to Clough, Earle, Perry, and Crust

Daniel F. Gucciardi[1, 2]*, Sheldon Hanton[3], and Clifford J. Mallett[2]

[1]*School of Physiotherapy, Curtin University*

[2]*School of Human Movement Studies, The University of Queensland*

[3]*Cardiff School of Sport, Cardiff Metropolitan University*

Authors' Notes

Gucciardi was supported by a University of Queensland Postdoctoral Research Fellowship during the preparation of this manuscript. Appreciation is extended to Mark Andersen, Scott Fleming, David Fletcher, Sandy Gordon, Dan Gould, Martin Hagger, Stephanie Hanrahan, Lew Hardy, Ben Jackson, Masato Kawabata, Steve Mellalieu and three anonymous reviewers for providing a critical review on earlier versions of our rejoinder. Any errors or inaccuracies reported in this article rest solely with Gucciardi, Hanton, and Mallett.

*Address correspondence to Daniel Gucciardi, School of Physiotherapy, Curtin University, GPO Box U1987, Perth, Western Australia. Email: daniel.f.gucciardi@gmail.com.

Submitted for Publication: January 18th 2013

Abstract

The measurement of mental toughness, which has gained increased popularity among scholars in the past decade, is an area of research that has typically lacked a synergy between theory and method. In a psychometric examination of the Mental Toughness Questionnaire-48 (MTQ48; Clough, Earle, & Sewell, 2002), we proposed several issues that can arise when theory is disconnected from method commensurate with current best practice (Gucciardi, Hanton, & Mallett, 2012). Clough, Earle, Perry, and Crust (2012) offered a critical commentary of our work, citing both substantive (e.g., inadequate literature review) and methodological (e.g., inappropriate samples) issues that they argued limited its contribution to progressing mental toughness measurement. In this article, we respond to these claims by drawing from theory and research. Although these discussions center on the MTQ48, we believe many of the issues have relevance to scholars and practitioners interested in the measurement of psychological variables as they pertain to sport, exercise, and other performance or achievement contexts.

Keywords: confirmatory factor analysis; construct validation; exploratory structural equation modeling; factorial validity; scale development

Progressing Measurement in Mental Toughness: A Response to Clough, Earle, Perry and Crust

Scholarly exchanges between researchers represent an important process by which to progress advancements in knowledge on substantive and methodological issues for scientific inquiry. It is therefore pleasing to see that our article (Gucciardi, Hanton, & Mallett, 2012) has stimulated a scholarly debate on the measurement of mental toughness. We thank Clough, Earle, Perry, and Crust (2012) for their critical evaluation which focused on mental toughness measurement issues. Despite the sole focus on the Mental Toughness Questionnaire-48 (MTQ48; Clough, Earle, & Sewell, 2002), we believe many of the issues that Clough, Earle, Perry et al. discussed are of importance to scholars and practitioners interested in the measurement of psychological variables as they pertain to sport, exercise, and other performance or achievement contexts. Thus, we greatly appreciate the opportunity to respond to the critique of our work, and are excited to contribute further to this intriguing debate.

In formulating their comments on our psychometric examination of the MTQ48, Clough et al. (2012) identified three main areas that relate to their perceptions of (a) an inadequate review of the literature, (b) an inadequate discussion of confirmatory factor analysis (CFA), and (c) the use of inappropriate samples. Our primary aims in this rejoinder are to address the criticisms leveled at our work within each of these three broad areas of concern, and use these discussion points as a foundation upon which to offer practical guidance to those scholars interested in measurement issues. We begin each section with a brief overview of Clough et al.'s comments about our original article. The narrative then turns to a discussion of key features of Clough and colleagues' evaluation that we believe require clarification and/or elaboration. Throughout the article we offer a primer on scale development issues pertinent to each of the discussion points.

Literature Review Issues

According to Clough et al. (2012), the first limitation of our evaluation of the MTQ48 was the "the lack of a truly comprehensive literature review" (p. 283). They argued that a

comprehensive review of the available literature is fundamental to an unbiased account of the current state of affairs. Inherent within their discussion of this limitation of our work is that "numerous errors and inaccuracies" (p. 283) occur when authors do not provide readers with a complete account of the literature that underpins their research. Although we agree on the importance of a literature review for setting the foundation of an empirical paper, several of the points Clough et al. offered to support their critique of our work require clarification.

First, Clough et al. (2012) argued that the primary aim of our paper was "clearly on criticizing the MTQ48" (p. 283). As we indicated in our paper, we were primarily interested in examining the psychometric properties of the measure of mental toughness used most frequently in the peer-reviewed literature. Specifically, we noted, "Despite its ongoing influence on research and practice as the most frequently adopted measure for most researchers and practitioners, the MTQ48 has yet to be subjected to a rigorous psychometric examination" (Gucciardi et al., 2012, p. 202). This summation of the available peer-reviewed literature was the driving force for the primary purpose of our study, which was "to examine the factorial validity of the MTQ48 in two broad achievement contexts" (p. 203). In addition to these statements, we also attempted to offer a concise, yet balanced review of both sport-general and sport-specific measures of mental toughness. In particular, we identified both the strengths and weaknesses of these available measures of mental toughness, two of which have involved the contributions of the lead co-author of our research team. We believe this balanced and collegial approach is consistent throughout our article. For example, we reported the data as generated from our statistical analyses, noting both the strengths (i.e., adequate internal reliability estimates for three of the four subscales, strong theoretical underpinning) and weaknesses (e.g., inconsistencies between the hypothesized model and data) of the MTQ48. Also, consistent with American Psychological Association (2010) guidelines, we provided our raw data to John Perry (one of the authors of the Clough et al. response) upon his request for verification purposes. Clough and colleagues do not appear to dispute the findings of our

analyses per se, only the choice of one of two analytical techniques we employed (i.e., confirmatory factor analysis).

Second, Clough et al. (2012) criticized our group for failing to "obtain as much information about [the MTQ48] as possible" (p. 283). A consistent theme in their narrative was that we overlooked a considerable body of work, such as a key doctoral thesis (Earle, 2007), a recently published book (Clough & Strycharczyk, 2012), and some unpublished evidence generated by researchers who have employed the MTQ48. We do not completely refute this claim; nevertheless, we would like to clarify and elaborate on our rationale for omitting such information. With regard to Clough and Strycharczyk's (2012) book, for example, it is worth noting that our paper was submitted, revised, and accepted for publication several months before their book was publicly available. Thus, we were unable to access and evaluate the information detailed in this book prior to the development of our article. Having obtained a copy and inspected the content of their book, our initial summation of the literature that the MTQ48 has yet to be subjected to a rigorous psychometric examination still holds and, therefore, the original rationale for our research remains. Specifically, although Clough and Strycharczyk overviewed a great deal of research that has examined the criterion-related validity of the MTQ48 across a variety of achievement contexts (e.g., sport, workplace, education), we could not find evidence of empirical data and a description of the criteria on which the adequacy of the measurement model was evaluated other than to say they analyzed their data "using complex statistical techniques such as factor analysis" and a "six-factor mental toughness model had emerged" (p. 42). Typically, this type of information is reported in detail in those books in which a major focus is to overview the foundations and development of a concept and its measurement tool.

As reflected in our previous work (Gucciardi, Mallett, Hanrahan, & Gordon, 2011), it also becomes apparent that we were aware of much of the unpublished research involving the MTQ48, particularly the work reported in the technical manual (AQR, 2007). The focus of our article, however, was on research that had been published in refereed, scholarly outlets because the peer

review process is highly regarded for ascertaining and maintaining the substantive and methodological quality of scientific work (cf. Bornmann, 2011; Brustad, 1999; Marsh, Jayasinghe, & Bond, 2011). Nevertheless, we acknowledge that the peer review process is not perfect and can therefore be biased in some respects (for a review, see Holt & Spence, 2012).

One type of publication bias, which is particularly important in this instance, refers to a phenomenon in which research articles founded on statistically significant results are more likely to be published than are studies that have nonsignificant findings. The issue of publication bias has been an important consideration for psychological science for at least three decades (e.g., Rosenthal, 1979), including sport and exercise psychology (e.g., Spence & Blanchard, 2001), not least because the phenomenon can result in a distorted representation of an area of inquiry. Although the generation of null results can influence the "publishability" of a research article, major methodological or measurement flaws can account for other studies that do not reach the pages of a peer-reviewed outlet (Ferguson & Brannick, 2012). Much of the unpublished work reported in the technical manual provides supportive evidence for the criterion-related validity of the MTQ48. As Clough et al. (2012) emphasized criterion-related validity as the "primary driver" (p. 285) of the construct validation process, inquisitive consumers of scientific knowledge might wonder why this research has not been disseminated in refereed outlets as is custom and practice, and also highly regarded in the scientific community (Bornmann, 2011; Brustad, 1999; Marsh, Jayasinghe et al., 2011). Owing to the limited amount of information reported in the technical manual (AQR, 2007) that relates to key substantive (e.g., justification for the nomological network) and methodological (e.g., choice of experimental design or analytical techniques) decisions, we and other scholars are unable to critically evaluate this work.

A central feature of Clough et al.'s (2012) critique is that we reported a number of mistakes as a result of a limited literature review. In stating that we made "numerous errors and inaccuracies" (p. 283), only one example is provided to support this claim. Specifically, they argued that our "assertion that "75% of the underlying model is hardiness theory" is simply untrue" (pp. 283-284).

They qualified this assertion by noting that their 4C's model of mental toughness was inductively derived from the views of key stakeholder's such as athletes, coaches, and sport psychologists, and that control, commitment, challenge, and confidence were qualities that emerged from qualitative mental toughness research. We do not refute the inductive nature of their theory development process (see Clough et al., 2002, p. 34) and the similarities of the model with subsequent mental toughness research (for reviews, see Gucciardi & Gordon, 2011); we are, however, confused with the claim that we have misinterpreted the contribution of hardiness theory to their 4Cs model of mental toughness.

In their original publication, Clough and colleagues (2002) noted, "Theoretical models within health psychology offer the practitioner insight into the constituents of mental toughness, especially the work associated with psychological hardiness" (p. 37). They further added "it was decided that if a model of mental toughness is to be useful it must…have its roots in established and robust psychological theory, rather than simply be a reflection of current practice" (p. 37). As a concluding statement, Clough and colleagues emphasized "the [4Cs] model developed in this chapter pays a healthy respect in theoretical terms to the "hardiness" approach utilized within health psychology" (p. 38). Elsewhere Clough and his colleagues have reiterated the importance of hardiness theory for the development of the 4Cs model of mental toughness. For example, Clough, Earle, and Strycharczyk (2008) noted that the "main thrust of theoretical development relates to hardiness" (p. 209) whereas Crust and Keegan (2010) acknowledged that "Clough et al. (2002) based their work in the theoretical foundations of hardiness" (p. 164). Hardiness theory, which dates back over 30 years (Kobasa, 1979; Maddi & Kobasa, 1984), is conceptualized as a cognitive personality variable consisting of a sense of control, commitment, and challenge. As is clearly evident in the aforementioned quotations, the three components of hardiness were an important consideration for the content of Clough and colleagues' mental toughness model and measurement tool (i.e., the MTQ48).

Clough and colleagues (2002) viewed the addition of confidence to the hardiness model as an important feature that linked "psychological theory and applied sport psychology" (p. 38). Moreover, they argued that confidence was "an important factor relating to sport performance" which "has not been considered as a distinct element in previous models of hardiness" (p. 38). Again, we do not dispute the importance of this extension of hardiness theory, as we have previously acknowledged the inclusion of confidence into mental toughness conceptualization is supported by evidence from qualititative research (e.g., Gucciardi et al., 2011). With the three components of control, commitment, and challenge being derived from hardiness theory (and subsequently supported by the views of key stakeholders; see Clough et al., 2002), it appears that confidence is the only unique dimension to the 4Cs conceptualization of mental toughness. Also, Crust and Azadi (2010) noted, "Initial qualitative work by Clough et al. involved athletes, coaches, and sport psychologists, and resulted in confidence being added to the three original components of hardiness" (p. 44). In a recent paper (Perry, Clough, Crust, Earle, & Nicholls, 2013), Clough and his colleagues have again acknowledged that confidence was added to the hardiness model in a paper published after their response paper (Clough et al., 2012). Given the authors' published statements about hardiness, it is not difficult to see how we would have arrived at this conclusion, and we still feel that the 4Cs of mental toughness is predicated primarily on hardiness theory.

In summary, we agree with Clough et al. (2012) that the literature review is fundamentally important for the production of high quality scholarly articles, but disagree that it must be *comprehensive*. In particular, it is not practically possible to include "everything" in the introduction to an empirical paper. Owing to page restrictions imposed by most refereed, scholarly outlets, it is important that authors develop a conceptually coherent and concise literature review in which they introduce the problem, develop the background, and state the purpose and rationale "in just a few pages" (American Psychological Association, 2010, p. 27). Rather than providing a "truly comprehensive literature review" (Clough et al., 2012, p. 283), therefore, scholars are encouraged to develop a selective review that encompasses the most relevant and appropriate information for

building a convincing rationale for the purpose of their study. To achieve this goal, Sternberg and Sternberg (2010) suggested that an

> …introduction should answer four basic questions: (a) what previous research led up to your research? (b) what does your research add to this previous research? (c) why is the addition made by your research important or interesting? and (d) how is the addition made? (p. 106).

In other words, researchers should identify a gap in the literature (e.g., conflicting theory and evidence, methodological flaws in previous research), justify the importance of addressing the gap (e.g., conceptual evolution, applied benefits), and introduce the way in which they intended on closing or reducing the gap (e.g., empirically test competing theories). We stand by our literature review in our paper as we believe that we adequately addressed each of these recommended criteria using peer-reviewed sources, and encourage readers to form their own critique of our work.

Confirmatory Factor Analysis

Clough et al. (2012) identified a "complete lack of a discussion of the strengths and the limitations of the confirmatory factor analysis (CFA) approach" (p. 283) as the second concern with our psychometric evaluation of the MTQ48. Inherent within their discussion of this issue is the notion that the MTQ48, as with most other measures of personality traits, was always destined to fail the CFA test. They argued for greater breadth in the construct validation process beyond that which focuses on internal structure (i.e., factorial validity) and noted criterion-related validity as the "primary driver" (p. 285) for assessing the usefulness of a psychometric tool. Although we agree with one aspect of their critique (i.e., CFA is not the only approach to examining the factorial validity of an instrument), much of their discussion on this point seems to be missing several key issues in our article.

First and foremost, as noted in our article, and reiterated here, CFA is a particularly pertinent data analytical technique when there is a strong theoretical base for the hypothesized measurement model (Russell, 2002; Williams, 1995). However, because CFA encompasses a number of highly

restrictive parameters in the estimation process (e.g., all nontarget loadings are constrained to be zero), the application of this analysis typically requires a clearly articulated theoretical structure (Asparohov & Muthén, 2009). Consistent with these viewpoints, we noted in our paper "CFA may not be suitably justifiable as an analytical approach for the assessment of the MTQ48" (Gucciardi et al., 2012, p. 203). Indeed, we dedicated almost one published page (p. 203) on a critical discussion of the substantive (e.g., theoretical rationale for choosing a data analysis technique) and empirical (e.g., model fit, parameter estimates) merits of both CFA and exploratory structural equation modeling (ESEM).

Second, Clough et al. (2012) argued that we relied primarily on "rough guidelines about goodness of fit as if they were "golden rules" (p. 284) for model fit and the assumption that "an instrument must "pass" the CFA test" (p. 285). With regard to this criticism, in our article we acknowledged the importance of exercising caution in strict adherence to model fit statistics when evaluating the usefulness of a statistical model: "However, caution has been urged in the strict adherence to such [model-fit] recommendations in psychometric evaluations of measures comprising 50 or more items loading onto five or more factors" (Gucciardi et al., 2012, p. 205). Even if one were to adopt the most liberal levels of acceptable fit for the multiple criteria of model fit, the hypothesized measurement model of the MTQ48 would still have been deemed inadequate in both our athlete and workplace samples. Nevertheless, we also considered standardized factor solutions to evaluate the significance and strength of the parameter estimates for each of the models tested. Model fit statistics aside, our examination of the factor solutions indicated a significant degree of inconsistency between the hypothesized four factor structure and our data in both the athlete and workplace samples (e.g., several significant unintended cross-loadings). With regard to the criticism that an instrument must pass the CFA test, our inclusion of ESEM as an additional assessment of factorial validity beyond CFA is consistent with recent recommendations (e.g., Hopwood & Donnellan, 2010; Schmitt, 2011) and represents evidence that we did not adhere to a unitary focus with a highly restrictive, hypothesis-testing analysis.

Within the broader context of their concerns regarding CFA for the psychometric evaluation of surveys, Clough et al. (2012) acknolwedged that it was not completely unexpected that the MTQ48, much like other complex, multidimensional instruments, would not fair well when subjected to this highly restrictive data analysis (cf. Marsh et al., 2009). As noted in our article, and reiterated previously in this response, we addressed this potential limitation by subjecting the MTQ48 to *both* CFA and ESEM. Although ESEM addresses some of the limitations of CFA (e.g, allows items to cross-load on unintended factors), Clough et al. failed to acknowledge this methodological feature in their critique of our article. Nevertheless, it is important to recognize that there are many examples of complex, multidimensional psychological tools that *have* received clear support for their factorial validity when subjected to CFA. Such tools include, but are not limited to, the Physical Self Description Questionnaire (see Marsh, Martin, & Jackson, 2010), Genos Emotional Intelligence Inventory (see Gignac, 2010), Motivation and Engagement Scale (see Liem & Martin, 2012), and the Flow Scales (see Jackson, Martin, & Eklund, 2008). Each of these psychometrically robust measures shares the common theme of a strong theoretical underpinning *and* empirical development commensurate with the evolution of substantive and methodological best practices.

Third, Clough et al. (2012) noted "an independent research group has [performed a CFA] and found that the MTQ48 was acceptable" and that we "appear to dismiss this work even, though Horsburgh et al. published their work in a quality peer-reviewed journal" (p. 286). We would like to think that we did not dismiss these findings. As previously discussed in this rejoinder, we reviewed all available peer reviewed research in our paper including the work of Horsburgh, Schermer, Veselka, and Vernon (2009). Nevertheless, in Horsburgh et al.'s article, the lack of detailed empirical data and methodological information (i.e., complete data analysis processes, criteria upon which model fit was assessed) means that readers are unable to fully evaluate the conclusions forwarded by the authors. For example, statements such as "the four-factor solution provided a better fit to the data than did a single factor" and "the pattern matrix suggested that the items fit

moderately well onto the designated factors" (Horsburgh et al., 2009, p. 102) do not provide sufficient information upon which to ascertain the veracity of these conclusions unless they are support by detailed empirical data such as model-fit statistics and factor loadings (cf. American Psychological Association, 2010, pp. 32-34).

We are also at odds with Clough et al.'s (2012) criticism that "placing [CFA] at the center of test development is an opinion rather than a fact or even best practice" (p. 285). In discussing the appropriateness of CFA for Big Five (i.e., personality) research, for example, Marsh et al. (2010) noted:

> However, many of the methodological and statistical advances in quantitative psychology in the last 2 decades are associated with latent-variable approaches such as CFA and structural equation models (SEMs). Hence, failure to embrace these new and evolving methodologies (throwing the baby out with the bathwater) would have dire consequences–particularly for a field of research so fundamentally based on factor analysis (p. 472).

Others have echoed these sentiments (e.g., Hinkin, 1995), with sport and exercise psychology scholars referring to CFA as the "preferred state-of-the-art method" (Hagger & Chatzisarantis, 2009, p. 513). Contrary to Clough et al.'s claim, the importance of CFA for test development and evaluation is not solely our opinion of a core component of current best practice, but is embedded within the psychology community where the MTQ48 is frequently cited (e.g., Marsh, 2007; McCrory, & Layte, 2012; Lafrenière, Verner-Filion, & Vallerand, 2012).

As we have acknowledged elsewhere (Gucciardi et al., 2011), other forms of reliability (e.g., test-retest) and validity (e.g., criterion) are important to the construct validity enterprise, but it is important that scale developers and evaluators establish support for the factorial validity of a test before proceeding to others forms of validity because internal structure has important implications for interpreting relationships with external variables (Gignac, 2009; Marsh, Martin et al., 2010). A

fundamental prerequisite and consideration to sound factorial validity is an adequately defined construct and its conceptual domain because:

> ...many researchers *think* they have a clear idea of what they wish to measure, only to find out that their ideas are more vague than they thought. Frequently, this realization occurs after considerable effort has been invested in generating items and collecting data—a time when changes are far more costly than if discovered at the outset of the process (DeVellis, 1991, p. 51).

Furthermore, we believe Clough and colleagues (2002) have provided limited information on their rationale for the conceptual underpinnings of the MTQ48 when they first introduced the tool, and uncertainty appears to remain in the literature despite Clough et al.'s (2012) apparent disagreement. For example, Clough et al. (2002) initially proposed a four-factor model of mental toughness, yet the 4Cs model has been extended to encapsulate a six-factor (Clough & Strycharczyk, 2012) and nine-factor solution (e.g., Horsburgh et al., 2009) with limited discussion on the rationale and evidence for these approaches. Also, Clough et al. (2012) acknowledged, "It is true that other researchers have used differing models, but this is acceptable within the design of the test" (p. 284). We could find no evidence or theory to substantiate this claim of acceptance with test variation among other scholars. For Clough et al. to suggest that factor analytic techniques are not essential for construct validity and that it is acceptable to use different models without empirical or theoretical substantiation is "problematic to the extent that it could lead to an "anything goes" mentality" (Hopwood & Donnellan, 2010, p. 341).

Despite their criticisms of the CFA approach, Clough et al. noted that their "own CFA, which is currently under review, reaches a different conclusion from that of the Gucciardi [*sic*] paper" (p. 286). Specifically, since the publication of our article and Clough and colleagues' (2012) response, two factorial validity assessments of the MTQ48 have emerged. First, Gerber et al. (in press) sampled high school students in the German-speaking part of Switzerland (n = 284; M_{age} = 18.30, SD = 4.17) and undergraduate exercise and health science students (n = 140; M_{age} = 20.00,

SD = 5.00) [nationality not reported]. CFA did not support the hypothesized six-factor structure of the MTQ48 in both the high school and university student samples, as well as a combined sample, although standardized factor loadings were supportive of the measurement model. In contrast, model fit indices generated using ESEM on the combined sample indicated that the six-factor measurement was adequate; however, standardized factor loadings identified several inconsistencies between the hypothesized measurement model and the data (i.e., significant cross-loadings, insignificant loadings on intended factors). Second, Perry et al. (2013) sampled senior managers (*n* = 4342), lower and middle managers (*n* = 1440), clerical/administrative workers (*n* = 1004), athletes (*n* = 442), and students (*n* = 978). Model fit indices and standardized factor loadings generated using CFA were generally supportive of the six-factor model across all samples, with the exception that both the CFI (.779 to .857) and TLI (.766 to .848) were below the minimum recommended level of .90 in all samples. In contrast, ESEM revealed support for the hypothesized six-factor model whereby all model fit indices surpassed minimum recommended levels (i.e., CFI, TLI >.90, RMSEA, SRMR <. 08) except for TLI (.884) in the athlete sample; however, standardized factor loadings identified several inconsistencies between the hypothesized measurement model and the data (i.e., significant cross-loadings, insignificant loadings on intended factors).

In summary, despite subsequently adopting the approach themselves (Gerber et al., in press; Perry et al., 2013), Clough et al. (2012) have questioned the use of CFA as an important methodological consideration for the measurement of mental toughness and other psychological constructs. Although we have offered several counter-arguments with many of the points Clough et al. noted to support their critique (e.g., inadequate discussion of the pros and cons of CFA, "golden rules"), as with others (e.g., Marsh et al., 2009) we have explicitly acknowledged that CFA is limited in some respects because it is a highly restrictive approach to data analysis. A fundamental implication of these methodological issues is the importance of substantive clarity when developing new constructs and theories (for a review, see MacKenzie, Podsakoff, & Podsakoff, 2011). When a clearly articulated theoretical model underpins measurement, CFA is considered the most

appropriate test (Hagger & Chatzisarantis, 2009; Russell, 2002; Williams, 1995); however, ESEM offers a viable alternative when the theoretical underpinnings are less established (Asparohov & Muthén, 2009; Marsh et al., 2009).

Samples and Methods

The use of inappropriate samples is the final criticism Clough et al. (2012) offered of our work. In particular, Clough et al. highlighted that although our "initial [athlete] sample is inadequate, the second [workplace] sample is perhaps worse" (p. 286). Specific criticisms of our samples included the use of incentives (i.e., course credit for students), inclusion of participants from two different nations (i.e., Australia, United Kingdom), online survey methodology, recruitment through an online panel of respondents, and the generalizability of the results from our samples. Although we agree that these issues are important considerations for ascertaining the quality of our work, several of the points offered to support Clough et al.'s critique are worthy of discussion.

We would like to clarify two points in Clough et al.'s (2012) discussion of our samples. First, at no point in our narrative did we "assume that [the athlete] sample is in any way representative of the general population" (p. 286) or state that our workplace sample was "representative of the Australian community in general" (p. 286). More broadly, we took great care to develop a well-balanced narrative that did not go beyond our design and data. Our concluding section offers an example of this approach: "The current findings provide *preliminary* [italics added] evidence that the psychometric properties of the MTQ 48 *may* [italics added] not be adequate, particularly with respect to its hypothesized underlying conceptual model" (Gucciardi et al., 2012, p. 211). Second, Clough et al. highlighted "The first sample used was students who take part in sport (although you have to search for this information, as it is not reported in the sample section)" (p. 286). In actual fact, we plainly depicted our athlete sample in the "Participants" section as individuals who participated in individual or team sports; we noted the recruitment of athletes via tertiary education settings in the "Procedures" section (see Gucciardi et al., 2012, p. 204).

With regard to the suitability of our athlete sample, Clough et al. (2012) highlighted: "Although it might be reasonable to use a student population to test out a range of psychological theories, tools, and techniques, it is highly questionable to use this type of sample for a construct validity study" (p. 286). They further added the athlete sample was inappropriate because it did not have sufficient variance across the variable of interest[1]. To clarify, total mental toughness scores in our samples ranged from 110 to 216 (M = 161.78, SD = 17.08) and 144 to 232 (M = 168.69, SD = 19.62) for the athletes and employees, respectively, on a possible range of 48 to 240. This range of total mental toughness is comparable with previous research that has employed the MTQ48 (e.g., Crust & Azadi, 2010; Crust & Keegan, 2010; Nicholls, Polman, Levy, & Backhouse, 2008). Item level analyses revealed that participants used the entire 5-point scale to respond to each item – with the exception of item 4 in the athlete sample – and all items were normally distributed in both the athlete (skew: -.82 to .56; kurtosis: -1.28 to 1.30) and employee (skew: -1.05 to .50; kurtosis: -1.11 to 2.92) samples (Kline, 2011). The criticism that our data did not have sufficient variance across the variable of interest could also mean that our sample did not contain athletes who were serious sport participants, or had sufficient variance with regard to the level of achievement or experience. An inspection of the participant demographics (Gucciardi et al., 2012, p. 204) indicates that such a criticism of our sample would be misleading. We sampled a roughly equal number of male (n = 354) and female athletes (n = 328) with varying levels of playing (i.e., 1 to 35 years; M = 9.11; SD = 4.41) and competitive experience (international = 9%; national = 26%; state or county = 25%; district or local = 39%).

Interestingly, an inspection of the MTQ48 technical manual (AQR, 2007) reveals that 64% (n = 619) of the development sample Clough and his colleagues employed to examine the factorial validity of the MTQ48 using principal components analysis with varimax rotation were students

[1] It is unclear what Clough et al. (2012) actually meant when they criticized our samples for not encompassing "the full range of the trait under investigation" (p. 286). Perry et al. (2013) have subsequently reported this aspect of their participant samples as being a key strength, with little discussion on what it means to include "the full domain of possible expressions of mental toughness" (p. 589) or empirical data to support this conclusion. Nevertheless, we interpreted this statement according to both empirical (e.g., item level descriptive statistics) and substantive issues (e.g., level of achievement or experience).

(see also Perry et al., 2013). Also, despite their claim regarding the inadequacy of student samples for the construct validation process, Clough et al. (2012) found it appropriate to cite research that involved students (McCoy, Marks, Carr, & Mbarika, 2004) to support their claim that our administration method (i.e., online survey) could have biased the results. Regardless of these conflicting views by Clough et al. it is important to reiterate that we recruited *athletes* via university settings such that a condition of participation in the study was that individuals were, at the time of completing the survey, engaged in competitive sport. This recruitment approach has been adopted by MTQ48 researchers (e.g., Crust & Keegan, 2010) and other sport psychology scholars (e.g., Davies, Lane, Devonport, & Scott, 2010). Our instructions both in the recruitment strategy, information sheet and the questionnaire package reinforced this aspect of our methodology to all participants. For example, the following quotation is taken directly from our information sheet: "This survey contains questions that relate to several psychological skills that are important for achievement in sport. It is essential that you provide your own honest opinion about how you currently view yourself as a sport performer/athlete." Additionally, we used the following instructional set before the MTQ48 items: "Please indicate how much you agree with each of the following statements as it relates to how you typically think, feel, and behave as an athlete or sport performer."

Clough et al. (2012) reported a number of criticisms with our athlete sample including the use of incentives, and no information on the percentages of Australian (n = 352) and British (n = 334) athletes. Clough et al. however, did not detail the specific implications of these criticisms for factorial validity other than to suggest our sample might not have been representative of the population. Nevertheless, we will address the general areas of concern. Incentives are commonly employed to increase response rates in survey research (Dillman, Smith, & Christian, 2009), and when implemented with student (e.g., Yeo & Frederiks, 2011) or student athlete (e.g., Williams & Cumming, 2012) samples, they do not appear to undermine the psychometric properties of measurement instruments. Additionally, online survey approaches with athletes have been shown to

produce higher and faster response rates, and yield fewer missing responses than hardcopy methods (Lonsdale, Hodge, & Rose, 2006). Nevertheless, despite the findings reported elsewhere, we cannot rule out the influence of incentives with the athlete sample in our paper.

Clough et al. (2012) also criticized our inclusion of participants from two different nations, noting that the "situation is made even worse by using samples from Australian and UK participants. No information is reported about the percentages of each of these [two nations]" (p. 286), but they did not explicitly detail why or how the inclusion of both Australian and British participants limits the robustness of our findings. As alluded to in our article, we recruited British athletes in an attempt to enhance our statistical power. Our rationale for recruiting Australian and then British athletes was that both are English speaking countries, each is well-versed in completing such psychological tests, and the United Kingdom and Australia is where the majority of mental toughness research has been conducted. Nevertheless, we have previously noted that it is risky for researchers to assume that the same construct is being assessed across different groups (Gucciardi et al., 2011). Additionally, including two potentially culturally different groups introduces the possibility that we obtained spurious correlations between items because of significant differences in levels reported. For example, if Australian athletes scored higher than British athletes on items 1 and 2, and items 1 and 2 are uncorrelated within each sample, then items 1 and 2 will be correlated when the samples are collapsed. Similarly, if items 1 and 2 are correlated in the "opposite" direction to the correlation induced by the cultural differences, then collapsing across samples could remove a significant correlation and in so doing reduce fit. Therefore, in the spirit of transparency and scholarly debate, we reanalyzed the data separately for the Australian and British samples to further assess the adequacy of the hypothesized MTQ48 model and the potential influence of integrating data from two different groups into the one analysis[2]. Collectively, the multiple criteria of model-data fit (see Table 1) and factor loadings (see Table 2) for both CFA and ESEM indicated that the

[2] It is worth noting that the Australian and British athletes also differed with respect to the recruitment process, such that the British athletes did not receive course credit for their participation but the Australia athletes were provided with this incentive.

hypothesized correlated, four factor model of the MTQ48 was unsatisfactory in both the Australian and British athletes.

The lack of information pertaining to the response rates was another criticism of our sampling strategy. Specifically, Clough et al. (2012) noted that "Because of the recruitment method used, it is not possible to calculate any meaningful response rates; therefore, it is impossible to have any feel about the adequacy of the sample" (p. 286). We do not refute this criticism of our work, as we were unable to determine the response rates because of the convenience sampling approaches. Interestingly, an inspection of those studies detailed in our original article that have employed the MTQ48 as a measure of mental toughness and subsequent factorial validity examinations (Gerber et al., in press; Perry et al., 2013) reveals that none included information on response rates; thus, one could direct the same criticism at this body of research. Nevertheless, it is important to recognize that sample representativeness can be more important than high response rates: "But it is not necessarily true that representativeness increases monotonically with increasing response rate...recent research has shown that surveys with very low response rates can be more accurate than surveys with much higher response rates" (Krosnick, 1999, p. 540). As previously noted, the demographics of our athlete participants revealed that we sampled a roughly equal number of male and female athletes from a variety of team and individual sports who were generally in their early twenties. Perhaps most important for representativeness, in our opinion, was that our athlete sample had varying degrees of experience in their sport and encompassed diverse expertise levels (i.e., international, national, state or county, district or local competition). These demographic details share considerable overlap with previous research involving the MTQ48 and a similar number of athletes (e.g., Nicholls et al., 2008).

Clough et al. (2012) were also critical of our sampling of employees using an online panel. Online panels consist of members of the general public who have agreed to occasionally take part in web-based studies. Benefits of online panels include short data collection times, availability of panelists' historical and profile data, increased response rates, and enhanced potential to easily

obtain large samples (for a review, see Göritz, 2007). Not surprisingly, online panels are increasingly being used for research across a variety of disciplines including the social (Tortora, 2009), psychological (Göritz, 2007), and medical sciences (Couper, 2007). Although we cannot be certain of the influence of this methodological approach on our data, elsewhere online panel recruitment methods have not affected the psychometric properties of psychological tools (see e.g., Lucas, Young, Zhdanova, & Alexander, 2010; Piccolo & Colquitt, 2006).

With regard to the use of online surveys, this mode of survey administration was cited by Clough et al. (2012) as a methodological limitation of our psychometric examination of the MTQ48. They cited McCoy et al. (2004) as evidence to support their claim that results can be biased when collected online. McCoy et al. however, conducted *t*-tests of raw scores and correlations of subscales between administration methods using a within-subjects design, and their analyses did not take measurement error into consideration, nor did they consider the internal structure of a measurement instrument. As our psychometric examination of the MTQ48 focused on internal structure (i.e., factorial validity), it is important to consider research that has examined this aspect of validity using rigorous statistical analyses that consider measurement equivalence. Specifically, researchers have revealed that administration method (i.e., online versus paper and pencil) has little to no effect on the psychometric integrity of tools when used in sport (e.g., Lonsdale et al., 2006), workplace (e.g., De Beuckelaer & Lievens, 2009), and education contexts (e.g., Chuah, Drasgow, & Roberts, 2006). Methodological concerns about online research have not been supported by several investigations (e.g., Gosling, Vazire, Srivastava, & John, 2004). Despite this criticism of our methodological approach, Clough and Strycharczyk (2012) have highlighted that the MTQ48 "is most commonly completed in online format" (p. 47; see also Perry et al., 2013).

Also related to the use of online survey methodology, Clough et al. (2012) highlighted that "it is not clear from the paper how security of the material was maintained, in what environment the questionnaires were completed, or how it could be verified who was actually completing the instrument" (p. 286). Both SocialSci and SurveyMonkey, the online survey platforms we employed

with the students and employees respectively, encrypt survey responses using secure sockets layers to protect the transmission of participant responses to the online database. As an additional measure to maintain participant confidentiality and anonymity, these online survey platforms do not collect the internet protocol addresses of respondents. Owing to the implementation of these processes for maintaining anonymity of the participating athletes and employees, we were unable to explicitly verify who completed the survey. To do so would have violated our human research ethics requirement for this study to maintain the anonymity of participants. It is important to note, however, that all athletes were recruited via university settings and only those individuals officially enrolled in an undergraduate course were provided with a secure web address. Additionally, SocialSci leverages both direct (e.g., qualifying questions prior to the main survey) and indirect information (e.g., analysis of survey and per page completion times across participants to determine if respondents are taking the survey seriously) about participants in their pool to gauge their honesty and determine their eligibility to participate in surveys. Nevertheless, we cannot detail the actual environment in which participants completed the questionnaire, other than to say that a computer or another mobile device (e.g., smartphone) and internet access was required.

The aim of our response in this paper is not only to respond to the criticisms of our work, but also to further the development of the mental toughness knowledge base. Therefore, as an additional check on the quality of our samples, we reexamined the data after removing participants who responded with the same score, or patterns of scores, to each item across the entire survey and the second half of the survey (e.g., boredom) because such noise could account for the poor model fit. We identified 1 athlete and 1 employee who responded with the same response category (3 or neutral) across the entire survey; no participants responded to the second half of the survey in the same way. Removal of the athlete and employee revealed that the data did not evidence a good fit with the hypothesized model: CFA of athlete sample, $\chi^2(1074) = 5508.86$, $p < .001$, CFI = .487, TLI = .461, SRMR = .104, RMSEA = .078 (90% CI = .076 to .080); CFA of workplace sample, $\chi^2(1074) = 4917.26$, $p < .001$, CFI = .521, TLI = .497, SRMR = .093, RMSEA = .075 (90% CI =

.073 to .077); ESEM of athlete sample, $\chi^2(942) = 2966.75$, $p < .001$, CFI = .766, TLI = .720, SRMR = .045, RMSEA = .056 (90% CI = .054 to .058); and ESEM of workplace sample, $\chi^2(942) = 2804.77$, $p < .001$, CFI = .768, TLI = .722, SRMR = .045, RMSEA = .056 (90% CI = .053 to .058). Collectively, these findings indicate that the removal of potentially problematic data points did little to improve the overall fit of the hypothesized MTQ48 model with the data.

In summary, Clough et al. (2012) have raised several issues related to our samples and recruitment strategies that have important methodological considerations for the measurement of mental toughness and other psychological constructs. Nevertheless, although we have offered evidence to suggest that Clough et al.'s criticisms of our samples and recruitment strategies (e.g., incentives, student athlete samples, online panels, survey administration) may not be as clear cut as they have suggested, we acknowledge that these issues require consideration in future research involving the MTQ48 and psychological assessment tools in general.

Progressing the Measurement of Mental Toughness: Concluding Thoughts

At the outset of their response paper, Clough et al. (2012) noted, "In reality, [Gucciardi et al.'s 2012] paper does little to progress the area under investigation" (p. 283). Having considered and responded to Clough and colleagues' critique of our article, as well as subjected it for critical assessment from several experts acknowledged on this manuscript, we believe the opposite is true. Validation is "the process of compiling evidence that supports the interpretations and uses of data and information collected" (Joint Committee on Standards for Educational Evaluation, 1994, p. 145). Furthermore, the International Test Commission (ITC) stated in their guidelines for measure use that users should "use tests only for those purposes which relevant and appropriate validity evidence is available" and avoid those tests "that have inadequate or unclear supporting technical information" (ITC, 1999, p.12). Examination of factorial validity is one of several important components in the validation process (e.g., Gignac, 2009). If no empirical information or data are available to support the hypothesized factor structure of a measurement tool, its validation process

cannot be progressed, and consequently users of the test cannot use the instrument confidently and interpret its scores appropriately.

Given that it is difficult to find a widely used tool in the psychological sciences that has not been factor analyzed at least twice, it begs the question as to why this analysis had not previously been performed on the MTQ48. Ideally, fundamental information of a multidimensional scale, including its factorial validity, should be provided by its scale developer using rigorous data analysis procedures commensurate with current best practices (e.g., Hagger & Chatzisarantis, 2009; Hinkin, 1995; Marsh et al., 2010). Unfortunately, however, at the time we initiated data collection for our study (March, 2011) as well as the publication of our manuscript, this process has not occured for the MTQ48 both in the original publication (cf. Clough et al., 2002; see also, Clough & Strycharczyk, 2012) and technical manual (AQR, 2007). Given that Clough and his colleagues, as well as those scholars who have employed the MTQ48 in their research, had not evaluated its factor structure using current best practices, it was inevitable that someone at sometime would test the factorial validity of the hypothesized model.

We believe our original article (Gucciardi et al., 2012) was an essential part of the validation process because the aim of the original paper was to provide information to judge whether the hypothesized factor structure was tenable. Although an independent group had offered support for the hypothesized model of the MTQ48 (Horsburgh et al., 2009), they did not offer verifiable information to justify their claims (i.e., complete data analysis processes, criteria upon which model fit was assessed, model fit statistics, solution estimates). Nevertheless, there was the possibility that we might have obtained information that would support the hypothesized factor structure of the MTQ48. However, the information we found through our data analyses was not one that supports the factor structure hypothesized by Clough and colleagues. In particular, we used multiple criteria from both CFA and ESEM to evaluate the validity of the hypothesized measurement and converged on our conclusions based on whether or not the instrument conformed to each of those criteria in two independent samples: multiple indices of model fit, solution estimates (e.g., cross-loadings),

discriminant validity, and face validity. Failing one of the multiple criteria does not provide impetus for abandoning or modifying an instrument; however, failing a number of these criteria across two data analyses and two independent samples raises many questions (cf. ITC, 2009, p. 12).

In summary, it is our hope that the scholarly debate in this journal has much to offer for progressing current approaches to the measurement of mental toughness, and more broadly that many of the key messages and issues discussed help scholars interested in assessment approaches in sport and exercise psychology. We also appreciate the time and effort Clough et al. have taken to consider our research and develop a response. Such "accelerated" scholarly debates are important processes for fast tracking knowledge development in areas of investigation that can otherwise take several years to occur. Indeed, aligned with our original discussion of directions for future research (Gucciardi et al., 2012), it appears that our recommendation to "report empirical data pertaining to the factorial validity of the MTQ 48 when this tool has been employed a measure of mental toughness" (p. 211) has already been accepted by peer reviewers (see Gerber et al., in press, supplementary information) and mental toughness scholars (Perry et al., 2013).

In this rejoinder, we have highlighted several reservations about the criticisms of our work offered by Clough et al. and substantiated these doubts with existing theory and research. We also believe many of the criticisms leveled at our work by Clough and colleagues are contradictory. For example, Perry et al. (2013) acknowledged the addition of confidence to hardiness theory as the basis for the 4Cs model once again after the publication of their original response paper; employed CFA (and ESEM) to examine the factorial validity of the MTQ48; did not report response rates and detailed information regarding the demographics of their participants (e.g., nationality, level of experience); used a student sample as part of a contstruct validity study; and collected data using both online and hardcopy methodologies. Nevertheless, to clarify, we do not argue that the MTQ48 is useless and should be discarded, but rather that the basis for its use (i.e., test development information, conceptual rationale, nomological validity), at the time of our original publication, had not yet been reported in detail in the peer reviewed literature. When our data are considered

alongside recent factorial validity studies (Gerber et al., in press; Perry et al., 2013) that have emerged since the publication of our article, it seems reasonable to conclude that there is some uncertainty regarding the adequacy of the hypothesized model and its operationalization in the MTQ48. Thus, a key recommendation for scholars who employ the MTQ48 as a measure of mental toughness is to report empirical evidence on its internal structure (i.e., factorial validity and internal reliability) on a sample-by-sample basis. Perhaps most important, we believe it would be worthwhile to revisit the conceptual underpinnings of the MTQ48, as well as the extent to which the tool reliably and validily captures this conceptualization of mental toughness. Indeed, the criteria for assessing whether an instrument measures the intended concept depend on the underlying definition (MacKenzie et al., 2011). Good science requires the testing of theoretical or methodological propositions. If existing theory or method withstands the test or challenge, it is strengthened; even if it does not, knowledge is advanced.

References

American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

AQR (2007). *Mental Toughness Questionnaire 48 technical manual.* Retrieved on November 19th from http://www.aqr.co.uk/sites/default/files/MTQ48%20Technical%20Manual%20Jan%202007.pdf.

Asparohov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16,* 397–438. doi: 10.1080/10705510903008204.

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology, 45,* 199–245.

Brustad, R. J. (1999). Editorial perspective: The contribution of the manuscript-review process to knowledge development in sport and exercise psychology. *Journal of Sport & Exercise Psychology, 21,* 307–312.

Chuah, S. C., Drasgow, F., & Roberts, B. W. (2006). Personality assessment: Does the medium matter? No. *Journal of Research in Personality, 40,* 359–376. doi: 10.1016/j.jrp.2005.01.006

Clough, P., Earle, K., Perry, J. L., & Crust, L. (2012). Comment on "Progressing measurement in mental toughness: A case example of the Mental Toughness Questionnaire 48" by Gucciardi, Hanton, and Mallett (2012). *Sport, Exercise and Performance Psychology, 1,* 283-287. doi: 10.1037/a0029771

Clough, P., Earle, K., & Strycharczyk, D. (2008). Developing resilience through coaching: MTQ48. In J. Passmore (Ed.), *Psychometrics in coaching: Using psychological and psychometric tools for development* (pp. 208–223). London, England: Kogan Page.

Clough, P., & Strycharczyk, D. (2012). *Developing mental toughness: Improving performance, wellbeing and positive behavior in others*. London, England: Kogan Page.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64,* 464–494. doi: 10.1086/318641

Crust, L., & Azadi, K. (2010). Mental toughness athletes' use of psychological strategies. *European Journal of Sport Science, 10,* 43–51. doi:10.1080/17461390903049972.

Crust, L., & Keegan, R. (2010). Mental toughness and attitudes to risk-taking. *Personality and Individual Differences, 49,* 164–168. doi:10.1016/j.paid.2010.03.026.

Davies, K. A., Lane, A. M., Devonport, T. J., & Scott, J. A. (2010). Validity and reliability of a Brief Emotional Intelligence Scale (BEIS-10). *Journal of Individual Differences, 31,* 198-208. doi: 10.1027/1614-0001/a000028

De Beuckelaer, A., & Lievens, F. (2009). Measurement equivalence of paper-and-pencil and internet organizational surveys: A large scale examination in 16 countries. *Applied Psychology – An International Review, 58,* 336-361. doi: 10.1111/j.1464-0597.2008.00350.x

DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.

Dillman, D. A., Smith, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method* (3rd ed.). Hoboken, NJ: Wiley.

Earle, K. (2007). *The construct of mental toughness: A psychometric and experimental analysis.* Unpublished doctoral dissertation, The University of Hull, United Kingdom.

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17,* 120–128. doi: 10.1037/a0024445

Gerber, M., Kalak, N., Lemola, S., Clough, P. J., Perry, J. L., Pühse, U., ... Brand, S. (in press). Are adolescents with high mental toughness levels more resilience against stress? *Stress and Health*. doi: 10.1002/smi.2447

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, NY: Springer.

Gignac, G. (2010). Seven-factor model of emotional intelligence as measured by Genos EI: A confirmatory factor analytic investigation based on self- and raterreport data. *European Journal of Psychological Assessment, 26,* 309–316. doi: 10.1027/1015-5759/a000041

Göritz, A. S. (2007). Using online panels in psychological research. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U.-D. Reips (Eds.), *The Oxford handbook of internet psychology* (pp. 473–485). Oxford, England: Oxford University Press.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59,* 93–104. doi: 10.1037/0003-066X.59.2.93

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. doi: 10.1037/a0027190

Gucciardi, D. F., & Gordon, S. (Eds.) (2011). *Mental toughness in sport: Developments in research and theory.* Abingdon, England: Routledge.

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

Hagger, M. S., & Chatzisarantis, N. L. D. (2009). Assumptions in research in sport and exercise psychology. *Psychology of Sport and Exercise, 10,* 511-519. doi:10.1016/j.psychsport.2009.01.004

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21,* 967–988. doi: 10.1177/014920639502100509

Holt, N. L., & Spence, J. C. (2012). A review of the peer review process and implications for sport and exercise psychology. *Athletic Insight, 4(1).* Available from http://www.athleticinsight.com.

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review, 14,* 332–346. doi: 10.1177/1088868310361240

Horsburgh, V. A., Schermer, J. A., Veselka, L., & Vernon, P. A. (2009). A behavioral genetic study of mental toughness and personality. *Personality and Individual Differences, 46,* 100-105. doi:10.1016/j.paid.2008.09.009.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior, 18,* 667–683. doi:10.1002/(SICI)1099-1379(199711)18:6_667::AID-JOB874_3.0.CO;2-T

International Test Commission (1999). *International guidelines for test use.* Retrieved October 31, 2012 from http://www.intestcom.org/upload/sitefiles/41.pdf.

Jackson, S. A., Martin, A. J., & Elkund, R. C. (2008). Long and short measures of flow: The construct validity of the FSS-2, DFS-2, and new brief counterparts. *Journal of Sport & Exercise Psychology, 30,* 561-587.

Joint Committee for Standards on Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs*. Newbury Park, CA: Sage.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NU: Guilford.

Kobasa, S. C. (1979). Stressful life events, personality, and health: An inquiry into hardiness. *Journal of Personality & Social Psychology, 37,* 1–11. doi: 10.1037/0022-3514.37.1.1

Krosnick, J. (1999). Survey research. *Annual Review of Psychology, 50,* 537–567. doi: 10.1146/annurev.psych.50.1.537

Lafrenière, M-A, K., Verner-Filion, J., & Vallerand, R. J. (2012). Development and validation of the Gaming Motivation Scale (GAMS). *Personality and Individual Differences, 53,* 827-831.

Liem, G. A., & Martin, A. J. (2012). The Motivation and Engagement Scale: Theoretical framework, psychometric properties, and applied yields. *Australian Psychologist, 47,* 3–13. Doi: 10.1111/j.1742-9544.2011.00049.x

Lonsdale, C., Hodge, K., & Rose, E. A. (2006). Pixels vs. paper: Comparing online and traditional survey methods in sport psychology. *Journal of Sport & Exercise Psychology, 28,* 100–108.

Lucas, T., Young, J. D., Zhdanova, L., & Alexander, S. (2010). Self and other justice beliefs, impulsivity, rumination, and forgiveness: Justice beliefs can both prevent and promote forgiveness. *Personality and Individual Differences, 49,* 851–856. doi: 10.1016/j.paid.2010.07.014

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

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

Marsh, H. W., Jayasinghe, U. W., & Bond, N. W. (2011). Gender differences in peer reviews of grant applications: A substantive-methodological synergy in support of the null hypothesis model. *Journal of Informetrics, 5,* 167-180. doi: 10.1016/j.joi.2010.10.004

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big-five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22,* 471-491. doi: 10.1037/a0019227

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

Marsh, H. W., Muthén, B., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16,* 439–476. doi: 10.1080/10705510903008220

McCrory, C., & Layte, R. (2012). Testing competing models of the Strengths and Difficulties Questionnaire's (SDQ's) factor structure for the parent-informant instrument. *Personality and Individual Differences, 52,* 882-887. doi:10.1016/j.paid.2012.02.011

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. doi:10.1016/j.paid.2007.11.011

McCoy, S., Marks, P., Carr, C., & Mbarika, V. (2004). Electronic versus paper surveys: Analysis of potential psychometric biases. In *Proceedings of the 37th Hawaii International Conference on System Sciences*, 5-8th January, 2004. Big Island, HI, USA. IEEE Computer Society.

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. doi: 10.1016/j.paid.2012.11.020

Piccolo, R. F., & Colquitt, J. A. (2006). Transformational leadership and job behaviors: The mediating role of core job characteristics. *Academy of Management Journal, 49,* 327–340.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86,* 638–641. doi:10.1037/0033-2909.86.3.638

Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis. *Personality and Social Psychology Bulletin, 28,* 1629–1646.

Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment, 29,* 304–321. doi: 10.1177/0734282911406653

Spence, J. C., & Blanchard, C. M. (2001). Publication bias in sport and exercise psychology research: The games we play. *International Journal of Sport Psychology, 32*, 386-399.

Sternberg, R. J., & Sternberg, K. (2010). *The psychologist's companion: A guide to writing scientific papers for students and researchers* (5th ed.). New York, NY: Cambridge University Press.

Tortora, R. (2009). Attrition in consumer panels. In P. Lynn (Ed.), *Methodology of longitudinal surveys* (pp. 235–249). Hoboken, NJ: Wiley.

Williams, L. J. (1995). Covariance structure modeling in organizational research: Problems with the method versus applications of the method. *Journal of Organizational Behavior, 16*, 225–234.

Williams, S. E., & Cumming, J. (2012). Sport imagery ability predicts trait confidence, and challenge and threat appraisal tendencies. *European Journal of Sport Science, 12,* 499-508. doi: 10.1080/17461391.2011.630102

Yeo, G. B., & Frederiks, E. R. (2011). Cognitive and affective regulation: Scale validation and nomological network analysis. *Applied Psychology – An International Review, 60,* 546-575. doi: 10.1111/j.1464-0597.2011.00447.x

Table 1. *Summary of fit indices for the hypothesized correlated, four-factor MTQ48 model examined with the Australian and British athletes*.

Models	χ^2	*df*	*p*	CFI	TLI	SRMR	RMSEA (90% CI)
Australian athletes (*n* = 352)							
CFA	3823.59#	1074	.000	.487	.461	.085	.085 (.082 to .088)
ESEM	2922.85	942	.000	.630	.557	.059	.077 (.074 to .080)
British athletes (*n* = 334)							
CFA	2124.34	1074	.000	.594	.574	.079	.054 (.051 to .057)
ESEM	1385.46	942	.000	.829	.795	.047	.038 (.033 to .042)

Note: χ^2 = chi square; df = degrees of freedom; CFI = comparative fit index; IFI = incremental fit index; TLI = Tucker-Lewis index; SRMR = standardized root mean residual; RMSEA = root mean square error of approximation; # denotes not positive definite covariance matrix; residual error terms were not correlated.

Table 2. *Standardized parameter estimates for the CFA and ESEM of the MTQ48 with the Australian (n = 352) and British athletes (n = 334).*

	Factor 1 (Challenge)		Factor 2 (Commitment)		Factor 3 (Control)		Factor 4 (Confidence)	
	ESEM	CFA	ESEM	CFA	ESEM	CFA	ESEM	CFA
Mt4	**.49/.41**	**.60/.51**	.11/.10		**.26/**.08		.02/.01	
Mt6	.24*/.11	**.35/.23**	**.47/**.05		**-.31/-**.01		.02/**.25**	
Mt14	.08/.14	**.31/.27**	**.29/**.16		-.16*/-.09		.17*/**.31**	
Mt23	**.42/.29**	**.65/.49**	.12/.22*		.03/.15		**.29/**.03	
Mt30	**.55/.37**	**.53/.36**	-.03/-.14		.08/**.25**		.06/-.12	
Mt40	.22/**.27**	**.38/.33**	.12/-.03		.09/.04		-.02/.14	
Mt44	**.51/.46**	**.62/.66**	.08/.20*		.21*/.17*		.04/-.06	
Mt48	**.56/.46**	**.57/.58**	-.08/.13		-.03/.15		**.23/-**.10	
Mt1	-.01/**.39**		-.12/.20	**.31/.45**	.09/.08		**.52/-**.13*	
Mt7	**.41/.36**		.08/**.25**	**.42/.49**	.14/.07		-.06/-.08	
Mt11	-.06/-.12		**.66/.48**	**.51/.39**	.12/.07		-.02/**.23**	
Mt19	.21/.17		.10/.21*	**.42/.35**	.15/.19*		.06/-.20*	
Mt22	.18/-.06		**.51/.46**	**.53/.33**	-.05/.01		-.03/-.13	
Mt25	.07/**.30**		**.36/.39**	**.61/.46**	**.33/-**.03		.10/**-.27**	
Mt29	.21/**.33**		.28/**.30**	**.58/.53**	.03/.01		.23/.00	
Mt35	.19/-.12		**.34/.56**	**.39/.38**	.11/.01		-.15/.11	
Mt39	.24/.12		.28*/**.56**	**.49/.56**	.29*/.12		-.12/-.16*	
Mt42	.15/.15		**.52/.40**	**.67/.53**	.06/.13		.15*/.03	
Mt47	.04/.19		**.34/**.25*	**.46/.40**	.16*/.07		.10/.15	
Mt2	.03/**.35**		.05/.04		-.04/.07	**.59/.26**	**.72/-**.02	
Mt5	.16/-.03		.01/.04		**.35/.57**	**.37/.26**	.10/-.16*	
Mt9	.01/-.07		**.32/.26**		-.02/.00	**.22/**.16*	.03/.10	
Mt12	.13/**.41**		.03/.00		.14/.09	**.68/.41**	**.64/**.13	
Mt15	.04/-.07		**.54/**.26*		-.12/.10	**.38/.41**	.10/**.37**	
Mt33	-.04/.00		.10/.13		-.21*/.00	.16*/**.34**	**.27/.36**	
Mt41	.06/.13		**.46/**.11		.13/**.24**	**.64/.55**	**.31/.34**	
Mt21	.10/.12		.19/-.02		.03/-.12	**.23/.39**	.02/**.58**	
Mt26	.15*/.04		.21*/.07		**-.67/-.27**	-.12/-.13	-.01/-.02	
Mt27	.31*/.01		.16/-.01		-.23/.06	**.32/.48**	.15/**.63**	
Mt31	**.67/.40**		-.06/.00		-.04/.06	**.36/.42**	.01/**.21**	
Mt34	**.34/**.08		-.12/.01		**-.50/-**.17	**-.21/-.26**	-.11/**-.28**	
Mt37	**.29/-**.09		**.35/.61**		.05/-.04	**.31/.25**	-.14/.07	
Mt45	**.39/**.20*		-.03/.02		.11/.17*	**.52/.46**	**.30/.30**	
Mt3	-.07/**.38**		.09/.05		.16*/.15		**.70/**.16*	**.67/.49**
Mt8	**.17/.27**		.08/.09		**.37/.25**		**.39/.25**	**.66/.58**
Mt10	-.14/-.17		**.38/**.21*		-.04/.14*		**.21/.32**	**.32/.32**
Mt13	.11/**.58**		-.06/-.10		-.02/-.02		**.64/**.11	**.53/.36**
Mt16	.20*/**.48**		.01/-.01		.07/.10		**.51/.23**	**.57/.51**
Mt18	.09/.02		**.41/.29**		.03/.04		**.33/.46**	**.61/.51**
Mt24	-.08/.01		.10/-.04		.05/-.02		-.01/**.37**	.07/**.24**
Mt32	-.14/.13		**.45/**.09		.10/.00		**.32/.46**	**.52/.43**
Mt36	.23/.00		.36*/.15*		-.03/.08		.07/**.54**	**.43/.49**
Mt17	.17*/-.03		-.02/.02		**.58/.53**		-.04/.00	**.35/.33**
Mt20	.18/.09		.05/.05		**.33/.55**		.14/-.12	**.42/.34**
Mt28	-.09/.01		**.30/**.09		.18/**.30**		.08/**.27**	**.32/.43**
Mt38	.02/.05		-.01/-.02		**.46/.66**		**.21/.04**	**.41/.46**
Mt43	.17*/-.04		-.07/-.01		**.43/.63**		-.08/.00	**.21/.38**
Mt46	.16/-.11		**.24/**.07		**.29/.46**		.05/.12	**.44/.34**

Note: Statistically significant parameter estimates at $p < .05$ are marked with an * and $p < .01$ are bolded; parameter estimates are for Australian/British athletes, respectively.