Multistudy Report

Development and Validation of an Improved Hardiness Measure

The Hardiness Resilience Gauge

Paul T. Bartone[1], Kelly McDonald[2], Braden J. Hansma[2], Jonathan Stermac-Stein[2], E. M. Romero Escobar[2], Steven J. Stein[2], and Rebecca Ryznar[3]

[1]Institute for National Strategic Studies, National Defense University, Washington, DC, USA
[2]Multi-Health Systems, Inc., Toronto, Canada
[3]Biomedical Sciences Department, Rocky Vista University, Parker, CO, USA

Abstract: Previous research shows that psychological hardiness is an important factor contributing to stress resilience in individuals. Of the various instruments available to measure hardiness, the most commonly used is the Dispositional Resilience Scale (DRS). Despite its demonstrated utility, the DRS-15 still has a number of serious limitations, including low subscale reliability and limited construct validity. The present work aims to create a new hardiness scale that addresses these limitations. A pool of new items plus the original DRS item set was administered to a census-matched stratified sample of N = 2,021 men and women across the United States. Items for the new scale were selected based on item distribution characteristics, item response theory plots, scale reliabilities, item-total correlations, and confirmatory factor analysis (CFA). CFA results showed the best fitting model reflected a hierarchical structure with three factors (commitment, control, and challenge) nested under a broad hardiness factor. This factor structure is replicated in two independent validation samples and also holds invariant across gender and age. The new scale shows much improved reliability coefficients (e.g., Cronbach's α of .93, .85, .84, and .89 for total hardiness, challenge, control, and commitment, respectively), as well as structural equivalence across gender and age. Validity is demonstrated in multiple samples via predictive associations of hardiness scores with theoretically relevant outcome measures, including coping, life satisfaction, anxiety, and depression. The Hardiness Resilience Gauge (HRG) possesses excellent reliability and validity and appears to be a more effective tool for measuring hardiness in adult populations.

Keywords: hardiness, measurement, stress, resilience, Hardiness Resilience Gauge

First identified by Kobasa (1979) in a study of stress and health in Chicago telephone executives, the concept of personality hardiness has proved to be an important factor influencing human resilience. A sizable body of literature now indicates that personality hardiness can enhance resilience, protecting some individuals against the ill effects of stress on health and performance (Bartone, 1989; Contrada, 1989; Eschleman et al., 2010; Kobasa & Puccetti, 1983; Wiebe, 1991). For example, in a study of Gulf War veterans, combat-exposed soldiers who were high in hardiness experienced fewer posttraumatic stress disorder (PTSD) symptoms than those low in hardiness (Bartone, 1999).

Hardiness is generally conceived as an attitudinal style or world view that develops early in life and is reasonably stable over time, though amenable to change under certain conditions (Bartone, 2006; Maddi & Kobasa, 1984). Hardy persons have a strong sense of life and work *commitment,* a greater feeling of *control,* and are more open to change and *challenges* in life. They tend to interpret stressful and painful experiences as a normal aspect of existence, part of life that is overall interesting and worthwhile. While hardiness commitment, control, and challenge show some conceptual similarity to other constructs such as organizational commitment, locus-of-control, and Big Five openness, they are theoretically and empirically quite distinct (Bartone et al., 2009; Maddi, 2002). For example, considerable work has shown that high hardiness is associated with active, problem-solving coping strategies and better mental health and adjustment, especially under stressful conditions (Eschleman et al., 2010), whereas low hardiness is linked to avoidance coping approaches, poor mental health including depression, anxiety and PTSD, and stress vulnerability (Bartone & Homish, 2020; Bartone et al., 2015; Thomassen et al., 2018). Hardiness has also been linked to greater well-being (Lambert et al., 1989) and relationship satisfaction (Nabizadeh & Mahdavi, 2016).

Of the various instruments available to measure hardiness, the most commonly used is the Dispositional Resilience Scale (DRS; Bartone, 1995, 2013; Windle, Bennett & Noyes, 2011). Although this scale has gone through a number of revisions, some notable limitations remain, including low sub-scale reliability. The present study was undertaken

to develop a new hardiness scale that addresses these previous limitations.

Measuring Hardiness

The measurement of hardiness was highly problematic in the early years of work with the concept. It was originally assessed by Kobasa (1979) with an amalgam of 18 different psychological scales that aimed at capturing the dimensions of commitment, control, and challenge. This original collection of over 100 items was later reduced to several shorter versions (Ouellette, 1993), but these still had a number of serious problems. These scales were composed exclusively of negative or "non-hardy" items, opening the door to contamination by neuroticism (e.g., "No matter how hard I try, my efforts will accomplish nothing"; Funk, 1992). Item response distributions were highly skewed, and many were found to be confounded with political attitudes and beliefs (e.g., "Government should guarantee jobs for all"; Bartone, 1984). Also, multiple studies were unable to replicate the theoretical three-factor structure of hardiness, further calling into question the validity of these scales (Funk, 1992).

A shorter and more coherent hardiness test with 50 items was subsequently developed by Bartone (1984, 1989) using samples of bus drivers and telephone company managers. Later, this scale was refined into a 45-item hardiness measure with a balance of positive and negative items and an equal number of items to measure the facets of commitment, control, and challenge (DRS; Bartone et al., 1989). In a critical review of hardiness theory and research, Funk (1992) recommended this scale as the best available hardiness measure at the time. Also, using the DRS, Sinclair and Tetrick (2000) confirmed a factor structure of three dimensions – commitment, control, and challenge – nested under a more general hardiness factor. The DRS was subsequently shortened and improved in various ways, resulting in a 30-item and 15-item version (Bartone, 1991, 1995). The DRS-15 has been used extensively in military and non-military samples, with fairly good results (Andrew et al., 2013; Bartone et al., 1989, 2008; Britt et al., 2001). A final revision of the short DRS-15 sought to improve scale reliabilities and eliminate linguistic bias in the wording of items (Bartone, 2013). The revised DRS-15 shows solid psychometric properties (Hystad et al., 2010) and evidence of predictive validity (Bartone, Valdes, et al., 2016; Johnsen et al., 2013).

Despite its success as a measure of hardiness, the DRS-15 still has several serious limitations. At five items each, the subscales of commitment, control, and challenge often show lower reliability coefficients than desired. For example, a recent study of 570 collegiate athletes found low Cronbach's α coefficients of .58, .67, and .67 for hardiness commitment, control and challenge, respectively, and only .69 for the total scale (Madrigal et al., 2016). This study also found that the theoretical three-factor hardiness model did not show a good fit for the data. Hystad et al. (2010) reported an even lower α coefficient of .62 for the hardiness challenge dimension, with somewhat better coefficients for commitment (.76) and control (.74). Similarly, low α coefficients were found in a sample of Norwegian navy cadets, with a low of .62 for challenge and .73 for commitment (Hystad et al., 2015). In a study of Norwegian undergraduate students, reliability coefficients again were low at .65 for commitment, .72 for control, .71 for the challenge, and .71 for the total hardiness scale (Hystad et al., 2009). Thus, scale reliability is a problem for the DRS-15, especially with regard to the challenge scale.

A further limitation of the DRS-15 is that the short 5-item scales may not fully capture the complexity of the hardiness facets, which is a construct validity weakness. For example, the challenge items for the DRS-15 all focus on the dimension of appreciation for variety and novelty in life and fail to address other conceptually important aspects of challenge. Missing is any attention to the tendency to view changes and disruptions as challenges to overcome and appraising failures as opportunities to learn and grow. The present work seeks to address these shortcomings by creating a new, more reliable, and construct valid hardiness measure with robust psychometric and structural properties across different populations.

Study 1: Scale Development and Initial Evaluation

In the first study, we describe the item development and selection process and the preliminary evaluation of the new scale on a large US-based sample. This group also provides the initial normative data for the new scale.

Materials and Methods

Item Development

A pool of 36 potential items was created that included the original 15 items of the DRS, with 21 new items. In developing new items, two authors (PB and KM) conducted an extensive review of the hardiness empirical and theoretical literature and then independently wrote 12–15 new items. New items aimed to measure the hardiness facets that were not redundant with the original 15 items of the DRS. Professional, academic, and theoretical jargon was avoided to create easily understood items that would be relevant for all adult respondents. Items were also designed to be free of linguistic bias and idiomatic expressions. Following extensive review and discussion, the most theoretically relevant and unique items were retained for further consideration.

This resulted in 21 new items, 8 for challenge, 6 for control, and 7 for commitment. In addition, 7 of the original 15 DRS items were slightly adjusted for clarity. For example, modifying adverbs (e.g., somewhat, really) were removed to make a point of the statement clearer for respondents. The end result of this process was 36 total items retained for more detailed analysis: 13 for challenge, 11 for control, and 12 for commitment.

Normative Sample

Using an online study panel obtained through Amazon Mechanical Turk (Mortensen & Hughes, 2018), a sample of N = 2,016 respondents answered a web-based survey that included the 36 hardiness items, as well as a series of demographic questions and several additional instruments (described below). The sample was representative of the adult population (18 years and older) of the United States and was matched to 2016 US census data for age, gender, race/ethnicity, education, geographic region, and employment status (United States Bureau of the Census, 2016). Participants were recruited from all 50 US states, with specific demographic targets.

To ensure data integrity, responses were screened for patterns of inconsistency or dubious validity (Goldammer et al., 2020). Cases showing evidence of careless responding were dropped from further analysis. Examples included unusually rapid (< 2 min to answer all 36 items) or delayed (> 35 min) completion of the survey. Cases showing identical responses to 13 or more consecutive items were also excluded, as these subjects appeared to be checking boxes without reading the items. Respondents who failed to answer four or more items were also dropped from further analysis. Applying these criteria, a total of 143 cases were thus excluded for spurious or incomplete responses. Of the remaining 1,873 cases, cells were created containing an equal number of cases for different age and gender groups, with proportional representation in geographic region, race, education, and employment groups. Table 1 presents additional details on the normative sample.

Table 1. Demographic variables and their distribution in the normative sample

	Frequency	
Variable	*N*	%
Gender		
Male	750	50.0
Female	750	50.0
Age group (years)		
18–24	300	20.0
25–34	300	20.0
35–54	300	20.0
55–64	300	20.0
65+	300	20.0
Median age (*SD*)	43 (18.7)	
Racial/ethnic group		
Asian	82	5.5
Black	177	11.8
Hispanic	241	16.0
White	957	63.8
Other	43	2.9
Geographic region		
Northeast US	264	17.6
Midwest US	372	24.8
Southern US	528	35.2
Western US	336	22.4
Education level		
No high school diploma	147	9.8
High school graduate	350	23.3
Some college or associate degree	505	33.7
Bachelor's degree	309	20.6
Graduate school	189	12.6
Employment status		
Employed or self-employed: Full-time	589	39.3
Employed or self-employed: Part-time	190	12.7
Retired	382	25.4
Student: Full-time	167	11.1
Student: Part-time	43	2.9
Unemployed	129	8.6
Total	1,500	

Item Selection

From the pool of 36 items, our goal was to identify the most psychometrically sound items, retaining at least eight items for each hardiness facet to assure construct validity and scale reliability. The following criteria guided the selection of the final items:

(1) *Response distribution*: An ideal test item should show good distribution characteristics with all possible response options endorsed and a lack of extreme skewness. Items were selected that displayed distributions in which all possible response options were endorsed to acceptable degrees (extreme responses were less frequent relative to other response options).

(2) *Item-total correlations*: The degree to which each item correlates with other items on its purported subscale was also examined. Selected items displayed moderate to high correlations with the remaining items on their respective subscales, with a minimum r = .30.

(3) *Confirmatory factor analyses (CFAs)*: CFA was applied to assess the degree to which each item contributed to its presumed subscale, as well as to determine the appropriateness of a three-factor hierarchical model. Items were retained that loaded .32 or greater on their intended factor while not cross-loading on other factors (Brown, 2015). An item was determined to cross-load if it: (a) loaded at .32 or greater on more

than one scale, and (b) the difference between loadings on two factors was < .10. Additional CFAs were conducted to identify an overall model with the best fit for the data. Three models were tested: (1) a single general factor; (2) three independent factors; (3) three independent factors nested under a general hardiness factor (hierarchical model).

(4) *Item-response theory (IRT)*: IRT was applied as an additional step in identifying well-performing items to retain. Ideally, good items should (a) yield adequate information about the underlying constructs (challenge, control, commitment) and (b) provide useful information about respondents at all levels of the construct being measured (Zanon et al., 2016). Item characteristic curves and item information curves were plotted and examined for desirable and undesirable properties. Items showing the most positive IRT patterns, such as equally dispersed trace distributions for polytomous responses and high levels of item information across all levels of the latent trait (*theta*), were retained for the final measure.

Norming Procedure

In establishing normative reference groups, the first step was to determine if any age or gender differences were present in the data. If significant differences were obtained for different respondent groups, separate norms would be appropriate. To test this, measurement invariance of the new Hardiness Resilience Gauge (HRG) scale was first assessed using multigroup CFA procedures (Horn et al, 1983; Jöreskog, 1971; Meredith, 1993). Following this, a series of analysis of variance (ANOVA) tests were performed to check for gender and age group mean differences. For measurement invariance, a factor structure consistent with hardiness theory (a general hardiness factor with three subdimensions) and confirmed with our own CFAs, was examined across gender (male and female) and age groups. Here, age was collapsed into five groups to improve interpretability (i.e., 18–24, 25–34, 35–54, 55–64, and 65+ years).

Measurement invariance was tested through a series of hierarchical tests where increasingly restrictive equality constraints were made on parameters across groups. The goal of this process was to determine whether the HRG met the conditions of scalar invariance (i.e., the indicator thresholds and unstandardized factor loadings are statistically equivalent across groups). For the invariance test, Wu and Estabrook's (2016) approach to model identification and measurement invariance testing was used. With this approach, threshold invariance is tested before factor loading invariance (Svetina et al., 2019). The sequence of models tested was: (1) configural model, equating the pattern of factor loadings across groups; (2) metric model, testing for weak invariance, or equating item thresholds across groups; and (3) scalar model, testing for strong invariance, or equating both thresholds and unstandardized factor loadings across groups. Changes in model fit were assessed through Satorra-Bentler scaled chi square difference tests (Satorra & Bentler, 2001) and decreases in robust goodness of fit indices (ΔGFI), such as ΔCFI and ΔRMSEA. Because our observed variables were ordinal and not continuous, and our factor structure was multidimensional, we used Svetina and Rutkowski's (2017) recommended cut-offs for ΔGFI, namely, for the metric model, we used a ΔRMSEA cut-off of $\leq$ .05, and for the scalar model a cut-off of ΔRMSEA $\leq$ .01 and ΔCFI $\geq -.002$.

For analysis of variance (ANOVA) tests, effect sizes are reported in addition to statistical significance. Effect sizes were assessed with Cohen's *d* statistic for gender (Cohen, 1988), applying Cohen's benchmark guidelines of .20 for small effects, .50 for medium, and .80 for large effect sizes. Since age was a categorical variable with multiple groups, the more appropriate partial-eta squared (η_p^2) statistic was used to assess effect size (Lakens, 2013). Here, guidelines suggested by Cohen (1988) indicate .01 as a small effect, .06 as medium, and .14 as large effect size.

Reliability and Validity

Reliability of the new scale was assessed using Cronbach's α coefficient, McDonald's ω, and also via a 3-week test-retest reliability analysis. McDonald's ω was used in addition to the more common Cronbach's α, since it is believed to be a more accurate indicator of reliability when assumptions regarding tau equivalence are not met (Hayes & Coutts, 2020). Construct and factorial validity were assessed with CFA to determine if the factor structure was consistent with hardiness theory (a general hardiness factor with three subdimensions). Convergent and divergent validity were evaluated by assessing Pearson correlations between the new hardiness test and scores on several theoretically relevant measures, including coping styles and life satisfaction. Here we also refer to Cohen's (1988) conventions for effect size, such that a correlation of .10 is considered small, .30 moderate, and .50 or more as large.

Measures

In addition to the DRS-15 and the HRG items under development, three additional measures were administered to assess the validity of the new instrument in the normative sample.

Coping was measured with two scales from the Coping Inventory for Stressful Situations (CISS; Endler & Parker, 1994). Relationships between scores on the HRG and the CISS subscales of Task-Oriented Coping and Emotion-Oriented Coping were explored. Task-Oriented Coping entails actively pursuing long-term solutions to address

problems that are causing stress. Emotion-Oriented Coping involves emotional reactions to stressful events (e.g., getting angry), self-preoccupation, and fantasizing (e.g., daydreaming). These types of reactions are generally maladaptive. It was expected that respondents higher in hardiness would also be higher in Task-Oriented Coping and lower in Emotion-Oriented Coping. Each scale contains 16 items, with Cronbach's α coefficients ranging from .87 to .92 for Task-Oriented Coping and .82 to .90 for Emotion-Oriented Coping scale (Endler & Parker, 1994). In the present sample, Cronbach's α for Task-Oriented Coping was .93, and for Emotion-Oriented Coping .90.

The Satisfaction With Life Scale (SWLS) is a 5-item measure designed to capture global judgments of a person's overall life satisfaction (Diener et al., 1985). Cronbach's α in the present sample was .90. Hardiness scores should be associated with increased life satisfaction.

Relationship satisfaction was assessed using a single-item indicator ("I am satisfied with my relationships") on a response scale ranging from *Strongly Disagree* to *Strongly Agree*. Previous research has found that assessing relationship satisfaction with a single-item is an adequate approach for measuring relationship satisfaction (Fülöp et al., 2020). Hardiness scores should be associated with higher levels of relationship satisfaction.

Data Analysis and Statistics

All analysis was performed in R Version 4.0.4 (R Core Team, 2020). As detailed above, steps in the evaluation of test items for the HRG included examination of response distributions (including IRT techniques performed with the lordif and mirt package – version 0.3-3 and 1.34, respectively; Choi et al., 2011; Chalmers, 2012), assessment of skewness and kurtosis, and item-total correlations. CFA was applied using the Lavaan package (version 0.6-8; Rosseel, 2012) to test several different measurement models for the HRG data, with robust model fit statistics used to determine the best-fitting model, including the Satorra-Bentler scaled chi square test, CFI, TLI, RMSEA, and SRMR. Lavaan was also used for testing measurement invariance. Reliability of the new HRG scale and subscales was assessed with Cronbach's α statistic as well as McDonald's ω (Hayes & Coutts, 2020) – both statistics were calculated using the psych package (version 2.1.9; Revelle, 2021). Pearson correlations were used to evaluate the association of hardiness scores with relevant outcome measures (using the stats package version 3.6.2; R Core Team, 2021)

Results

Factor Structure

Application of all item selection criteria resulted in the retention of 28 items, 10 challenges, 10 commitments, and 8 controls (sample items for each subscale can be found in Table 2). The factor structure of the 28 items was expected to parallel the established hierarchical factor structure of the previous DRS scale, in which the three dimensions of challenge, control, and commitment are nested under the higher-order hardiness construct. CFA was performed using data from the normative sample. Three models were tested. The first was a single-factor model in which all items were loaded onto a single hardiness factor. The second model was a three-factor solution in which the facets of challenge, control, and commitment were uncorrelated. The third model was the hierarchical one in which the three facets were nested under a general hardiness factor. For all models, a polychoric variance-covariance matrix was estimated, and Weighted Least Squares Mean and Variance estimation was used. This was done to account for the ordinal nature of the data, which also did not fully meet assumptions of multivariate normality and homoscedasticity. Also, for this reason, we used the Satorra-Bentler scaled chi-square test (Satorra & Bentler, 2001), and the more stringent robust goodness of fit indices where applicable to determine the best fitting model, including the Satorra-Bentler scaled chi square test (Satorra & Bentler, 2001), the Comparative Fit Index (CFI; Bentler, 1990), the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973), the Standardized Root Mean Square Residual (SRMR; Hu & Bentler, 1999), and the Root Mean Square Error of Approximation (RMSEA; Steiger & Lind, 1980). Robust CFI and TLI values > .90 and SRMR and robust RMSEA values < .10 were used as indicators of adequate model fit (Hu & Bentler, 1999).

Table 2. Sample items from HRG Subscales

Subscale	Item
Challenge	I find the positives in any life change. Even if I fail at something, I look for ways to improve.
Control	I am confident I can accomplish whatever. I set out to do. I am responsible for my own success in life.
Commitment	I have a clear sense of purpose in my life. I feel energized about life.

Note. For access to the full Hardiness Resilience Gauge, please contact the publisher, Multi-Health Assessments (2018). https://www.storefront.mhs.com/collections/hrg

Results confirmed that the hierarchical model provided the best fit to the data, superior to alternative models. For the hierarchical model, robust fit indices were CFI = .91, TLI = .90, RMSEA = .09, and SRMR = .06. The comparable non-robust fit indices were considerably higher, at CFI = .98, TLI = .98, and RMSEA = .08. Complete model statistics are provided in Table 3.

Measurement Invariance

Measurement invariance analysis following Svetina and Rutkowski's (2017) approach revealed that males and

Table 3. Goodness of fit indices for three alternative HRG models

	One-factor	Three-factor orthogonal model	Three-factor hierarchical model
Number of parameters	112	112	115
Scaled χ^2	7,470.81	27,291.95	4,415.53
df	350	350	347
p	< .001	< .001	< .001
CFI	.84	.40	.91
TLI	.83	.35	.90
RMSEA	.12	.23	.09
SRMR	.08	.31	.06

Note. CFI = Comparative Fit Index; *df* = Degree of Freedom; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; TLI = Tucker-Lewis Index. Robust estimates for CFI, TLI, and RMSEA are reported due to some deviation from multivariate normality and homoscedasticity. The non-robust estimates were considerably higher (hierarchical model, CFI = .98, TLI = .98, and RMSEA = .08).

Table 4. HRG distribution by gender in the normative sample

	Males (*N* = 750)		Females (*N* = 750)				
	M	*SD*	*M*	*SD*	*F*(1, 1,490)	*p*	Cohen's *d*
Total hardiness	99.4	15.5	100.6	14.4	0.06	.808	0.09
Challenge	99.7	14.8	100.4	15.2	0.77	.381	0.05
Control	100.3	17.1	102.1	15.7	1.26	.262	0.12
Commitment	99.6	15.7	100.5	14.3	0.13	.723	0.06

Note. Recommended guidelines for evaluating Cohen's |*d*| are 0.20 = small, 0.50 = medium, 0.80 = large (Cohen, 1988).

females were invariant on the HRG (i.e., the conditions for strong invariance were met; Table 7, top portion). ANOVA results also showed that males (*N* = 750) and females (*N* = 750) did not differ significantly on their total standardized hardiness scores, nor on any of the three subscales. Further, none of the observed gender differences reached the minimum criterion for even a small effect size. These results suggest that the HRG is measuring males and females in the same way and that hardiness as measured by the HRG is similar for both males and females. Descriptive statistics, significance tests, and effect sizes are displayed in Table 4.

Age effects were also found to be negligible. Here, the sample was stratified into five age groups of *N* = 300 each, as follows: 18–24 years, 25–34 years, 35–54 years, 55–64 years, and 65+ years. Results verified measurement invariance across age groups (see the bottom portion of Table 7). It was also found that respondents of different ages did not differ significantly on their total standardized hardiness scores, nor on any of the three subscales – this finding was corroborated through examination of effect sizes, none of which reached the minimum criterion for even a small effect size. Taken together, the HRG appears to measure individuals of different age groups similarly, and hardiness scores on the HRG do not appear to substantially vary with age. Table 5 provides additional descriptive statistics, significance tests, and effect sizes for age.

Finally, there was no significant age by gender interactions, and none of the interactions reached even the small effect size level. Given the negligible age and gender effects (i.e., measurement invariance and no group mean differences), a single normative group representing the US general population was deemed most appropriate. Standard scores with a mean of 100 and a standard deviation of 15 were estimated for the challenge, control, and commitment subscales. For the challenge and commitment subscales, the distributions of standard scores showed appropriate dispersion and shape. For the control subscale, the normative means and standard deviations of raw scores were statistically smoothed at the high end of the distribution to ensure a similarly appropriate dispersion and shape (Roid, 1992; Zachary & Gorsuch, 1985). To compute total hardiness scores, standard scores of the three subscales were summed and standardized to a mean of 100 and a standard deviation of 15.

Reliability

Internal Consistency

Internal consistency is typically measured using Cronbach's α (Cronbach, 1951), which is a function of both the interrelatedness of the test items and the length of the test (John & Benet-Martinez, 2000). In the normative data, Cronbach's α for total hardiness was .93, indicating high reliability. The α values for the subscales were also high, at .89 for commitment, .84 for control, and .85 for challenge.

Hayes and Coutts (2020), among others, have argued for the use of McDonald's ω (1999) rather than Cronbach's α as a more robust measure of scale reliability (Peters, 2014). Unlike Cronbach's α, McDonald's ω does not

Table 5. HRG distribution by age in the normative sample

	18–24 (*N* = 300)		25–34 (*N* = 300)		35–54 (*N* = 300)		55–64 (*N* = 300)		65+ (*N* = 300)				
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*(4, 1,490)	*p*	η_p^2
Total hardiness	100.5	15.1	101	14.4	98.9	15.4	100.6	15.2	99	14.8	1.70	.148	.003
Challenge	100.7	14.7	101.5	14.5	99.2	15.2	100.0	15.4	98.7	15.1	2.98	.018	.004
Control	102.7	16.5	102.2	15.8	100.6	16.8	101.7	16.5	98.8	16.6	2.16	.077	.007
Commitment	99.1	15.4	100.1	14.7	98.6	15.4	101.3	15.0	101	14.4	0.62	.646	.005

Note. Guidelines for evaluating η_p^2 are .01 = small, .06 = medium, .14 = large (Cohen, 1988).

assume tau equivalence (the variance between a scale item and the latent variable measured by the scale is approximately equal across items in the scale). When tau equivalence is not met, Cronbach's α is likely to be a negatively biased estimator of reliability (Sijtsma, 2009). An examination of the standardized factor loadings, as well as fitting a CFA model in which factor loadings were constrained to be equal across items, revealed that assumptions of tau equivalence were not met for all HRG scales in the normative sample (i.e., model fit was worse than an unconstrained model). Thus, McDonald's ω is the more appropriate reliability coefficient. Table 8 (top portion) displays the McDonald's ω and Cronbach's α coefficients for the HRG and its subscales.

Test-Retest Reliability

In assessing test-retest reliability, a 3-week interval was deemed optimal. A shorter interval would increase the opportunity for memory effects to influence responses (Downie & Heath, 1970), while a longer interval provides the chance for real developmental shifts to affect responses. A subset of 168 randomly chosen individuals was selected from the normative sample to complete the HRG test twice over a three-week interval (mean interval = 22.1 days, *SD* = 1.6 days). The test-retest correlation for HRG total hardiness was high at $r = .81$. The hardiness subscales also showed good test-retest reliability with challenge at $r = .80$, control at $r = .74$, and commitment at $r = .79$. Effect sizes for the differences over time were small, indicating that the differences between Time 1 and Time 2 scores did not reveal any meaningful change in scores. Descriptive statistics, effect sizes, and test-retest correlations are shown in Table 6.

Validity

Relationship to Other Constructs

In addition to factorial validity as established by the CFA, the validity of the new scale was further evaluated by examining its association with other relevant psychological constructs. These analyses provide evidence for convergent and discriminant validity, evaluating the extent to which the new scale (HRG) is measuring the intended hardiness construct.

Hardiness Resilience Gauge and Coping Styles

As expected, results showed that HRG scores correlated positively with Task-Oriented Coping, $r = .68$, $p < .001$ and negatively with Emotion-Oriented Coping $r = -.30$, $p < .001$. An examination of the relationships between the HRG subscales and the CISS coping styles revealed similar results. Commitment, control, and challenge scale scores were all strongly positively correlated with Task-Oriented Coping and negatively correlated with Emotion-Oriented Coping. With Task-Oriented Coping, correlations with commitment, control and challenge were .61, .55, and .62 respectively (all $p < .001$). With Emotion-Oriented Coping, correlations with commitment, control, and challenge were −.31, −.22, and −.26, respectively (all $p < .001$). Overall, HRG scores were related to coping styles in expected ways. Respondents with higher HRG scores were more likely to use positive coping strategies and less likely to use negative coping strategies compared to those with low HRG scores.

Hardiness Resilience Gauge and Life Satisfaction

People who are high in hardiness experience a range of positive life outcomes, including less life and work stress and increased psychological well-being (Eschleman et al., 2010). As expected, total hardiness was positively related to SWLS scores, $r = .48$, $p < .001$. Thus, persons higher in hardiness reported more life satisfaction. Similar findings were observed when looking at the HRG subscales individually. For Satisfaction with Life, correlations were .56, .39, and .33 with commitment, control, and challenge, respectively, all $p < .001$.

HRG and Relationship Satisfaction

HRG total hardiness scores correlate as expected with Relationship Satisfaction, $r = .46$, $p < .001$. Those high in hardiness are more satisfied with the quality of their social relationships. For the hardiness facets, correlations are also positive with Relationship Satisfaction, at .52, .38, and .32 with commitment, control, and challenge, respectively, all $p < .001$. The strongest correlation is with commitment, which makes theoretical sense considering that this hardiness facet includes a commitment to one's social world.

Table 6. Three-week test-retest reliability for the HRG (N = 168)

		Time 1		Time 2			
	N	*M*	*SD*	*M*	*SD*	Cohen's *d*	Correlation between HRG Time 1 and Time 2
Total hardiness	168	101.1	18.7	103.7	17.6	−0.14	.81
Challenge	168	101.7	18.9	104.2	18.2	−0.13	.80
Control	168	99.8	18.1	102.7	17.1	−0.16	.74
Commitment	168	102.7	18.0	104.1	16.6	−0.08	79

Discussion

This first study documents the development and evaluation of an improved hardiness scale, the HRG. Applying careful item selection strategies, 28 items were retained for the final scale, representing the three hardiness facets of commitment, control, and challenge. The HRG shows the predicted hierarchical structure of three factors nested under a more general hardiness factor and proved to be invariant across gender and age groups. In a normative sample of N = 1,500, which was census-matched to the United States population, the new scale shows superior reliability for total hardiness, as well as for its subscales. The new scale also shows appropriate construct validity, as assessed through factor analysis, and convergent and divergent validity, as assessed through correlations with other measures, including coping styles and life satisfaction. This preliminary evidence is supported by additional validation efforts using two independent, non-US samples. This is described in detail in Study 2 below.

Study 2: Cross-Validation of the Hardiness Resilience Gauge on Independent Samples

Two additional independent samples were obtained to further evaluate the psychometric and structural properties of the new hardiness scale and assess its validity.

Canadian Sample

Using an online research panel, a sample of N = 394 respondents answered a web-based survey that included the HRG and the DRS-15 (there were 5 identical items between the two measures that were administered only once). Participants also completed a series of demographic questions and additional measures (described below). All respondents resided in Canada, and the sample was targeted to contain an equal representation of males and females. To ensure that participants' ages contained sufficient variability, the sample was evenly split between participants who fell within the 18–44 age range and participants who were above 45 years of age. Respondents were omitted if they violated at least two of the following criteria: (1) Excessively short (< 5 min) or long survey durations (> 40 min), (2) A large number of repeated consecutive responses to the survey questions (> 61% of the scale for the HRG and 40% for the DRS-15), (3) Large summed absolute differences (defined as > 2 standard deviations above the mean) between forward and reverse scored items for both the HRG and DRS-15, and (4) Being identified as a multivariate outlier as assessed through Mahalanobis Distance. Participants were also excluded if they had incomplete responses to either the DRS-15 or HRG. This resulted in the exclusion of 31 participants, leaving a final sample size of 363 participants. Of these, 69.7% identified as White, 21.4% as Asian, 2.8% as Black, and 6.1% as other. In terms of education, 20.1% were high school graduates, 35.4% some college, 32% bachelor's degree, and 9.9% graduate school. 63.4% were employed full or part-time, 24% retired, and 10.2% unemployed.

International Sample

An anonymized dataset containing N = 4,994 was provided by MHS – Multi-Health Systems (the HRG publisher) from their data archive. Respondents had completed the HRG and several other measures as part of a professional development activity or research project. Although a detailed breakdown of nationality is not available for this sample, we do know that respondents came from multiple countries, including Canada, Chile, China, Estonia, Finland, Germany, Greece, Hong Kong, Ireland, Korea, Lithuania, South Africa, Thailand, United Kingdom, and the United States of America. All participants completed the assessment in English. In addition to the HRG, participants were asked a series of optional questions on demographics and work-related outcomes. Excluded from the sample were any respondents who did not provide demographic information (e.g., gender, occupation) or answer the work-related outcome questions. This accounted for the majority of those excluded. Participants were also omitted if they had incomplete responses for the HRG. Of the N = 1,313 remaining participants, 14 cases were removed for violating at least two of the screening metrics defined above. The final sample consisted of N = 1,299 participants with complete data on the HRG, demographics, and work-related questions. Of these participants, 73.2% identified as female and 26.8% identified as male.

Measures

Canadian Sample

General anxiety was measured with the GAD-7 (General Anxiety Disorder-7; Spitzer et al., 2006). A 7-item scale, the authors report a Cronbach's α of .92 and test-retest reliability of .83. Earlier research on hardiness and anxiety has shown that hardiness is associated with lower levels of anxiety (Kovács & Borcsa, 2017; Kowalski & Schermer, 2019). Thus, we expected that people scoring higher on the HRG would report lower levels of anxiety. In the present sample, Cronbach's α for the GAD-7 was .94.

Depression was measured with the 9-item PHQ-9 scale (Patient Health Questionnaire-9; Kroenke et al., 2001). The authors report Cronbach's α ranging from .86 to .89. Consistent with previous research on hardiness and depression, it was expected that people who score higher on the HRG would report lower levels of depression (Bartone & Homish, 2020; Maddi et al., 2006; Ng & Lee, 2020; Sinha & Singh, 2009). In the present sample, Cronbach's α for the PHQ-9 was .92.

Coping in this sample was measured with 5 scales drawn from the Carver COPE Inventory (Carver, 1997). These are short 2-item scales, with reported Cronbach's α coefficients of .68 for Active Coping, .64 for Positive Reframing, .54 for Denial, .65 for Behavioral Disengagement, and .90 for Substance Abuse. In the present sample, Kendall's tau coefficients are reported as well as Cronbach's α, because the tau coefficient often provides a more accurate estimate of reliability for 2-item scales (Eisinga et al., 2013). Cronbach's α and Kendall's tau coefficients (in parentheses) were .80 (.63) for Active Coping, .80 (.60) for Positive Reframing, .46 (.31) for Denial, .79 (.68) for Behavioral Disengagement, and .91 (.83) for Substance Abuse. Because hardy people tend to use more effective coping strategies in dealing with stress, it was expected that people who score higher on the HRG would also score higher on the Active Coping and Positive Reframing subscales, and lower on Denial, Behavioral Disengagement, and Alcohol/Substance Abuse subscales.

International Sample

Stress appraisal was measured with five items. Three of the items related to positive stress appraisal, asking participants to indicate how strongly they believed stress is something that is useful and beneficial. An example item is: "There are benefits to experiencing stress." Two of the items related to negative stress appraisal, asking participants to indicate how strongly they believed stress is something to be avoided and that stress prevents them from doing the things that they want to do. Responses were on a 5-point scale, ranging from *Strongly Disagree* to *Strongly Agree*. The five questions were summed to create an aggregate measure (both scales for the negative appraisal questions were reversed). Cronbach's α was acceptable at α = .74. In line with previous research, we hypothesized that people scoring higher on hardiness would be more likely to see stress as potentially beneficial and not something that is negative and to be avoided (Florian et a, 1995; Skomorovsky & Sudom, 2011; Stein & Bartone, 2020).

Work performance was measured with two items: "I am engaged at work," and "I am successful in my job." Responses were on a 5-point Likert scale, ranging from *Strongly Disagree* to *Strongly Agree*. The two questions were summed to create an aggregate measure. Cronbach's α (α = .70) was acceptable given the 2-item nature of the scale, while Kendall's tau (τ = .36) indicated a moderate correlation between the two measures. It was hypothesized that higher hardiness scores would be associated with greater work performance (i.e., higher engagement and satisfaction).

Data Analysis and Statistics

The same statistical procedures described in Study 1 above for the normative sample were also applied with the cross-validation samples in order to further assess the reliability and validity of the new HRG hardiness scale.

Results

Measurement Invariance

The earlier result showing measurement invariance for gender in the normative sample was confirmed in the International cross-validation sample. That is, males and females were invariant in terms of how they responded on the HRG (i.e., the conditions for strong invariance were met). These results are presented in Table 7 (middle portion). As reported earlier, HRG responses were found to be invariant across age groups in the Normative sample (Table 7, bottom portion). Due to the relatively small sample size, we were unable to test measurement invariance for gender or age in the Canadian sample, and age information was not available in the International sample.

Reliability

Cronbach's α values were obtained in both cross-validation samples. For the Canadian sample, α values were .95 for total hardiness, .91 for commitment, .86 for control, and .86 for challenge. For the International sample, these values were .91 for total hardiness, .87 for commitment, .77 for control, and .82 for challenge. Examination of the standardized factor loadings, as well as fitting a CFA model in which factor loadings were constrained to be equal across items, showed that the assumption of tau equivalence was not met for HRG scales in both cross-validation samples. As with the Study 1 normative sample, McDonald's ω was considered to be the more appropriate reliability coefficient for these samples. For comparison purposes, Table 8 displays

Table 7. Measurement invariance tests for gender and age groups

Model	Scaled χ^2	*df*	CFI	RMSEA	$\Delta\chi^2$
Normative: Male vs. Female					
Configural	4,551.10	694	.914	.086	
Metric	4,590.10	722	.914	.085	$\Delta\chi^2(28) = 30.08$, $p > .05$
Scalar	4,409.91	747	.918	.081	$\Delta\chi^2(25) = 32.11$, $p > .05$
International: Male vs. Female					
Configural	3,302.32	694	.901	.076	
Metric	3,305.05	708	.902	.075	$\Delta\chi^2(14) = 15.94$, $p > .05$
Scalar	3,194.66	733	.907	.072	$\Delta\chi^2(25) = 26.68$, $p > .05$
Normative: All 5 age groups					
Configural	5,383.50	1,735	.921	.084	
Metric	5,469.28	1,843	.921	.081	$\Delta\chi^2(108) = 108.68$, $p > .05$
Scalar	5,233.49	1,943	.929	.075	$\Delta\chi^2(100) = 106.65$, $p > .05$

Note. The metric model includes equality constraints across thresholds, whereas the scalar model includes equality constraints across thresholds and factor loadings (Svetina & Rutkowski, 2017). Chi square is Satorra-Bentler scaled chi square. Delta parameterization was used for all models.

both the McDonald's ω and Cronbach's α coefficients for the HRG and its subscales in all three samples.

Validity

Factor Structure

Once again, the factor structure of the HRG was expected to parallel the hierarchical model demonstrated in the (Study 1) normative sample, in which the three dimensions of challenge, control, and commitment are nested under the higher-order construct of hardiness. The same CFA procedures used with the normative sample were also applied for both cross-validation samples in order to verify the best-fitting model. Results confirmed that the 3-factor hierarchical model showed the best fit in both samples, although goodness of fit indices were slightly below the acceptable threshold for the International sample. For the Canadian sample, the fit indices for the three-factor hierarchical model were: CFI = .93, TLI = .92, RMSEA = .09 and SRMR = .07. For the International sample, the fit indices were: CFI = .89, TLI = .88, RMSEA = .08 and SRMR = .07. These results are summarized in Table 9.

Relationships to Other constructs – Validation Samples

Canadian Sample

Hardiness Resilience Gauge and Generalized Anxiety. As expected, total hardiness was negatively correlated with Generalized Anxiety Disorder scores as assessed through the GAD-7, $r = -.38$, $p < .001$. Thus, people who are higher in hardiness have lower self-reported scores of Generalized Anxiety Disorder. All subscales of the HRG were significantly negatively associated with GAD-7 scores, with the commitment scale showing the strongest negative correlation at $r = -.39$, $p < .001$.

Hardiness Resilience Gauge and Depression. Also, as expected, total hardiness was negatively correlated with depression scores on the PHQ-9, $r = -.38$, $p < .001$.

Table 8. McDonald's ω and Cronbach's α reliability coefficients for the HRG and its subscales

	Commitment	Control	Challenge	Hardiness total
Normative sample				
McDonald's ω	.92	.90	.89	.87
Cronbach's α	.85	.84	.89	.93
Canadian sample				
McDonald's ω	.94	.91	.90	.91
Cronbach's α	.91	.86	.86	.95
International sample				
McDonald's ω	.92	.86	.88	.85
Cronbach's α	.87	.77	.82	.91

Respondents who are higher in hardiness also report fewer depression symptoms. All subscales of the HRG were significantly negatively correlated with PHQ-9 Depression scores, with the commitment scale again showing the strongest negative correlation, $r = -.41$, $p < .001$.

Hardiness Resilience Gauge and Coping. The HRG and five coping styles were examined. As expected, total hardiness was positively correlated with both Active Coping, $r = .55$, $p < .001$, and Positive Reframing, $r = .51$, $p < .001$. Also as expected, total hardiness was negatively correlated with Behavioral Disengagement, $r = -.16$, $p < .01$ and Substance Use, $r = -.11$, $p < .05$, although the latter relationships were weaker than for the positive coping indicators. Total hardiness score also showed a weak but positive correlation with Denial, $r = .11$, $p > .05$.

International Sample

Hardiness Resilience Gauge and Stress Appraisal. As expected, total hardiness scores were positively correlated with Stress Appraisal, $r = .43$, $p < .001$. Thus, people who are more hardy are more likely to interpret stress as something that is potentially positive rather than uniformly negative. The

Table 9. Goodness of fit indices for three HRG factor models for the Canadian and International samples

	One-factor	Three-factor orthogonal model	Three-factor hierarchical model
Canadian sample			
Number of parameters	111	111	114
Scaled χ^2	1,805.31	9,810.07	1,324.77
df	350	350	347
p	< .001	< .001	< .001
CFI	.90	.33	.93
TLI	.89	.28	.92
RMSEA	.11	.27	.09
SRMR	.08	.37	.07
International sample			
Number of parameters	112	112	115
Scaled χ^2	5,423.98	15,540.15	3,385.76
df	350	350	347
p	< .001	< .001	< .001
CFI	.81	.44	.89
TLI	.80	.39	.88
RMSEA	.11	.18	.08
SRMR	.09	.28	.07

Note. CFI = Comparative Fit Index; *df* = Degree of Freedom; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; TLI = Tucker-Lewis Index. CFI, TLI and RMSEA values are robust estimates. Also, the Canadian sample had 1 less parameter estimate than the International sample due to one less threshold parameter being estimated for an item (i.e., not all response options were endorsed for that item, so it was collapsed with an adjacent category).

challenge subscale showed the strongest correlation with Stress Appraisal, $r = .45$, $p < .001$.

Hardiness Resilience Gauge and Work Performance. Also as expected, total hardiness scores were positively correlated with work performance, $r = .42$, $p < .001$. Thus, people who are more hardy are also more likely to report being successful and more engaged in their jobs. The commitment sub-scale showed the strongest correlation with Work Performance, $r = .45$, $p < .001$.

For ease of reference, correlations between HRG scales and all criterion indicators for all three samples are summarized in Table 10.

Correlations With Earlier Hardiness Test

As an additional check on validity, we assessed the Pearson correlation between the scales of the new HRG and the previous hardiness measure, the DRS-15, in the Canadian sample ($N = 363$). HRG scores were first standardized to correct the unequal number of items in the subscales. Results showed the HRG correlated strongly with the DRS, for overall hardiness ($r = .87$), for the commitment scale ($r = .84$), for the control scale ($r = .91$), and for the challenge scale ($r = .61$).

Hardiness Resilience Gauge vs. Dispositional Resilience Scale in Predicting Outcomes

The relationship between the HRG and various outcome variables in the Canadian sample was explored while controlling for DRS-15 scores. The HRG should predict the outcomes above and beyond the DRS-15 if it is to offer enhanced predictive utility over the original assessment. In other words, does the HRG show incremental predictive validity beyond what the previous measure accounts for? For these analyses, five highly similar items between the HRG and DRS-15 were removed from the HRG raw score in order to reduce potential multicollinearity between the scales. Hierarchical linear regressions were performed using the stats package (version 3.6.2) in R (R Core Team, 2021), 7 regressions with the total raw scores for both the HRG and DRS-15, and 7 regressions with the raw scores for just the challenge subscale of both instruments. The challenge scale was of special interest because that is where the greatest differences exist between the DRS-15 and the HRG. For all regressions, DRS-15 raw scores were entered in the first block, and HRG raw scores were entered in the second block. The change in R^2 was examined.

For total hardiness, results showed that the HRG accounted for significant unique variance in Positive Reframing Coping, Active Coping, and Denial Coping. For the challenge subscale, the HRG accounted for significant unique variance in PHQ-9 Depression, GAD-7 Anxiety, Positive Reframing Coping, Active Coping, and Denial Coping. HRG total scores did not account for significant unique variance in PHQ-9, GAD-7, Behavioral Disengagement Coping, or Substance Use. HRG challenge likewise did not account for significant unique variance in either the Behavioral Disengagement Coping or Substance Use. Table 11 summarizes these results.

Table 10. Correlations between the HRG hardiness scales and various criterion indicators for all three samples

	Commitment	Control	Challenge	Hardiness total
Normative sample				
Task-oriented coping (*N* = 1,471)	.61***	.55***	.62***	.68***
Emotion-oriented coping (*N* = 1,473)	−.31***	−.22***	−.26***	−.30***
Life satisfaction (*N* = 1,500)	.56***	.39***	.33***	.48***
Relationship satisfaction (*N* = 1,463)	.52***	.38***	.32***	.46***
Canadian sample				
Depression – PHQ-9 (*N* = 346)	−.41***	−.35**	−.26***	−.38***
Anxiety – GAD-7 (*N* = 351)	−.39***	−.33***	−.32***	−.38***
Active coping (*N* = 353)	.51***	.46***	.51***	.55***
Positive reframing (*N* = 353)	.48***	.38***	.52***	.51***
Denial (*N* = 353)	.12*	−.03	.22***	.11*
Behavioral disengagement (*N* = 353)	−.14**	−.19***	−.08	−.16**
Substance use (*N* = 352)	−.13*	−.12*	−.06	−.11*
International sample				
Stress appraisal (*N* = 1,270)	.31***	.35***	.45***	.43***
Work performance (*N* = 1,260)	.45***	.30***	.35***	.42***

Note. *$p < .05$; **$p < .01$; ***$p < .001$.

Discussion

In this Study 2, we explored the reliability and validity of the HRG in two independent samples, one consisting of Canadian nationals and the other made up of respondents from multiple countries who completed the HRG as part of a professional development activity or for a research study. As with Study 1, results showed high-reliability coefficients in both samples, as well as good evidence for validity. For HRG total hardiness and the three subscales, Cronbach's α ranged from .77 to .95, whereas McDonald's ω ranged from .85 to .94. The HRG also showed measurement invariance across gender, lending further support to its general applicability. Factorial validity was also confirmed, with a three-factor hierarchical model showing the best fit in both independent samples. The HRG also showed predicted associations with relevant outcome measures, including anxiety, depression, positive and negative coping strategies, stress appraisals, and work performance. Finally, the HRG was found to account for significant unique variance in a number of the outcome variables while controlling for the DRS-15; this was especially true for the challenge subscale, which underwent the most extensive revision in the development of the HRG. Overall results from Study 2 provide strong support for the reliability and validity of the HRG as a measure of psychological hardiness.

General Discussion

Since it was first reported by Kobasa (1979), multiple studies have shown that psychological hardiness is an important factor influencing individual resilience, health, and human performance under stress. Nevertheless, research and applications involving hardiness have been somewhat hampered by the lack of a fully reliable and comprehensive, standardized tool for measuring it. This is due in part to the complex nature of the hardiness construct, which contains multiple facets that are themselves multidimensional. The present work addresses this gap, providing a hardiness measure with increased reliability, shown in terms of internal consistency (Cronbach's α = .93; McDonald's ω = .87) and 3-week test-retest reliability (*r* = .81). Reliability coefficients for the three hardiness facets of commitment, control, and challenge, are also uniformly high. Further, the HRG shows enhanced validity, with better content area coverage for the hardiness facets of challenge, control, and commitment. In particular, challenge has been poorly measured in earlier hardiness tests, where items have focused almost exclusively on attitudes toward variety and novelty, and neglecting other theoretically important aspects such as the tendency to regard change and disruption as interesting life challenges and opportunities to learn. The new HRG challenge scale incorporates these additional components of challenge, thus providing a more construct-valid indicator of this important hardiness dimension.

The HRG hardiness measure also shows the expected hierarchical structure of three distinct factors nested under a general hardiness dimension. This supports the theoretical notion that the three factors work together, creating a whole that is more than the sum of its parts. For example, having a strong sense of control – the belief that one can influence outcomes – can reinforce and encourage the sense of challenge or willingness to take risks and try new things. Likewise, having a strong sense of purpose or meaning in life can encourage and facilitate a sense of control

Table 11. Regressions with both HRG and DRS hardiness measures predicting various outcome indicators

	Test statistic	*p*	R^2	Δ*F*
Total HRG and DRS				
PHQ9 ~ DRS	*F*(1, 344) = 115.30	< .001	.25	Δ*F*(1) = 1.04, *ns*
PHQ9 ~ HRG + DRS	*F*(2, 343) = 58.19	< .001	.25	
GAD7 ~ DRS	*F*(1, 349) = 106.40	< .001	.23	Δ*F*(1) = 0.17, *ns*
GAD7 ~ HRG + DRS	*F*(2, 348) = 53.15	< .001	.23	
AC ~ DRS	*F*(1, 351) = 103.40	< .001	.23	Δ*F*(1) = 36.08, *p* < .001
AC ~ HRG + DRS	*F*(2, 350) = 74.90	< .001	.30	
PR ~ DRS	*F*(1, 351) = 80.13	< .001	.19	Δ*F*(1) = 35.85, *p* < .001
PR ~ HRG + DRS	*F*(2, 350) = 61.97	< .001	.26	
D ~ DRS	*F*(1, 351) = 0.0006	*ns*	.00	Δ*F*(1) = 16.88, *p* < .001
D ~ HRG + DRS	*F*(2, 350) = 8.44	< .001	.05	
BD ~ DRS	*F*(1, 351) = 18.87	< .001	.05	Δ*F*(1) = 0.95, *ns*
BD ~ HRG + DRS	*F*(2, 350) = 9.91	< .001	.05	
SU ~ DRS	*F*(1, 350) = 12.61	< .001	.03	Δ*F*(1) = 1.45, *ns*
SU ~ HRG + DRS	*F*(2, 349) = 7.04	< .001	.04	
Challenge subscale for HRG and DRS				
PHQ9 ~ DRS	*F*(1, 344) = 13.75	< .001	.04	Δ*F*(1) = 11.53, *p* < .001
PHQ9 ~ HRG + DRS	*F*(2, 343) = 12.85	< .001	.07	
GAD7 ~ DRS	*F*(1, 349) = 25.07	< .001	.07	Δ*F*(1) = 16.56, *p* < .001
GAD7 ~ HRG + DRS	*F*(2, 348) = 21.37	< .001	.11	
AC ~ DRS	*F*(1, 351) = 19.05	< .001	.05	Δ*F*(1) = 106.09, *p* < .001
AC ~ HRG + DRS	*F*(2, 350) = 65.42	< .001	.27	
PR ~ DRS	*F*(1, 351) = 16.11	< .001	.04	Δ*F*(1) = 106.09, *p* < .001
PR ~ HRG + DRS	*F*(2, 350) = 65.42	< .001	.27	
D ~ DRS	*F*(1, 351) = 3.24	*ns*	.01	Δ*F*(1) = 2.02, *ns*
D ~ HRG + DRS	*F*(2, 350) = 9.37	< .001	.05	
BD ~ DRS	*F*(1, 351) = 0.53	*ns*	.001	Δ*F*(1) = 15.37, *p* < .001
BD ~ HRG + DRS	*F*(2, 350) = 1.28	*ns*	.01	
SU ~ DRS	*F*(1, 350) = 1.23	*ns*	.004	Δ*F*(1) = 0.24, *ns*
SU ~ HRG + DRS	*F*(2, 349) = 0.74	*ns*	.004	

Note. AC = Active Coping; BD = Behavioural Disengagement; D = Denial; DRS = Dispositional Resilience Scale; GAD7 = General Anxiety Disorder-7 Scale; HRG = Hardiness Resilience Gauge; PHQ9 = Patient Health Questionnaire-9; PR = Positive Reframing; SU = Substance Use. The tilde (~) symbol indicates that the variable on the left was regressed on the right-hand side variable(s).

over one's own destiny. Past research applications have fruitfully examined the influence of overall hardiness as well as the three facets separately, although results with the facets are sometimes ambiguous (Eid et al., 2008; Johnsen et al., 2009). And while the three facets generally tend to intercorrelate and go up or down together, this may not always be true. In some cases, people can be high in certain hardiness facets and low in others. For example, in their study of Norwegian Army soldiers deployed to Kosovo, Johnsen et al. (2014) found that while for most subjects their hardiness facet scores were internally consistent (all high, medium, or low), some subjects showed deviant or unbalanced profiles. The largest deviant group was the "Rigid Controls," who were low in hardiness challenge despite being high in commitment and control. Also identified were "Sensation-seekers" who were high in challenge but low in commitment and control. In this study, soldiers with unbalanced "Rigid-Control" hardiness profiles also reported more health problems, similar to the low hardy group. In another study that looked at hardiness profiles, Norwegian navy cadets with unbalanced hardiness profiles – commitment, control, and challenge scores not in agreement – also showed more extreme and unhealthy physiological stress reactions (Sandvik et al., 2013). With the improved hardiness facet scales in the HRG, similar investigations into the influence of the separate hardiness facets are more feasible and should yield more clear and valid results.

One criticism of earlier hardiness measures is that the tests contained a male bias and did not apply equally well to women (Riska, 2002; Klag & Bradley, 2004). This is an issue of measurement equivalence or invariance. Measurement invariance across gender has been previously shown for the DRS hardiness scale but not for all items

(Kardum et al., 2012). Hystad (2012) demonstrated gender equivalence in the DRS-15 but noted some non-equivalence in the control scale. The new HRG hardiness scale is an advance in this regard, with demonstrated measurement invariance across two large and independent samples.

Another important consideration is if measures such as the HRG are invariant or hold equally well across age groups. Our results show the HRG displays measurement invariance across the broad age range of 18–65 years. Future studies will be needed to determine the validity of the HRG in younger age groups.

In both studies and all three of the present samples, the HRG showed expected associations with relevant criterion indicators. Hardiness was negatively associated with anxiety (PHQ-9) and depression (GAD-7) scores. Previous studies have also linked hardiness to lower anxiety. For example, in a study of university undergraduates, Kowalski and Schermer (2019) found that hardiness was negatively correlated with anxiety, even after controlling for neuroticism. It is worth noting that PTSD is also an anxiety disorder, and there are multiple studies showing that hardiness is a protective factor against PTSD (e.g., Bartone, 1999; Escolas et al., 2013; Thomassen et al., 2018).

Consistent with previous studies, the present research also finds fairly strong associations of HRG hardiness scores with depression. For example, Maddi et al. (2006) found that hardiness was negatively related to depression in a sample of US Army War College students. Other studies have found a link between hardiness and depression in a variety of samples (Bartone & Homish, 2020; Ganellen & Blaney, 1984; Ng & Lee, 2020). HRG hardiness scores also showed the expected positive association with measures of well-being, life satisfaction, and satisfaction with relationships. The positive relation of hardiness to well-being has also been frequently reported (e.g., Kowalski & Schermer, 2019; Bartone & Bowles, 2021). In the present study, the HRG subscale of commitment showed the strongest correlations with relationship satisfaction and life satisfaction. This makes sense considering that hardiness commitment involves a strong engagement with the social world and a sense of passion and deep involvement in life (Kobasa & Maddi, 1977; Stein & Bartone, 2020).

People high in hardiness tend to rely on more adaptive coping styles that involve taking action to address the source of their stress (Eschleman et al., 2010). Likewise, they tend not to use avoidance or emotional coping strategies that are not aimed at solving problems. Therefore, we expected that higher scores on the HRG would be related to more adaptive coping styles. This was largely the case with HRG hardiness scores. In Study 1, hardiness correlated quite strongly with task-oriented coping and negatively with emotion-oriented coping, and these results were consistent across the HRG subscales (Table 10). This trend also holds in Study 2 with the Canadian sample, where HRG hardiness was positively associated with active coping and positive reframing, while negatively correlated with behavioral disengagement and alcohol/substance abuse.

A somewhat surprising finding is that the HRG challenge scale shows a small but positive correlation with denial coping. This may reflect the tendency of high challenge people to welcome change and disregard risk to some extent when facing uncertain situations. Just such a situation is presented by the COVID pandemic, which was a dominant concern when the present data were being collected in early 2021. It is also the case that when facing stressful situations, especially those outside of our control, some amount of denial can be adaptive. For example, Horowitz (2001) has described the normal, healthy grief process following the loss of a loved one as involving periods of denial, when one puts the loss out of mind in order to go on with life and accomplish necessary tasks, and periods of intrusion when one is fully aware and experiencing the grief pain of loss. Similarly, when facing stressors like an infectious disease pandemic that is largely outside of individual control, a tendency to minimize the threat and get on with the tasks of living could be equally adaptive. Still, the weight of the evidence in the present studies shows that HRG hardiness is associated with positive problem-solving coping strategies, and negatively associated with avoidance coping styles. Future research is needed to identify under what conditions different coping styles are most effective in dealing with stressful situations.

Study 2 also showed, in a large International sample, that persons high in HRG hardiness tend to make more positive stress appraisals, seeing potential benefits even in difficult situations. These individuals also rate themselves as more engaged and successful in their work. While hardiness theory clearly indicates that hardiness should facilitate growth from stress, there is to date only modest empirical support for this. One recent study addressing this issue examined US Army recruiters, who face considerable pressure in their jobs to meet monthly production quotas (Bartone & Bowles, 2021). The authors report that hardiness in a large national sample of recruiters was related to post-traumatic growth and increased well-being. Further study is needed to address the question of when stress exposure can lead to actual growth and improvement and what role hardiness may play in this process.

In comparing the new HRG hardiness scale with the previous short DRS-15 version, the correlations obtained here were all quite high (.84 or above), with the exception of the challenge scale, which was .61. This was expected considering that the previous challenge scale had weak construct validity, which the new scale attempts to correct with better coverage of the challenge dimension. For example, the old challenge scale did not address the aspect of

learning from experience and failures, which the HRG challenge scale now includes. So the lower correlation between the HRG and DRS-15 challenge scales is expected, and reflects the improved construct validity of the new scale.

A final important question is: does the new HRG hardiness measure provide incremental predictive validity over the earlier version? Based on a series of regression analyses performed with the Canadian sample, the answer appears to be yes. The HRG scale showed significantly increased power to predict active, problem-solving coping and positive reframing. The HRG challenge scale, in particular, showed incremental validity in predicting anxiety, depression, and positive and negative coping approaches. The HRG thus shows improved predictive power over the most recent earlier hardiness test.

Limitations

Several limitations of this research should be mentioned. In assessing predictive validity, the present study relies on cross-sectional data in which relevant outcome indicators were collected at the same time point as HRG hardiness scores. A preferred design would be a longitudinal one in which outcome measures are collected at a later point in time, thus allowing for more clear inferences regarding causal directionality.

Another potential limitation in the present work concerns our reliance on an online survey methodology using Mechanical Turk (MTurk). While this methodology has become quite common in the social sciences, some questions have been raised about the representativeness of samples obtained in this way. For example, some researchers have pointed out that MTurk samples tend to be younger, more unemployed, and more depressed (Ophir et al., 2020; Paolacci & Chandler, 2014). However, these same studies indicate that these limitations can be largely overcome by the use of careful screening methods. In the present study, we took steps to exclude any respondents with mental health problems, and any who showed spurious response patterns or lack of attention to the task. Further, in the normative sample (Study 1), we deliberately sampled particular demographic groups for age, gender, employment status, education, and geographic region, thus providing greater assurance regarding generalizability of results.

In Study 2, the anonymized nature of the International sample data imposed some limitations due to the lack of demographic data. Lacking information on age, we were unable to test for HRG measurement invariance across age groups. In addition, because all participants in the International sample completed the survey in English, and we do not know if English was their primary language, it is possible that there were differences in English proficiency within this sample which could have impacted how they responded to the questionnaire. The relatively small size of the Canadian sample also prevented testing for measurement invariance across age. However, the HRG did prove to be invariant across age groups in the larger, normative sample. Future studies should explore this issue further in other groups, and also assess the applicability of the HRG in teenagers and adolescents.

Conclusion

While hardiness has generally been conceived as a fairly stable trait, there is evidence that it can also be increased through training programs and life experiences (Bartone, Eid, et al., 2016; Stein & Bartone, 2020). Considering the growing interest in programs that aim specifically to increase the stress resilience qualities associated with hardiness, it is all the more important that a highly valid and reliable measure of hardiness be available to researchers and practitioners. The HRG was developed to help meet this need.

The HRG is fairly brief at 28 items and easy to administer. It was designed to be easily adaptable for use in other languages and free of cultural bias and idiomatic expressions. The HRG shows appropriate convergent and discriminant validity, as well as a good fitting hierarchical model. Likewise, it demonstrates good generalizability across multiple and diverse samples and is valid for women and men and different age groups. The HRG thus provides an important advance in measuring the hardiness construct, and should prove to be a useful tool in research, clinical and consulting settings.

References

Andrew, M. E., Mnatsakanova, A., Howsare, J. L., Hartley, T. A., Charles, L. E., Burchfiel, C. M., McCanlies, E. C., & Violanti, J. M. (2013). Associations between protective factors and psychological distress vary by gender: The Buffalo Cardio-Metabolic Occupational Police Stress Study. *International Journal of Emergency Mental Health, 15*(4), 277–288.

Bartone, P. T. (1984). *Stress and health in Chicago Transit Authority bus drivers (Unpublished doctoral dissertation)*. Department of Behavioral Sciences, University of Chicago.

Bartone, P. T. (1989). Predictors of stress-related illness in city bus drivers. *Journal of Occupational Medicine, 31*, 657–663. https://doi.org/10.1097/00043764-198908000-00008

Bartone, P. T. (1991, June). *Development and validation of a short hardiness measure*. Paper presented at the American Psychological Society Annual Convention, Washington, DC. http://www.hardiness-resilience.com/docs/aps91b.pdf

Bartone, P. T. (1995, June). *A Short Hardiness Scale*. Poster presented at the American Psychological Society Annual Convention, New York, NY. https://www.researchgate.net/publication/331875947_A_Short_Hardiness_Scale_updated

Bartone, P. T. (1999). Hardiness protects against war-related stress in Army reserve forces. *Consulting Psychology Journal, 51*, 72–82. https://doi.org/10.1037/1061-4087.51.2.72

Bartone, P. T. (2006). Resilience under military operational stress: Can leaders influence hardiness? *Military Psychology, 18*(Suppl.), S131–S148. https://doi.org/10.1207/s15327876mp1803s_10

Bartone, P. T. (2013). *Cross-cultural adaptation of the DRS-15 Dispositional Resilience Scale: A short hardiness measure*. Final report on Fulbright Research Fellowship. http://www.hardiness-resilience.com/docs/Crosscultural adaptation of DRS Fulbright.pdf

Bartone, P. T., & Bowles, S. V. (2021). Hardiness predicts post-traumatic growth and well-being in severely wounded servicemen and their spouses. *Military Medicine, 186*(5–6), 500–504. https://doi.org/10.1093/milmed/usaa250

Bartone, P. T., Eid, J., & Hystad, S. W. (2016). Training hardiness for stress resilience. In N. Maheshwari & V. Kumar (Eds.), *Military psychology: Concepts, trends and interventions* (pp. 231–248). Sage.

Bartone, P. T., Eid, J., Hystad, S. W., Jocoy, K., Laberg, J. C., & Johnsen, B. H. (2015). Psychological hardiness and avoidance coping are related to risky alcohol use in returning combat veterans. *Military Behavioral Health, 3*, 274–282. https://doi.org/10.1080/21635781.2015.1085931

Bartone, P. T., Eid, J., Johnsen, B. H., Laberg, J. C., & Snook, S. A. (2009). Big Five personality factors, hardiness, and social judgment as predictors of leader performance. *Leadership & Organization Development Journal, 30*(6), 498–521. https://doi.org/10.1108/01437730910981908

Bartone, P. T., & Homish, G. G. (2020). Influence of hardiness, avoidance coping, and combat exposure on depression in returning war veterans: A moderated-mediation study. *Journal of Affective Disorders, 265*, 511–518. https://doi.org/10.1016/j.jad.2020.01.127

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

Bartone, P. T., Ursano, R. J., Wright, K. M., & Ingraham, L. H. (1989). The impact of a military air disaster on the health of assistance workers: A prospective study. *Journal of Nervous and Mental Disease, 177*(6), 317–328. https://doi.org/10.1097/00005053-198906000-00001

Bartone, P. T., Valdes, J. J., & Sandvik, A. (2016). Psychological hardiness predicts cardiovascular health. *Psychology, Health & Medicine, 21*, 743–749. https://doi.org/10.1080/13548506.2015.1120323

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246. https://doi.org/10.1037/0033-2909.107.2.238

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. https://doi.org/10.1037/1076-8998.6.1.53

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford Press.

Carver, C. S. (1997). You want to measure coping but your protocol's too long: Consider the brief COPE. *International Journal of Behavioral Medicine, 4*(1), 92–100. https://doi.org/10.1207/s15327558ijbm0401_6

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and monte carlo simulations. *Journal of Statistical Software, 39*(8), 1–30. https://doi.org/10.18637/jss.v039.i08

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (1st ed.). Erlbaum.

Contrada, R. J. (1989). Type A behavior, personality hardiness, and cardiovascular responses to stress. *Journal of Personality and Social Psychology, 57*, 895–903. https://doi.org/10.1037/0022-3514.57.5.895

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–335. https://doi.org/10.1007/BF02310555

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Eid, J., Johnsen, B. H., Bartone, P. T., & Nissestad, O. A. (2008). Growing transformational leaders: Exploring the role of personality hardiness. *Leadership & Organization Development Journal, 29*(1), 4–23. https://doi.org/10.1108/01437730810845270

Eisinga, R., Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health, 58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Endler, N. S., & Parker, J. D. A. (1994). Assessment of multidimensional coping: Task, emotion, and avoidance strategies. *Psychological Assessment, 6*(1), 50–60. https://doi.org/10.1037/1040-3590.6.1.50

Eschleman, K. J., Bowling, N. A., & Alarcon, G. M. (2010). A meta-analytic examination of hardiness. *International Journal of Stress Management, 17*(4), 277–307. https://doi.org/10.1037/a0020476

Escolas, S. M., Pitts, B. L., Safer, M. A., & Bartone, P. T. (2013). The protective value of hardiness on military posttraumatic stress symptoms. *Military Psychology, 25*(2), 116–123. https://doi.org/10.1037/h0094953

Florian, V., Mikulincer, M., & Taubman, O. (1995). Does hardiness contribute to mental health during a stressful real-life situation? The roles of appraisal and coping. *Journal of Personality and Social Psychology, 68*, 687–695. https://doi.org/10.1037/0022-3514.68.4.687

Funk, S. C. (1992). Hardiness: A review of theory and research. *Health Psychology, 11*, 335–345. https://doi.org/10.1037/0278-6133.11.5.335

Fülöp, F., Bőthe, B., Gál, E., Cachia, J., Demetrovics, Z., & Orosz, G. (2020). A two-study validation of a single-item measure of relationship satisfaction: RAS-1. *Current Psychology*. https://doi.org/10.1007/s12144-020-00727-y

Ganellen, R. J., & Blaney, P. H. (1984). Hardiness and social support as moderators of the effects of life stress. *Journal of Personality and Social Psychology, 47*(1), 156–163. https://doi.org/10.1037/0022-3514.47.1.156

Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. *The Leadership Quarterly, 31*(4), Article 101384. https://doi.org/10.1016/j.leaqua.2020.101384

Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629

Horn, J. L., McArdle, J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *Southern Psychologist, 1*(4), 179–188.

Horowitz, M. J. (2001). *Stress response syndromes: Personality styles and intervention* (4th ed.). Aronson.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new

alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hystad, S. W. (2012). Exploring gender equivalence and bias in a measure of psychological hardiness. *International Journal of Psychological Studies, 4*(4), 69–79. https://doi.org/10.5539/ijps.v4n4p69

Hystad, S. W., Eid, J., Johnsen, B. H., Laberg, J. C., & Bartone, P. T. (2010). Psychometric properties of the Revised Norwegian Dispositional Resilience (Hardiness) Scale. *Scandinavian Journal of Psychology, 51*, 237–245. https://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*(5), 421–429. https://doi.org/10.1080/00313830903180349

Hystad, S. W., Olsen, O. K., Espevik, R., & Säfvenbom, R. (2015). On the stability of psychological hardiness: A three-year longitudinal study. *Military Psychology, 27*, 155–168. https://doi.org/10.1037/mil0000069

Johnsen, B. H., Bartone, P., Sandvik, A. M., Gjeldnes, R., Morken, A. M., Hystad, S. W., & Stornæs, A. (2013). Psychological hardiness predicts success in a Norwegian Armed Forces border patrol selection course. *International Journal of Selection and Assessment, 21*, 368–375. https://doi.org/10.1111/ijsa.12046

Johnsen, B. H., Eid, J., Pallesen, S., Bartone, P. T., & Nissestad, O. A. (2009). Predicting transformational leadership in naval cadets: Effects of personality hardiness and training. *Journal of Applied Social Psychology, 39*(9), 2213–2235. https://doi.org/10.1111/j.1559-1816.2009.00522.x

Johnsen, B. J., Hystad, S. W., Bartone, P. T., Laberg, J. C., & Eid, J. (2014). Hardiness profiles: Defining clusters of the Dispositional Resilience Scale and their relation to soldiers' health. *Military Behavioral Health, 2*(2), 123–128. https://doi.org/10.1080/21635781.2013.867789

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*(4), 409–426. https://doi.org/10.1007/BF02291366

Kardum, I., Hudek-Knežević, J., & Krapić, N. (2012). The structure of hardiness, its measurement invariance across gender and relationships with personality traits and mental health outcomes. *Psihologijske Teme, 21*(3), 487–507.

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. https://doi.org/10.1348/135910704773891014

Kobasa, S. C. (1979). Personality and resistance to illness. *American Journal of Community Psychology, 7*(4), 413–423. https://doi.org/10.1007/BF00894383

Kobasa, S. C., & Maddi, S. R. (1977). Existential personality theory. In R. J. Corsini (Ed.), *Current personality theories* (pp. 243–276). Peacock.

Kobasa, S. C., & Puccetti, M. A. (1983). Personality and social resources in stress resistance. *Journal of Personality and Social Psychology, 44*(4), 839–850. https://doi.org/10.1037/0022-3514.45.4.839

Kovács, I. K., & Borcsa, M. (2017). The relationship between anxiety, somatic symptoms and hardiness in adolescence. *Romanian Journal of Applied Psychology, 19*(2), 42–49.

Kowalski, C. M., & Schermer, J. A. (2019). Hardiness, perseverative cognition, anxiety, and health-related outcomes: A case for and against psychological hardiness. *Psychological Reports, 122*(6), 2096–2118. https://doi.org/10.1177/0033294118800444

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology, 4*, Article 863. https://doi.org/10.3389/fpsyg.2013.00863

Lambert, V. A., Lambert, C. E. Jr., Klipple, G. L., & Mewshaw, E. A. (1989). Social support, hardiness and psychological well-being in women with arthritis. *Image: The Journal of Nursing Scholarship, 21*, 128–131. https://doi.org/10.1111/j.1547-5069. 1989.tb00116.x

Maddi, S. R. (2002). The story of hardiness: Twenty years of theorizing, research, and practice. *Consulting Psychology Journal: Practice and Research, 54*, 173–185. https://doi.org/10.1037/1061-4087.54.3.175

Maddi, S. R., Brow, M., Khoshaba, D. M., & Vaitkus, M. (2006). Relationship of hardiness and religiousness to depression and anger. *Consulting Psychology Journal: Practice and Research, 58*(3), 148–161. https://doi.org/10.1037/1065-9293.58.3.148

Maddi, S. R., & Kobasa, S. C. (1984). *The hardy executive*. Jones-Irwin.

Madrigal, L., Gill, D. L., & Eskridge, K. M. (2016). Examining the reliability, validity and factor structure of the DRS-15 with college athletes. *Psihologijske Teme, 25*(2), 263–280.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Erlbaum.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543. https://doi.org/10.1007/BF02294825

Mortensen, K., & Hughes, T. L. (2018). Comparing Amazon's Mechanical Turk platform to conventional data collection methods in the health and medical research literature. *Journal of General Internal Medicine, 33*, 533–538. https://doi.org/10.1007/s11606-017-4246-0

Multi-Health Systems, Inc. (2018). *Hardiness Resilience Gauge* [Measurement instrument]. MHS. https://www.storefront.mhs.com/collections/hrg

Nabizadeh, F., & Mahdavi, A. (2016). Relationship between hardiness and marital satisfaction in women with breast cancer. *Archives of Breast Cancer, 3*, 92–96. https://doi.org/10.19187/abc.20163392-96

Ng, S. M., & Lee, T. M. C. (2020). The mediating role of hardiness in the relationship between perceived loneliness and depressive symptoms among older. *Aging & Mental Health, 24*(5), 805–810. https://doi.org/10.1080/13607863.2018.1550629

Ophir, Y., Sisso, I., Asterhan, C. S. C., Tikochinski, R., & Reichart, R. (2020). The Turker blues: Hidden factors behind increased depression rates among Amazon's Mechanical Turkers. *Clinical Psychological Science, 8*(1), 65–83. https://doi.org/10.1177/2167702619865973

Ouellette, S. C. (1993). Inquiries into hardiness. In L. Goldberger & S. Breznitz (Eds.), *Handbook of stress: Theoretical and clinical aspects* (2nd ed., pp. 77–100). Free Press.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science, 23*(3), 184–188. http://www.jstor.org/stable/44318769

Peters, G. J. (2014). The alpha and omega of scale reliability and validity. *The European Health Psychologist, 16*(2), 56–69.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2021). *Functions in stats (version 3.6.2). Stats 3.6.2*. https://www.rdocumentation.org/packages/stats/versions/3.6.2

Revelle, W. (2021). *Procedures for psychological, psychometric and personality research. Psych 2.1.9*. https://cran.r-project.org/web/packages/psych/psych.pdf

Riska, E. (2002). From Type A man to the hardy man: Masculinity and health. *Sociology of Health & Illness, 24*, 347–358. https://doi.org/347-358.10.1111/1467-9566.00298

Roid, G. H. (1992, April). *Smoothing score distributions using moments: Applications in the norming of tests*. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, CA.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Sandvik, A. M., Bartone, P. T., Hystad, S. W., Phillips, T. M., Thayer, J. F., & Johnsen, B. H. (2013). Psychological hardiness predicts neuroimmunological responses to stress. *Psychology, health & medicine, 18*(6), 705–713. https://doi.org/10.1080/13548506.2013.772304

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

Sijtsma, K. (2009). On the use, misuse and the very limited usefulness of cronbach's alpha. *Psychometrika, 74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sinclair, R. R., & Tetrick, L. E. (2000). Implications of item wording for hardiness structure, relation with neuroticism, and stress buffering. *Journal of Research in Personality, 34*(1), 1–25. https://doi.org/10.1006/jrpe.1999.2265

Sinha, V., & Singh, R. N. (2009). Immunological role of hardiness on depression. *Indian journal of psychological medicine, 31*(1), 39–44. https://doi.org/10.4103/0253-7176.53314

Skomorovsky, A., & Sudom, K. A. (2011). Role of hardiness in the psychological well-being of Canadian Forces officer candidates. *Military Medicine, 176*(1), 7–12. https://doi.org/10.7205/milmed-d-10-00325

Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of internal medicine, 166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Steiger, J. H., & Lind, J. M. (1980, May). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Stein, S. J., & Bartone, P. T. (2020). *Hardiness: Making stress work for you to achieve your life goals*. Wiley.

Svetina, D., & Rutkowski, L. (2017). Multidimensional measurement invariance in an international context: Fit measure performance with many groups. *Journal of Cross-Cultural Psychology, 48*(7), 991–1008. https://doi.org/10.1177/0022022117717028

Svetina, D., Rutkowski, L., & Rutkowski, D. (2019). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using Mplus and the lavaan/semTools packages. *Structural Equation Modelling: A Multidisciplinary Journal, 27*(1), 1–20. https://doi.org/10.1080/10705511.2019.1602776

Thomassen, Å. G., Hystad, S. W., Johnsen, B. H., Johnsen, G. E., & Bartone, P. T. (2018). The effect of hardiness on PTSD symptoms: A prospective mediational approach. *Military Psychology, 30*(2), 142–151. https://doi.org/10.1080/08995605.2018.1425065

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1–10. https://doi.org/10.1007/BF02291170

United States Bureau of the Census. (2016). *United States Census 2016*. Government Printing Office.

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

Windle, G., Bennett, K. M., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and Quality of Life Outcomes, 9*, Article 8. https://doi.org/10.1186/1477-7525-9-8

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika, 81*(4), 1014–1045. https://doi.org/10.1007/s11336-016-9506-0

Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology, 41*(1), 86–94. https://doi.org/10.1002/1097-4679

Zanon, C., Hutz, C. S., Yoo, H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica, 29*, Article 18. https://doi.org/10.1186/s41155-016-0040-x

History
Received November 3, 2020
Revision received January 14, 2022
Accepted February 3, 2022
Published online April 1, 2022
EJPA Section / Category Positive Psychology & Assessment

Acknowledgment
Portions of this paper were presented at the 54th International Applied Military Psychology Seminar, May 6–10, 2020, Sarajevo, Bosnia, and Herzegovina.

Conflict of Interest
Paul T. Bartone receives royalties for the Hardiness Resilience Gauge.

Open Science
We report how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. If we use inferential tests, we report exact *p*-values, effect sizes, and 95% confidence or credible intervals.
Open Data: The information needed to reproduce all of the reported results are not openly accessible due to privacy concerns.
Open Materials: The information needed to reproduce all of the reported methodology is not openly accessible. The material is available on request from MHS Assessments.
Preregistration of Studies and Analysis Plans: This study was not preregistered.

ORCID
Paul T. Bartone
https://orcid.org/0000-0002-5767-0086

Paul T. Bartone
Institute for National Strategic Studies
National Defense University
Fort Lesley J. McNair
300 5th Ave., SW
Washington, DC 20319
USA
bartonep@gmail.com