Exploring Gender Equivalence and Bias in a Measure of Psychological Hardiness

Sigurd W. Hystad[1]

[1] Department of Psychosocial Science, University of Bergen, Bergen, Norway

Correspondence: Sigurd W. Hystad, Department of Psychosocial Science, University of Bergen, P.O. Box 7807, 5020 Bergen, Norway. Tel: 47-5-553-289. E-mail: sigurd.hystad@psysp.uib.no

Abstract

One of the most pervasive criticisms found in the hardiness literature concern the question whether the construct is equally important for men and women. Using a multi-group confirmatory factor analytic approach, this question was explored from a more fundamental perspective by examining measurement equivalence across gender in a measure of hardiness, the 15-item Dispositional Resilience Scale [DRS-15; Bartone, P. T. (1995). *A short hardiness scale.* Paper presented at the Annual Convention of the American Psychological Society, New York.]. Although results suggested some non-equivalence related to the control subscale, follow-up analyses examining gender bias in the two non-equivalent items showed that the effect of gender was minimal. The gender effects found indicated that women had a greater tendency to endorse these items compared to men. Given the stringent criteria used to test for equivalence and minimal evidence of bias found, it is concluded that the results largely point to equivalence across gender in the DRS-15.

Keywords: Psychological hardiness, gender differences, measurement equivalence, multi-group confirmatory factor analysis, item bias

1. Introduction

The personality disposition known as hardiness describes a generalized style of functioning characterized by a strong sense of commitment (the ability to see the world as interesting and meaningful), control (the belief in one's own ability to influence the course of events), and challenge (seeing new experiences as exciting opportunities for personal growth; Bartone, 2000). It is conceptualized as a constellation of personality characteristics that function as a resilience resource in the encounter with stressful life events (Kobasa, 1979; Kobasa, Maddi, & Kahn, 1982; Kobasa, Maddi, & Zola, 1983). What started as a longitudinal study of American business executive (c.f., Maddi & Kobasa, 1984) has by now grown into an impressive body of research demonstrating the stress mitigating effects of hardiness (see, e.g., Bernas & Major, 2000; Hystad, Eid, Laberg, Johnsen, & Bartone, 2009; Soderstrom, Dolbier, Leiferman, & Steinhardt, 2000).

The hardiness-stress mechanism most likely involves a combination of cognitive, physiological, and behavioral processes. For example, Maddi and Hightower (1999) have suggested that hardiness encourages a kind of coping process that renders stressful events less harmful, labeled transformational coping. That is, high-hardy individuals are more likely to react to stressful events with increased interaction, effort, and active attempts to find solutions. Part of this transformational coping involves the interpretation or meaning that people attach to events around them (Ouellette, 1993). People characterized by high levels of hardiness believe they can control or influence events, tend to interpret stressful events in positive and constructive ways, and construe these events as challenges and valuable learning opportunities. These adaptive cognitions are in turn believed to result in lower levels of organismic strain in response to potentially threatening events (Kobasa, Maddi, Puccetti, & Zola, 1985). For instance, Dolbier and colleagues (Dolbier et al., 2001) have suggested that these adaptive appraisals protects individuals from the immune-suppressive effects of stress, and thereby enabling them to uphold a healthy status.

1.1 Gender Differences in Hardiness Research

The accumulating evidence aside, the construct of hardiness has not escaped criticism. One of the most pervasive lines of critique concerns the question whether hardiness is equally important for men and women. From a

social-constructive and feminist standpoint, Riska (2002) has criticized the construct as merely a way to confirm and legitimize traditional masculinity. In Riska's view, hardiness is a product of the socio-political climate of its decade (the 1980s), reflecting the traditional white, middle-class, and masculine values prevalent in the US at the time. She describes a transition from the Type A man that captured the ideological construct of traditional masculinity prevalent in the 1950s, to the hardy man, which served to demedicalize and give new legitimacy to traditional male behavior. In this transition, hardiness became a key to re-evaluate the core features of traditional masculinity; men could again be ambitious, competitive, and in control, while remaining healthy, as opposed to the unhealthy and coronary-prone Type A man.

Many of the concerns regarding gender differences found in the literature stem from the fact that the original hardiness measure was developed based on a sample of exclusively male executives (c.f. Maddi & Kobasa, 1984), and that early empirical investigations tended to focus primarily on men. In later studies that have included female participants, inconsistent or equivocal results have been reported. Some have found that hardiness moderates the ill effects of stress on health for men, but not for women (Benishek & Lopez, 1997; Klag & Bradley, 2004; Shepperd & Kashani, 1991), while others have found similar effects for the two sexes (King, King, Fairbank, Keane, & Adams, 1998; Robitschek & Kashubeck, 1999; Rosen, Wright, Marlowe, Bartone, & Gifford, 1999).

Several explanations have been forwarded to explain the gender differences. It has been suggested that the coping strategies typically employed by men and women could account for some of the observed differences (Klag & Bradley, 2004; Williams, Wiebe, & Smith, 1992). More precisely, men stereotypically cope with stress by using problem-focused strategies, whereas women are believed to employ avoidance strategies to a larger degree (Tamres, Janicki, & Helgeson, 2002). Moreover, men and women are often believed to differ in regard to how a particular stressor is appraised and in what they consider stressful (Baum & Grunberg, 1991). Combined, the argument is that, compared to hardy men, equally hardy women use less beneficial cognitive and behavioral coping strategies.

However, this explanation seems insufficient on several grounds. Firstly, there are reasons to believe that the stereotypical coping patterns of men and women are not as clear cut as presented above (Tamres et al., 2002). And secondly, research have demonstrated gender differences even when no differences in coping were present (Klag & Bradley, 2004).

A more plausible explanation could be that the stressors that have been investigated in many hardiness studies are predominantly male oriented (Wiebe & Williams, 1992). Several studies have focused on achievement-oriented stressors, whereas social or interpersonal stressors have been researched less often. Many studies may thus contain a masculine bias in the gender relevance of stressors. To illustrate, Wiebe (1991) subjected male and female participants to an experimentally induced evaluative threat (achievement-oriented), and recorded the participants' affective and physiological response. The few gender differences Wiebe found showed that hardiness was a protective factor for males, but not females, and this could be related to the achievement-oriented stressor used.

Yet studies that have focused on female participants and included interpersonal or social stressors have managed to demonstrate beneficial effects of hardiness. For example, Feinauer, Mitchell, Harper, and Dane (1996) found that among victims of childhood sexual abuse, high-hardy women had significantly fewer distressing symptoms (e.g., depression, sleep disturbance, sexual dysfunctions, and symptoms of post-traumatic stress disorder). Likewise, Foster and Dion (2003) demonstrated that women who scored high on hardiness reported less anxiety after being exposed to both hypothetical and personal encounters with gender discrimination in a laboratory.

1.2 The Issue of Measurement Equivalence

While the above arguments all raise important and valid points, they fail to address a more fundamental question; are we actually measuring the same underlying construct in men and women with current measures of hardiness? Great strides have been made within the field of cross-cultural psychology in stressing the importance of instrument equivalence in cross-group comparisons. It is now generally acknowledged that measurement equivalence needs to be established before any meaningful comparisons between different groups can be made (Byrne & Watkins, 2003; van de Vijver & Leung, 1997). However, measurement equivalence, and the closely related term measurement bias, has not featured as prominently in hardiness research.

Equivalence is a generic term describing different aspect concerning the comparability of a construct or measuring instrument across two or more groups. For instance, *measurement* equivalence refers to whether the association between the construct or constructs being measured and the indicators used to measure it are the same across groups (Byrne & Watkins, 2003). In other words, measurement equivalence concern the degree to

which the content of the items comprising the scale are being perceived in the same way across groups. Similarly, *structural* equivalence refers to whether relations among the constructs, as tapped by its subscales, are the same across groups (Note 1). Structural equivalence thus concerns the underlying theoretical or factorial structure of a given measuring instrument.

In a similar vein, the term bias can be thought of as a generic term reflecting all nuisance factors that can possibly threaten multi-group comparability. For instance, the term *construct* bias suggests the notion that the construct being measured is understood differently across the groups being studied, for example because its indicators are differentially appropriate across groups. The term *item* bias (also commonly referred to as differential item functioning or DIF) implies that a particular item elicits a differential meaning of their content across groups. The implication of item bias is that the item does not discriminate between people with different standings on the trait in the same way across groups. In other words, people with the same standing on the underlying trait should have the same item score irrespective of group membership (van de Vijver & Leung, 1997).

As an extension of the critique forwarded by Riska (2002), one could therefore argue that hardiness measures contain (subtle) gender biases and do not adequately capture or measure the construct in women. For example, the behavior of working hard that is often used as a marker of hardiness may not be an appropriate indicator among women. In Riska's terms, this indicator would reflect traditional masculine values and behaviors. As of yet, no study has systematically examined potential gender bias at the measurement level. This is somewhat surprising since measurement equivalence is a prerequisite if any substantial claims about gender differences are to be made. Before we can establish whether differences exist in the hardiness-health relationship, we need to ascertain that the same underlying construct is being measured in men and women.

1.3 Research Aim

The aim of the current study was to fill this gap in the literature by examining a commonly used hardiness measure for equivalence across samples of male and female participants. To achieve this aim, a confirmatory factor analytic (CFA) approach was used, wherein a baseline model was compared to successively more restricted models. Next, provided with findings of non-equivalent items, an analysis of variance (ANOVA) procedure suggested by van de Vijver and Leung (1997) was used to further explore the presence of bias related to the items in question.

2. Method

2.1 Participants

In order to obtain a sufficiently large and diverse study sample, participants from four pre-existent surveys were pooled into one sample. Participants in these surveys had all completed the same hardiness measure. The first sample consisted of employees from the Norwegian Armed Forces who had completed hardiness questionnaires as part of an annual personnel survey in 2008 (n = 7522, 1265 females). The second sample included undergraduate students enrolled in introductory psychology courses at the University of Bergen who had completed the hardiness scale as part of a larger study in 2007 (n = 289, 226 females). The third sample comprised participants who had attended leader development programs under the auspices of the Norwegian Armed Forces in the period 1994 to 2007 (n = 157, 88 females). Finally, the fourth sample included candidates applying for admission into different officer candidate schools and military academies in Norway in 2008 (n = 257, 37 females).

Combined, the sample size totaled 8 225 participants (1 616 females; 6 609 males). Of the female participants, 240 (14.9%) were 24 years old or younger, 308 (19.1%) were between the age 25 and 34, 343 (21.2%) were between the age 35 and 44, 710 (43.9%) were between the age 45 and 54, and 15 (0.9%) were older than 54 years. Of the male participants, 262 (4%) were 24 years old or younger, 1 295 (19.6%) were between the age 25 and 34, 1 780 (26.9%) were between the age 35 and 44, 3 256 (49.3%) were between the age 45 and 54, 10 (0.2%) were older than 54 years, and six participants did not report their age. Thus, apart from in the youngest category (24 years or younger), male and female participants were reasonably similar in respect to age distribution.

2.2 The Hardiness Instrument

Hardiness was measured with the 15-item Dispositional Resilience Scale (DRS-15; Bartone, 1995). The DRS-15 consists of five items each to measure the control (e.g., "By working hard you can nearly always achieve your goals"), commitment (e.g., "Most of my life gets spent doing things that are meaningful"), and challenge (e.g., "Changes in routines are interesting to me") dimensions of hardiness. After reversing six negatively keyed items,

a total hardiness score can be obtained by summing responses to all items. In addition to a total hardiness score, three subscale scores can be created by summing the relevant five items for each of the commitment, control, and challenge dimensions. All items are scored on a four-point scale, ranging from *not at all true* to *completely true*.

Previous research has confirmed the DRS-15 to be a valid and useful tool in both military and non-military samples (e.g., Britt, Adler, & Bartone, 2001; Clark, 2002; Vogt, Rizvi, Shipherd, & Resick, 2008). In the present study, an adapted Norwegian version of the DRS-15 was used. This scale has previously been validated for use in the Norwegian population and language (Hystad, Eid, Johnsen, Laberg, & Bartone, 2010). In two recent studies, this scale predicted the likelihood of sickness absence from work (Hystad, Eid, & Brevik, 2011) and was negatively related to risk for alcohol abuse in military personnel (Bartone, Hystad, Eid, & Brevik, 2012).

Cronbach's alphas for the total score in this study were .78 and .76 for men and women, respectively. For men, the Cronbach's alphas for the subscales were .74, 75, and .62 for commitment, control, and challenge, respectively, while for women, the alphas were .73, .73, and .67 for commitment, control, and challenge, respectively. These reliability coefficients are within the range typically reported for the 15-item scale and subscales (usually in the .60–.70 range; e.g., Bartone, Roland, Picano, & Williams, 2008; Britt et al., 2001).

2.3 Statistical Analyses

Hystad et al. (2010) have demonstrated that the DRS can best be represented as a hierarchical structure comprising a general hardiness dimension and three first-order factors corresponding to the sub-dimensions of commitment, control, and challenge. To test the equivalence of this theoretical structure across male and female participants, version 6 of the statistical program EQS was used (Bentler, 2001). Based upon previous suggestions and empirical investigations (Byrne, Shavelson, & Muthén, 1989; Vandenberg & Lance, 2000), the following steps were taken in the present study:

1) A test of configural equivalence, wherein equal factor structures are tested. This is achieved by specifying the same pattern of fixed and free factor loadings in each group, and aims at examining whether the hardiness instrument evoke the same cognitive frame of reference for female and male respondents. This also serves as a baseline model to which subsequent, more restricted models can be compared.

2) A test of measurement equivalence, in which factor loadings for like items are invariant across groups. Accordingly, step 1 was repeated with imposed equality constraints on the factor loadings for like items. In essence this examines whether the associations between like items and the underlying construct are the same across groups, and thus whether the construct indicators (i.e., the items) are calibrated to the underlying construct in the same manner.

3) A test of structural equivalence, in which the underlying theoretical structure of the instrument is tested for equivalence. This entails that the relations among the construct, as tapped by the hardiness measure's subscales, are equivalent across groups. In this model all specified constraints from the previous step are retained while concurrently testing for equivalence of the relations between the latent factors.

In each of these steps, all error variances were allowed to vary freely between the two groups. The requirement that error variances be equal between groups is considered excessively stringent and of little practical value, and thus test of equivalence generally do not include constraints between errors (Byrne & Watkins, 2003).

Provided with evidence of non-equivalent items, the next step included scrutinizing further each item in question by using the ANOVA approach described by van de Vijver and Leung (1997). In this approach, item score is treated as the dependent variable, while gender and score levels are the independent variables. Score levels are composed on the basis of total score on the instrument or its subscales, and ideally, all possible score levels should be scrutinized. Most often, however, it is impossible to separate all score levels due to insufficient data in many levels. Based on van de Vijver and Leung's recommendation of at least 50 persons per score level, nine different levels were created. Next, 2 x 9 two-way ANOVAs, with gender (2 levels) and score level (9 levels) as independent variables, and item score as the dependent variables, were performed.

3. Results

3.1 Baseline Models

Testing for equivalence begins with separate estimations of well-fitting baseline models for each group. Following completion of this task, the multi-group model, comprising the baseline models for both men and women, is tested for equivalence across groups.

Given evidence of some kurtosis related to some DRS items, all analyses were based on the robust estimation

method available in EQS that corrects for non-normality. This method provides a robust Satorra-Bentler chi square (S-Bχ^2; Satorra & Bentler, 2001), accompanied by robust versions of the root mean square error of approximation (*RMSEA) and Comparative Fit Index (*CFI). A value of .90 for the *CFI is commonly taken as the rule-of-thumb lower limit cut-point of acceptable fit, whereas *RMSEA values less or equal to .08 represent reasonable fit (Byrne & Watkins, 2003). In addition to these robust fit statistics, we inspected Jöreskog and Sörbom's (1996) Goodness-of-Fit Index (GFI) and the standardized root mean residual (SRMR). Generally, a GFI ≥ .90 and SRMR ≤ .08 indicate good model fit (Hu & Bentler, 1999; Kline, 1998). Goodness-of-fit statistics related to the baseline models revealed an adequately, albeit marginally, well-fitting model for both women (S-Bχ^2 (86) = 665.87; *CFI = .88; GFI = .94; *RMSEA = .066; SRMR = .071) and men (S-Bχ^2 (86) = 2 399.69; *CFI = .88; GFI = .94; *RMSEA = .064; SRMR = .074).

In testing for the baseline models for both groups, inspection of the Lagrange Multiplier (LM) test provided by EQS argued for the specifications of error covariance between two item pairs for both groups (DRS6 and DRS8; DRS16 and DRS18). Error correlation between item pairs can be justified because it often indicates perceived redundancy in item content or represents non-random error due to method effects (Byrne, Baron, & Campell, 1993). On these grounds, we considered it theoretically justified to include error correlations between said items because they are negatively keyed (DRS6 and DRS8) and/or belong to the same subscale (DRS6 and DRS8; DRS16 and DRS18). The multi-group model to be tested for equivalence therefore contains two common item pair covariances. This model is illustrated in Figure 1.
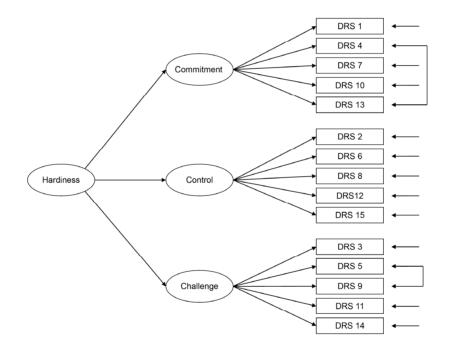

Figure 1. Hypothesized equivalent multi-group model of the dispositional resilience scale

3.2 Test for Equivalence

Nested models (i.e., models that are hierarchically related to one another in that their parameter sets are subsets of one another) are usually compared by computing the difference in chi square values ($\Delta\chi^2$) for the two models. This $\Delta\chi^2$ value is distributed as χ^2, with degrees of freedom equal to the difference in degrees of freedom (Δdf), and non-significant values indicate equivalence between models. When the analyses are based on robust estimation, Satorra and Benter (2001) have shown that the difference in S-Bχ^2 (ΔS-Bχ^2) can be corrected and used in the same way as the $\Delta\chi^2$ value to compare models. Results from the model that allowed all parameters to be freely estimated across gender, and the model that constrained factor loadings and the common error covariance to equality, yielded a ΔS-Bχ^2(15) of 39.45, with $p < 0.001$.

Table 1. Goodness-of-fit statistics and summary of equivalence testing of the dispositional resilience scale across gender

Model	S-Bχ²	df	*CFI	SRMR	*RMSEA	*RMSEA C.I.
1. No constraints imposed	3 097.39	171	.884	.072	.065	.063-.067
2. Factor loadings and common error covariance[b] constrained	3 136.90	186	.883	.073	.062	.060-.064
3. Factor loadings, common error covariance[c], and second-order loadings constrained	3 128.08	186	.883	.074	.062	.060-.064
Model Comparison	ΔS-Bχ²[a]	Δdf	Δ*CFI			
1 vs. 2	39.45***	15	.001			
1 vs. 3	28.11*	15	.001			

Note. S-Bχ² = Satorra-Bentler scaled chi square; GFI = Goodness-of-Fit Index; *CFI = robust version of the Comparative Fit Index; *RMSEA = robust version of the root mean square error of approximation; C.I. = confidence interval; SRMR = standardized root mean residual.

[a] Corrected ΔS-Bχ² values are reported. [b] Error covariance between item pairs DRS6-DRS8 and DRS16-DRS18. [c] Error covariance between DRS16 and DRS18.
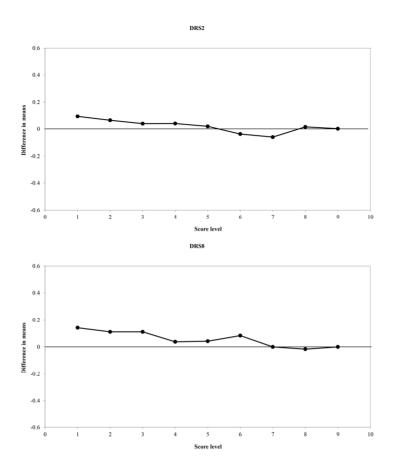
* $p < .05$. *** $p < .001$.

3.3 Tests for Item Bias

The LM test χ^2 statistics assigned to each constrained parameter indicated that the non-equivalence related to one error covariance and two factor loadings (DRS2: "By working hard you can nearly always achieve your goals" and DRS8: "I don't think there's much I can do to influence my own future"), both belonging to the control subscale. Faced with these results, the relations between the latent factors were next tested for (structural) equality, while allowing the error covariance and factor loadings associated with DRS2 and DRS8 to vary freely among groups, as suggested by Byrne et al. (1989). The results from this test of partial equivalence indicated non-equivalence related to the second-order loading from control to the general hardiness factor (LM test χ^2 = 10.67, p = .001). Results from the testing of equivalence are summarized in Table 1.

The results from the test of structural equivalence echoed the finding from the analysis related to the first-order factor loadings, suggesting that the control dimension and related items might not function equivalently across gender. To further explore the potential non-equivalence of the two control items, tests for evidence of item bias were performed following the procedures proposed by van de Vijver and Leung (1997), as described in section 2.3. According to van de Vijver and Leung, significant differences related to the main effect of gender points to uniform bias (i.e., bias that is constant across score levels), while a significant interaction between gender and score level indicates non-uniform bias (i.e., bias that is not constant across score levels). Given the relatively large sample size used in the present study, significant effects are likely to emerge for trivial differences between the genders. Consequently, the extent of bias was evaluated based on inspection of effect sizes (η_p^2) for the main and interaction effects, where values of .01, .06, and .14 are considered small, medium, and large, respectively (Cohen, 1988).

Results from the ANOVA showed that item DRS8 demonstrated a significant effect of gender, p = .004, η_p^2 = .001. None of the interaction effects between gender and score level, or the main effect of gender for item DRS2, was significant. In reviewing the effect size for DRS8, however, the extent of bias could be characterized as negligible. In other words, the effect of gender accounted for only 1‰ of variance in item score.

Figure 2 gives a more illustrative picture of the two non-equivalent items. The horizontal axes of this figure represents the different score levels computed for the control subscale, while the vertical axes represents the difference in mean value resulting from subtraction of item mean score for female participants from the item mean score for male participants. A biased item is expected to exhibit a pattern in which the plotted score level points depart from zero in a systematic pattern. For example, if the plot of points remains consistently above or below zero, the item is said to be uniformly biased towards one of the groups, depending on whether the plot is above or below.

Figure 2. Non-equivalent items related to the control subscale of the dispositional resilience scale

Positive scores on the vertical axes indicate higher mean scores for women and negative scores on the vertical axes indicate higher mean scores for men

Turning to figure 2, it is evident that although the plotted points for DRS8 remain somewhat consistently above zero, this pattern is not strong. The line connecting the mean difference at each score level is close to zero (representing equal item mean scores), and is near to tangent to the zero-line at the higher score levels. A final point worth mentioning is that, to the extent that the item is biased, it favors female participants. That is, at every score level except level 8, female had mean item scores larger than or equal to male participants. Also evident in Figure 2, the plotted points for the other non-equivalent item (DRS2), exhibit a more inconsistent pattern, interchangeably positioned above and below zero, as expected from the non-significant ANOVA results.

4. Discussion

This article aimed to fill a gap in the literature by examining equivalence in a commonly used hardiness scale. It was argued that before we have established whether hardiness scales actually measure the same construct across gender, it is impossible to draw any decisive inferences about any differences found in the literature. The results from the analyses of equivalence are indicative of some gender differences. Adequately well-fitting baseline models were established for both genders, suggesting that the DRS evoked the same cognitive frame of reference for both female and male participants. However, two items belonging to the control subscale were found not to be equivalent across gender. This entails that the associations between these items and the underlying construct of control were not equivalent for female and male participants. Echoing this result, the test for structural equivalence showed that the relationship between the control subscale and the general hardiness factor was non-equivalent across gender. Examining the non-equivalent items in more detail, however, revealed that the amount of bias was small at best.

It is interesting to note that the difference between male and female participants all appeared in the control subscale. The control dimension of hardiness involves the perception of your ability to affect the course of events,

and is assessed by statements involving hard work and individual effort to achieve goals and affect your surroundings (e.g., "How things go in my life depends on my own actions). Following Riska (2002), these indicators could be said to reflect traditional masculine values. There is also empirical evidence to suggest that personal control holds different importance to the identity of men and women. For example, it is frequently found that those with traditionally masculine characteristics are generally more affected by success in instrumental activities, as opposed to those with traditionally female characteristics who are generally more affected by interpersonal success (Waelde, Silvern, & Hodges, 1994). Moreover, in a study of adolescents, Margalit and Eysenck (1990) found that boys' development of identity focused on individual achievement and task-oriented behavior, while girls' development focused on issues of relationships and social behavior.

Yet a male emphasis on personal control does not seem to explain the gender differences found in this study. As the ANOVA analyses showed, the modest signs of bias found in the non-equivalent control items favored the female participants. In other words, given the same score on the underlying control factor, female participants nevertheless had a higher mean item score than male participants. In practical terms, this means that women had a higher propensity to endorse these items compared to men. Also, while not included in the results section, but implied from the results from the ANOVA analyses, women had significantly higher mean scores on the control subscale ($t(8223) = 4.03$, $p < .001$). A likely explanation for these results pertains to the particular samples used. With one exception, the samples were drawn from military populations, or, in one case, candidates applying for officer candidate schools. Perhaps the women in these populations have internalized traditional male values in order to succeed in predominantly male dominated domains, and to such a degree that these values eventually exceeded those of their male counterparts.

It should also be noted that the analyses conducted in this article represents a conservative test of equivalence. Specifically, the χ^2 is sensitive to sample size. Because of this, the $\Delta\chi^2$ (and $\Delta S\text{-}B\chi^2$) value is also sensitive to sample size and tends to yield significant values even for trivial differences between groups (Cheung & Rensvold, 2002). For this reason, there is an increasing tendency to rely on two other criteria when evaluating equivalence (Byrne, 2006). Firstly, the more restricted or constrained model is deemed equivalent if it exhibits an adequate fit to the data. Based on this criterion, equivalence across gender could perhaps be said to have been established in the present study. The more restricted models that constrained parameters to equality across groups demonstrated adequate fit on par with the unconstrained baseline model. As Table 1 in section 3.2 shows, the *CFI and SRMR values increased somewhat compared with the baseline model; the *RMSEA value, however, was smaller in both constrained models. The second alternative criterion involves inspecting change in fit statistics other than the χ^2. Cheung and Rensvold (2002) examined 20 different goodness-of-fit statistics and concluded that the ΔCFI was a robust statistic relatively unaffected by sample size. They also suggested that the ΔCFI value should not exceed .01. Again, based on this criterion, equivalence across gender is supported in the present study, as evident in the negligible ΔCFI value of .001 (see Table 1).

In conclusion, the results from the present study support some gender difference at the measurement level of psychological hardiness. As previously argued, earlier studies have maybe somewhat prematurely jumped to conclusions and argued that hardiness is more important for men, without establishing whether the same construct is actually measured in both genders. The present study suggests that there might in fact be some gender differences in hardiness, relating to how the associations between the indicators and the underlying construct of control is perceived; and as an extension to this, how the association between the control factor and the general hardiness dimension is perceived.

However, it is important to note that, while these differences might be of statistical significance, their substantial or practical value might be less certain. Judged by less stringent criteria, for example, the results from the present study largely point to equivalence across gender. In addition, the amount of gender bias according to the ANOVA was negligible, accounting for minimal amount of variance in item scores. At any rate, it behooves the researcher to take the issue of measurement equivalence into account when exploring gender differences in hardiness.

References

Bartone, P. T. (1995). A short hardiness scale. Paper presented at the Annual Convention of the American Psychological Society, New York.

Bartone, P. T. (2000). Hardiness as a resiliency factor for United States Forces in the Gulf War. In J. M. Violanti, D. Paton & C. Dunning (Eds.), *Posttraumatic stress intervention: Challenges, issues and perspectives* (pp. 115-133). Springfield, IL: Charles C Thomas Publisher Ltd.

Bartone, P. T., Hystad, S. W., Eid, J., & Brevik, J. I. (2012). Psychological hardiness and coping style as risk/resilience factors for alcohol abuse. *Military Medicine, 177*, 517-524.

Bartone, P. T., Roland, R. R., Picano, J. J., & Williams, T. J. (2008). Psychological hardiness predicts success in US Army Special Forces candidates. *International Journal of Selection and Assessment, 16*, 78-81. http://dx.doi.org/10.1111/j.1468-2389.2008.00412.x

Baum, A., & Grunberg, N. E. (1991). Gender, stress, and health. *Health Psychology, 10*, 80-85. http://dx.doi.org/10.1037/0278-6133.10.2.80

Benishek, L. A., & Lopez, F. G. (1997). Critical evaluation of hardiness theory: Gender differences, perception of life events, and neuroticism. *Work and Stress, 11*, 33-45. http://dx.doi.org/10.1080/02678379708256820

Bentler, P. M. (2001). *EQS: Structural equations program manual*. Encino, CA: Multivariate Software.

Bernas, K. H., & Major, D. A. (2000). Contributors to stress resistance - Testing a model of women's work-family conflict. *Psychology of Women Quarterly, 24*, 170-178. http://dx.doi.org/10.1111/j.1471-6402.2000.tb00198.x

Britt, T. W., Adler, A. B., & Bartone, P. T. (2001). Deriving benefits from stressful events: The role of engagement in meaningful work and hardiness. *Journal of Occupational Health Psychology, 6*, 53-63. http://dx.doi.org/10.1037/1076-8998.6.1.53

Byrne, B. M. (2006). *Structural equation modeling with EQS* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Byrne, B. M., Baron, P., & Campell, T. L. (1993). Measuring adolescent depression: Factorial validity and invariance of the Beck Depression Inventory across gender. *Journal of Research on Adolescence, 3*, 127-143. http://dx.doi.org/10.1207/s15327795jra0302_2

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456-466. http://dx.doi.org/10.1037/0033-2909.105.3.456

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*, 155-175. http://dx.doi.org/10.1177/0022022102250225

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

Clark, P. C. (2002). Effects of individual and family hardiness on caregiver depression and fatigue. *Research in Nursing & Health, 25*, 37-48. http://dx.doi.org/10.1002/nur.10014

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Dolbier, C. L., Cocke, R. R., Leiferman, J. A., Steinhardt, M. A., Schapiro, S. J., Nehete, P. N., … Sastry, J. (2001). Differences in functional immune responses of high vs. low hardy healthy individuals. *Journal of Behavioral Medicine, 24*, 219-229. http://dx.doi.org/10.1023/A:1010762606006

Feinauer, L. L., Mitchell, J., Harper, J. M., & Dane, S. (1996). The impact of hardiness and severity of childhood abuse on adult adjustment. *American Journal of Family Therapy, 24*, 206-214. http://dx.doi.org/10.1080/01926189608251034

Foster, M. D., & Dion, K. L. (2003). Dispositional hardiness and women's well-being relating to gender discrimination: The role of minimization. *Psychology of Women Quarterly, 27*, 197-208. http://dx.doi.org/10.1111/1471-6402.00099

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research, 18*, 177-144. http://dx.doi.org/10.1080/03610739208253916

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. http://dx.doi.org/10.1080/10705519909540118

Hystad, S. W., Eid, J., & Brevik, J. I. (2011). Effects of psychological hardiness, job demands, and job control on sickness absence: a prospective study. *Journal of Occupational Health Psychology, 16*, 265-278. http://dx.doi.org/ 10.1037/a0022904

Hystad, S. W., Eid, J., Johnsen, B. H., Laberg, J. C., & Bartone, P. (2010). Psychometric properties of the revised Norwegian Dispositional Resilience (hardiness) Scale. *Scandinavian Journal of Psychology, 51,* 237-245. http://dx.doi.org/ 10.1111/j.1467-9450.2009.00759.x

Hystad, S. W., Eid, J., Laberg, J. C., Johnsen, B. H., & Bartone, P. T. (2009). Academic stress and health: Exploring the moderating role of personality hardiness. *Scandinavian Journal of Educational Research, 53*,

421-429. http://dx.doi.org/10.1080/00313830903180349

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: Users's reference guide*. Chicago, IL: Scientific Software International.

King, L. A., King, D. W., Fairbank, J. A., Keane, T. M., & Adams, G. A. (1998). Resilience-recovery factors in post-traumatic stress disorder among female and male Vietnam veterans: Hardiness, postwar social support, and additional stressful life events. *Journal of Personality and Social Psychology, 74*, 420-434. http://dx.doi.org/10.1037/0022-3514.74.2.420

Klag, S., & Bradley, G. (2004). The role of hardiness in stress and illness: An exploration of the effect of negative affectivity and gender. *British Journal of Health Psychology, 9*, 137-161. http://dx.doi.org/10.1348/135910704773891014

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

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

Kobasa, S. C., Maddi, S. R., & Kahn, S. (1982). Hardiness and health - A prospective study. *Journal of Personality and Social Psychology, 42*, 168-177. http://dx.doi.org/10.1037/0022-3514.42.1.168

Kobasa, S. C., Maddi, S. R., Puccetti, M. C., & Zola, M. A. (1985). Effectiveness of hardiness, exercise and social support as resources against illness. *Journal of Psychosomatic Research, 29*, 525-533. http://dx.doi.org/10.1016/0022-3999%2885%2990086-8

Kobasa, S. C., Maddi, S. R., & Zola, M. A. (1983). Type-A and hardiness. *Journal of Behavioral Medicine, 6*, 41-51. http://dx.doi.org/10.1007/BF00845275

Maddi, S. R., & Hightower, M. (1999). Hardiness and optimism as expressed in coping patterns. *Consulting Psychology Journal: Practice and Research, 51*, 95-105. http://dx.doi.org/10.1037/1061-4087.51.2.95

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

Margalit, M., & Eysenck, S. B. G. (1990). Prediction of coherence in adolescence: Gender differences in social skills, personality and family climate. *Journal of Research in Personality, 24*, 510-521. http://dx.doi.org/10.1016/0092-6566(90)90036-6

Ouellette, S. C. (1993). Inquiries into hardiness. In L. Goldberger & S. Breznitz (Eds.), *Handbook of Stress. Theoretical and Clinical Aspects* (2nd ed.). New York: The Free Press.

Riska, E. (2002). From Type A man to the hardy man: masculinity and health. *Sociology of Health & Illness, 24*, http://dx.doi.org/347-358. 10.1111/1467-9566.00298

Robitschek, C., & Kashubeck, S. (1999). A structural model of parental alcoholism, family functioning, and psychological health: The mediating effects of hardiness and personal growth orientation. *Journal of Counseling Psychology, 46*, 159-172. http://dx.doi.org/10.1037/0022-0167.46.2.159

Rosen, L. N., Wright, K., Marlowe, D., Bartone, P., & Gifford, R. K. (1999). Gender differences in subjective distress attributable to anticipation of combat among US Army soldiers deployed to the Persian Gulf during Operation Desert Storm. *Military Medicine, 164*, 753-757.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrica, 66*, http://dx.doi.org/507-514. 10.1007/BF02296192

Shepperd, J. A., & Kashani, J. H. (1991). The relationship of hardiness, gender, and stress to health outcomes in adolescents. *Journal of Personality, 59*, 747-768. http://dx.doi.org/10.1111/j.1467-6494.1991.tb00930.x

Soderstrom, M., Dolbier, C., Leiferman, J., & Steinhardt, M. (2000). The relationship of hardiness, coping strategies, and perceived stress to symptoms of illness. *Journal of Behavioral Medicine, 23*, 311-328. http://dx.doi.org/10.1023/A:1005514310142

Tamres, L. K., Janicki, D., & Helgeson, V. S. (2002). Sex difference in coping behavior: A meta-analytic review and examination of relative coping. *Personality and Social Psychology Review, 6*, 2-30. http://dx.doi.org/10.1207/S15327957PSPR0601_1

van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: SAGE Publications.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-70. http://dx.doi.org/10.1177/109442810031002

Vogt, D. S., Rizvi, S. L., Shipherd, J. C., & Resick, P. A. (2008). Longitudinal investigation of reciprocal relationship between stress reactions and hardiness. *Personality and Social Psychology Bulletin, 34*, 61-73. http://dx.doi.org/ 10.1177/0146167207309197

Waelde, L. C., Silvern, L., & Hodges, W. F. (1994). Stressful life events: Moderators of the relationships of gender and gender roles to self reported depression and suicidality among college students. *Sex Roles, 30*, 1-22. http://dx.doi.org/ 10.1007/BF01420737

Wiebe, D. J. (1991). Hardiness and stress moderation: A test of proposed mechanisms. *Journal of Personality and Social Psychology, 60*, 89-99. http://dx.doi.org/10.1037/0022-3514.60.1.89

Wiebe, D. J., & Williams, P. G. (1992). Hardiness and health – A social psychophysiological perspective on stress and adaptation. *Journal of Social and Clinical Psychology, 11*, 238-262. http://dx.doi.org/10.1521/jscp.1992.11.3.238

Williams, P. G., Wiebe, D. J., & Smith, T. W. (1992). Coping processes as mediators of the relationship between hardiness and health. *Journal of Behavioral Medicine, 15*, 237-255. http://dx.doi.org/10.1007/BF00845354

Note

Note 1. There exist slight terminological differences in the literature when referring to different aspects of equivalence testing. The term *metric* or *measurement unit* equivalence is often used when testing the hypothesis that the factor loadings for like items are equivalent across groups (Horn & McArdle, 1992; van de Vijver & Leung, 1997). The term *structural* equivalence is often used interchangeably with the term *construct* equivalence to refer to a situation where the same construct is measured across groups (often assessed based on comparison of factor solutions; van de Viijver & Leung, 1997). This article adheres to Byrne et al's (1989) distinction, and considers all relationships between measured variables and latent constructs as aspects of *measurement* invariance, and all relationships concerning the latent construct themselves as aspects of *structural* invariance.