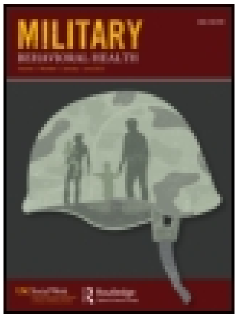


Military Behavioral Health

ISSN: 2163-5781 (Print) 2163-5803 (Online) Journal homepage: http://www.tandfonline.com/loi/umbh20

# Confirmatory Factor Analysis of a Measure of Comprehensive Airman Fitness

Gary L. Bowen, Todd M. Jensen & James A. Martin

MILITARY BEHAVIORAL HEALTH
2016, VOL. 0, NO. 0, 1–11
http://dx.doi.org/10.1080/21635781.2016.1199984

Routledge
Taylor & Francis Group

# Confirmatory Factor Analysis of a Measure of Comprehensive Airman Fitness

Gary L. Bowen[a], Todd M. Jensen[a], and James A. Martin[b]

[a]University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; [b]Bryn Mawr College, Bryn Mawr, Pennsylvania

**ABSTRACT**

The U.S. Air Force has committed significant resources to implementing policies and programs consistent with the Department of Defense's concept of *total force fitness*. A 12-item measure of Comprehensive Airman Fitness was proposed and empirically examined, using component measures of mental fitness, physical fitness, social fitness, and spiritual fitness from the Support and Resiliency Inventory. Results confirm that the components of airman fitness can be conceptualized as pieces of a total fitness construct and that the measure is invariant across subgroups. Implications for policy and practice are discussed, and an agenda for future research is presented.

**KEYWORDS**
Comprehensive Airman Fitness; Total Force Fitness; U.S. military; U.S. Air Force; Support and Resiliency Inventory



The concept of *total force fitness* was developed by the Department of Defense in 2009 to focus efforts on promoting "health, resilience, and optimal performance" in the context of demands and challenges faced by military personnel and their families (Institute for Alternative Futures, 2009, p. 2). Admiral Mike Mullen, Chairman of the Joint Chiefs of Staff (2007–2011), spearheaded the initiative "Total Force Fitness for the 21st Century," which had application to all of the service branches (Jonas et al., 2010). Integrating components of fitness, health, resilience, and readiness, the concept of total force fitness included eight domains, including four domains of mind (psychological, behavioral, spiritual, and social) and four domains of body (physical, medical, nutritional, and environmental; Jonas et al., 2010; M. Mullen, 2010).

In response to Admiral Mullen's leadership and parallel to efforts in the U.S. Army (Comprehensive Soldier Fitness; Casey, 2011), the U.S. Air Force, which is the focus of this investigation, embraced the concept of Comprehensive Airman Fitness, including the specification of four fitness domains represented in the Department of Defense model: mental, physical, social, and spiritual. Through its contracting relationship with RAND Corporation, the Air Force has given considerable attention to conducting literature reviews across the domains of Total Force Fitness (see Meadows, Miller, & Robson, 2015, for an overarching summary of this work).

Air Force efforts were initiated in the Air Combat Command in 2010, followed shortly thereafter by efforts in the Air Mobility Command (J. Michel, personal communication, May 15, 2015); the Air Force officially launched its Comprehensive Airman Fitness program on March 30, 2011 (Gonzalez, Singh, Schell, & Weinick, 2014). Air Force Instruction 90–506, 2 April 2014, established requirements for the program with the airman population broadly defined to encompass all Air Force military members, Air Force civilian employees, and all Air Force military and civilian employee family members.

Air Force Instruction 90–506 defines Comprehensive Airman Fitness as an "integrated framework" rather than a "stand-alone program" or "specified training class" (p. 3). As specified in the instruction, the program "encompasses many cross-functional education and training efforts, activities, programs, and other equities that play a contributory role in sustaining a fit, resilient, and ready force" (p. 3). The Deputy Chief of Staff for Manpower, Personnel and Services (HQ USAF/A1) is the focal point for activities related to implementation of this instruction.

Unlike the U.S. Army, which made assessment a key aspect of the Comprehensive Soldier Fitness program from the beginning (Peterson, Park, & Castro, 2011), the Air Force has not established a metric for assessing Airman fitness and its related components, although Air

**CONTACT** Gary L. Bowen glbowen@email.unc.edu School of Social Work, The University of North Carolina at Chapel Hill, 325 Pittsboro Street, Chapel Hill, NC 27599.

The views and opinions contained in this article are those of the authors and should not be construed as official Department of the Air Force position, policy, or decision, unless so designated by other authorized documents. Data reported in this article were collected by Flying Bridge Technologies, Inc., Charlotte, N.C., under Task Order Number 9Q1 SFSRAB001, General Services Administration. The Office of Human Research Ethics at the University of North Carolina at Chapel Hill determined that the proposed secondary analysis of these data was exempt from human subject review.

Force Instruction 90-506 calls for such measures: Comprehensive Airman Fitness "metrics/indicators derived from defined measures and self-reported data provided in community-based Air Force surveys will be used to provide commanders a view of the comprehensive fitness of an organization" (p. 11). Even the comprehensive, biennial Air Force Community Assessment (Air Force Instruction 90–501, 15 October 2013), which is used at the installation level, in part, to assess member needs and strengths, does not specifically capture and report information on these four fitness domains. Other sources of data, such as the Air Force Climate Survey and the Caring for People Survey, are available but none of these sources focus explicitly on assessing the four fitness components (Meadows et al., 2015).

The Support and Resiliency Inventory, which is a brief (12–15-min), anonymous, self-administered, web-based, Air Force–sponsored assessment tool, offers potential promise for deriving a Comprehensive Airman Fitness metric from existing self-report data from Air Force active duty members (Bowen & Martin, 2011). Such a measure of total force fitness and its related components has important implications for Air Force leaders and practitioners charged with understanding, promoting, and supporting Comprehensive Airman Fitness through the Community Action Information Board and Integrative Delivery System at each Air Force installation (Air Force Instruction 90–501, 15 October 2013), as well as for researchers interested in monitoring and evaluating policy and program interventions to promote fitness. Both evidence-informed policy and practice and intervention research depend on reliable and valid measures of intended outcomes (Fraser & Galinsky, 2010; E. J. Mullen, 2004).

Using a sample of active-duty Air Force members who completed the Support and Resiliency Inventory during a 2-week period in January of 2012, we examined the factor structure of the 12-item fitness measure using confirmatory factor analysis. We hypothesized that the 12 items represent four first-order latent factors (mental fitness, physical fitness, social fitness, and spiritual fitness) and a higher order latent factor (total fitness). We also examined the results of the confirmatory factor analysis for measurement invariance by important sociodemographic variables—military pay grade, gender, marital status, and deployment in the past 12 months.

## The support and resiliency inventory

The Support and Resiliency Inventory was originally developed and pilot-tested with sponsorship from the Air Force Space Command Family Matters Office (2004–2007) in the context of its unit services outreach strategy (Orthner, Bowen, & Mancini, 2003). In its early years of development, the inventory was known as the Unit Assets Inventory, with parallel versions for Air Force members and the civilian spouses of these members, including the Personal Assets Inventory (Huebner, Mancini, Bowen, & Orthner, 2009). The Support and Resiliency Inventory received Air Force–wide sponsorship in 2008, which continued until 2013, from the Airman and Family Services Division as part of its Community Readiness Consultant Practice Strategy (Bowen, Martin, Liston, & Nelson, 2009). Sponsorship for the Support and Resiliency Inventory shifted to the HQ Air Force Resilience Division in 2013/2014 for use by Air Force installation community support coordinators who are the focal point ("specialist and facilitator") for Comprehensive Airman Fitness at the installation level (Air Force Instruction 90–506, April 2, 2014, p. 7). At present, the tool is available to installations via individual contracts through a private corporation in Charlotte, North Carolina.

In 2011, the Support and Resiliency Inventory was revised, in part, to better capture the concept of Comprehensive Airman Fitness; 12 items are now used to assess the four domains of Airman fitness (three items per domain). This process has involved a bit of trial and error, including informative work with the U.S. Marine Corps to develop a similar metric (Bowen & Martin, 2013a). It is important to note that the three items on the social fitness domain were recently shifted from a focus on "willingness to seek help from others" to "the ability to depend on support from others," which we believe better reflects the nominal definition of social fitness in Air Force Instruction 90-506 (see Table 1). This shift did not result in a need to revise the Support and Resiliency Inventory—all items are included on the 2011 revision (Bowen & Martin, 2011), although the individual and group summary reports have not been revised to incorporate this change.

Despite approximately 59,000 Air Force members and employees completing the Support and Resiliency Inventory from 2011 to 2014, to date, the conceptual integrity of the fitness measure has not received sufficient empirical attention. Exploratory factor analysis of earlier data

**Table 1.** U.S. air force definitions of four fitness domains.

| Fitness domain | Definition |
|---|---|
| Mental fitness | The ability to effectively cope with unique mental stressors and challenges. |
| Physical fitness | The ability to adopt and sustain healthy behaviors needed to enhance health and well-being |
| Social fitness | The ability to engage in healthy social networks that promote overall well-being and optimal performance. |
| Spiritual fitness | The ability to adhere to beliefs, principles, or values needed to persevere and prevail in accomplishing missions. |

*Source.* Air Force Instruction 90-506, April 2, 2014, Comprehensive Airman Fitness, pp. 15–16.

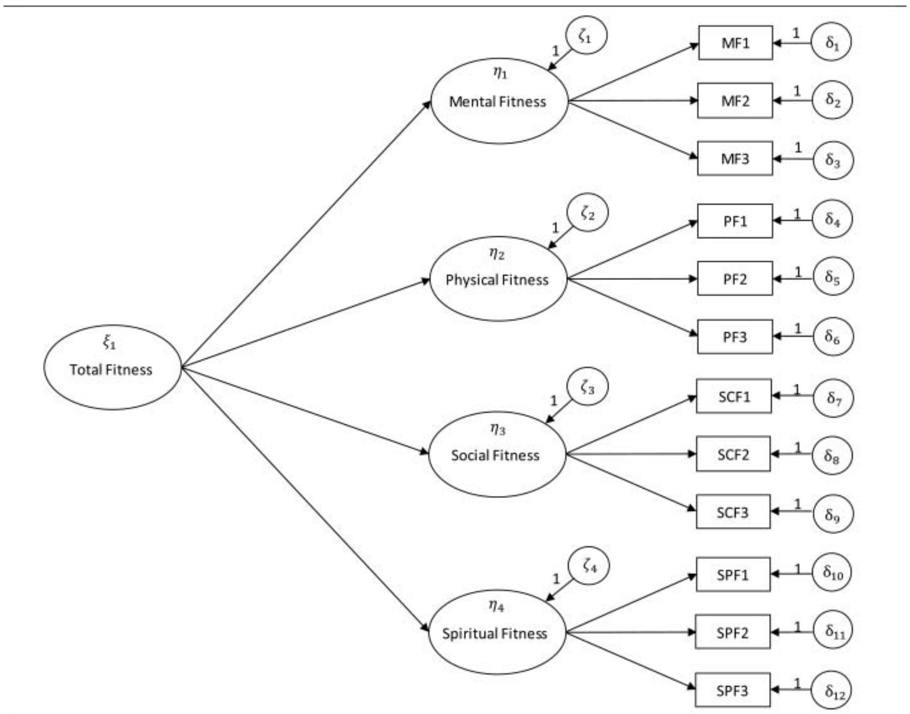
supported the presence of the four distinct fitness factors (mental, physical, social, and spiritual) and high levels of internal consistency were demonstrated within factors (Bowen & Martin, 2013b). Yet, questions remain as to whether the four fitness measures can be used to represent a useful total score, which is a particularly efficient way to measure fitness in empirical research and a simple way to display results for practitioners. Questions also remain regarding the relative invariance of the measure (component and total) across population subgroups. Measurement invariance across key population subgroups suggests that an instrument reliably captures the same phenomenon for members of each subgroup—a desirable characteristic of any assessment tool used within a diverse target population, such as active-duty Air Force members. Valid comparisons between population subgroups depend upon invariance in the measure at hand, which is more often assumed than confirmed in studies.

## Hypothesized model

Total fitness is a multicomponent factor that includes four domains: mental, physical, social, and spiritual. Table 1 includes nominal definitions of these four fitness domains, as defined in Air Force Instruction 90-506 (pp. 15–16).

Figure 1 illustrates the hypothesized model that is tested in this investigation. The model shows a total of 12 observed variables associated with four first-order latent factors (mental, physical, social, and spiritual). In addition, the model shows a second-order factor structure in which the four first-order latent factors load onto a higher order latent factor, total fitness.

The measurement invariance of the conceptual model is examined in the context of pay grade (E1–E4, E5–E6, E7–E9, O1–O3, O4–O10), gender (female, male), marital status (single, married), and deployment in the past 12 months (no, yes). These variables are frequently used in research to study variation in outcomes for military members and their families. Of these variables, pay grade is most often used in military studies for purposes of making subgroup comparisons on outcomes of interest (e.g., Bowen, Jensen, Martin, & Mancini, 2016; Bowen, Mancini, Martin, Ware, & Nelson, 2003). Special attention is often directed to junior enlisted members (E1–E4). Compared with their more senior counterparts (mid-level and senior enlisted and officers), junior enlisted members (E1–E4) have less influence over the nature of their day-to-day assignments and job responsibilities and less supervisory and leadership responsibilities for and authority over others (Hamaoka et al., 2014). Many junior enlisted members also struggle with the demands associated with new marriages and young children in the context of military policies that actively encourage members to pursue marriage and parenthood at a young age (Lundquist & Xu, 2014).




*Note:* For metric calibration, the variances/error variances of first- and second-order factors are fixed to 1.

**Figure 1.** Hypothesized second-order factor model.

## Method

### Source of data

In January 2012, Air Force Chief of Staff General Norton A. Schwartz (2008–2012) and Chief Master Sergeant of the Air Force James A. Roy (2009–2013) directed a 1-day stand down for all Air Force units worldwide to focus on member, unit, and community resiliency. The stand down was in direct response to the uptick in the U.S. Air Force suicide rate in the first 2 weeks of 2012.

Associated with this mandatory event, individuals and units (including military members and Air Force civilian employees) were offered the opportunity to complete the web-based Support and Resiliency Inventory as a means of facilitating stand-down discussions (Department of the Air Force, Office of the Chief of Staff, January 12, 2012). For 2 weeks (January 12, 2012–January 26, 2012), 11,885 Air Force members and civilian employees voluntarily completed the Support and Resiliency Inventory in support of this command directive. All responses were anonymous.

Although information from the Support and Resiliency Inventory was intended to inform the design, delivery, and evaluation of program and services to promote the fitness and resilience of total force Air Force members and civilian employees at the unit, base or Major Command level, this administration had a specific purpose—to allow total force Air Force members to examine their own fitness and resilience profile as part of the stand-down conversation. The inventory was well suited for this purpose because respondents were able to download a graphical summary of their responses at the end of the survey, including their fitness profile. A web-based worksheet provided respondents with an opportunity to develop an individual plan of action for increasing their success in adapting to life challenges and meeting military life and duty responsibilities.

### Sample profile

The present analysis focuses on the 8,730 respondents from the larger sample who reported that they were currently serving on active duty (regular component). Civilian employees and members of the Air Force Reserve and Air National Guard were not included in the present analysis—these employees and members face a rather unique set of occupational circumstances and challenges (Redmond et al., 2015). Also, active duty members who were currently deployed were deleted from the sample ($n$ = 209). Respondents represented all of the major commands in the Air Force with the exception of the Air Force Global Strike

**Table 2.** Variable and sample description for the full sample ($N$ = 8,730).

| Variable | *N* | *M* | *SD* | Skewness | Kurtosis | Min | Max | Missing values |
|---|---|---|---|---|---|---|---|---|
| Fitness variables | | | | | | | | |
| Mental fitness | | | | | | | | |
| MF1 | 8,556 | 7.11 | 2.63 | −0.83 | 2.95 | 0 | 10 | 1.99% |
| MF2 | 8,582 | 8.15 | 2.08 | −1.36 | 4.77 | 0 | 10 | 1.70% |
| MF3 | 8,582 | 7.77 | 2.26 | −1.19 | 4.07 | 0 | 10 | 1.70% |
| Physical fitness | | | | | | | | |
| PF1 | 8,596 | 7.24 | 1.90 | −0.65 | 3.60 | 0 | 10 | 1.53% |
| PF2 | 8,591 | 7.85 | 2.16 | −1.07 | 3.84 | 0 | 10 | 1.59% |
| PF3 | 8,595 | 7.58 | 1.88 | −0.77 | 3.62 | 0 | 10 | 1.55% |
| Social fitness | | | | | | | | |
| SCF1 | 8,479 | 8.11 | 2.59 | −1.44 | 4.24 | 0 | 10 | 2.88% |
| SCF2 | 8,488 | 7.67 | 2.57 | −1.08 | 3.43 | 0 | 10 | 2.77% |
| SCF3 | 8,510 | 7.02 | 2.82 | −0.75 | 2.61 | 0 | 10 | 2.52% |
| Spiritual fitness | | | | | | | | |
| SPF1 | 8,461 | 8.70 | 1.86 | −1.97 | 7.56 | 0 | 10 | 3.08% |
| SPF2 | 8,451 | 8.60 | 1.87 | −1.88 | 7.28 | 0 | 10 | 3.20% |
| SPF3 | 8,369 | 8.27 | 2.24 | −1.66 | 5.69 | 0 | 10 | 4.14% |
| Grouping variables | | | | | | | | |
| Deployed during past 12 months (1 = yes) | 8,664 | 0.23 | | | | 0 | 1 | 0.76% |
| Gender (1 = male) | 8,664 | 0.79 | | | | 0 | 1 | 0.76% |
| Marital status (1 = married) | 8,730 | 0.59 | | | | 0 | 1 | 0.00% |
| Pay grade | 8,650 | | | | | | | 0.92% |
| E1–E4 | | 0.31 | | | | | | |
| E5–E6 | | 0.33 | | | | | | |
| E7–E9 | | 0.14 | | | | | | |
| O1–O3 | | 0.10 | | | | | | |
| O4 and higher | | 0.11 | | | | | | |
| Other characteristics | | | | | | | | |
| Age (years) | 8,658 | | | | | | | 0.82% |
| Younger than 26 | | 0.43 | | | | | | |
| 26-35 years | | 0.30 | | | | | | |
| 36 and older | | 0.26 | | | | | | |
| Parent or stepparent (1 = yes) | 8,730 | 0.51 | | | | | | |

Command, the Air Force Special Operations Commands, and the Air Force Reserve Command, although it was not possible to determine the major command of respondents who used the portal-based self-administration rather than the unit-based administration of the Support and Resiliency Inventory.

Table 2 includes a profile description of the full sample. Overall, respondents approximated the profile of the Air Force active duty population (Department of Defense, Office of the Deputy Assistant Secretary of Defense, Military Community and Family Policy, 2012). The modal respondent was male (79%), married (59%), a parent or stepparent (51%), had not been deployed in the past 12 months (77%), and in either the junior-enlisted (31%) or mid-enlisted (33%) pay grade profile group. Approximately two in five respondents were younger than age 26 years (43%).

**Table 3.** Observed indicators for each first-order latent construct.

| Construct/item label | Description |
|---|---|
| Mental fitness ($\alpha = .90$) | |
| MF1 | I look forward to beginning each day. |
| MF2 | I keep a positive outlook on life. |
| MF3 | I enjoy most days. |
| Physical fitness ($\alpha = .86$) | |
| PF1 | I maintain a healthy diet. |
| PF2 | I exercise on a regular basis. |
| PF3 | I maintain a healthy lifestyle. |
| Social fitness ($\alpha = .79$) | |
| SCF1 | I can depend on support from one or more extended family members, if I need it. |
| SCF2 | I can depend on support from one or more friends, if I need it. |
| SCF3 | I can depend on support from one or more members of my unit (or place of work), if I need it. |
| Spiritual fitness ($\alpha = .94$) | |
| SPF1 | I have a guiding set of principles or beliefs. |
| SPF2 | I attempt to live in accordance with a guiding set of principles or beliefs. |
| SPF3 | I draw strength from a set of guiding principles or beliefs. |

*Note*. All dimensions range from 0 (*not at all*) to 10 (*completely*).

## Measures

### Substantive variables

We used 12 items to assess the four first-order constructs in the empirical model: mental fitness (3 items: MF1, MF2, and MF3), physical fitness (3 items: PF1 PF2, and PF3), social fitness (3 items: SCF1, SCF2, and SCF3), and spiritual fitness (3 items: SPF1, SPF2, and SPF3). Table 3 presents the items that corresponded to each construct. Modeled after Cantril's (1965) self-anchoring ladder scale, each item was assessed on the same 11-point slider scale ranging from 0 (*not at all*) to 10 (*completely*). Although the design of the rating scale was driven more by the design of the online survey and the ease of using a slider scale on a handheld device, Lozano, Garcia-Cueto, and Muniz (2008) reported that, in general, the reliability and validity of a measure increase as the number of response options increase. When comparing 5-point scales with 11-point scales, Dawes (2002) found modest mean differences between the two scale formats when the 5-point scale was rescaled for comparison; however, the 11-point scale had a greater amount of variance (coefficient of variation) than did the 5-point scale. As Dawes (2002) concluded, scales that produce greater variance have benefit in examining the relations among variables.

Table 2 presents descriptive statistics (means, standard deviations) for these measures, and Table 4 presents the associated correlation matrix. The alpha coefficients for the summary measures ranged from a low of .79 for social fitness to a high of .94 for spiritual fitness.

### Grouping variables

Four grouping variables were used for the conduct of measurement invariance tests. Participants were partitioned into five pay grade groups, representing each the following levels: (a) E1–E4, (b) E5–E6, (c) E7–E9, (d) O1–O3, and (e) O4 and higher. Gender was a binary measure, representing either male or female participants. Marital status was a binary measure such that those who indicated being married were grouped together, and those who indicated being single and never married, legally separated, divorced, or widowed were grouped together. Deployment in the past 12 months was a binary measure that partitioned participants on the basis of whether they had been deployed in the past 12 months.

**Table 4.** Correlation matrix for observed indicators.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Mental fitness | | | | | | | | | | | |
| 1 MF1 | | | | | | | | | | | |
| 2 MF2 | 0.71* | | | | | | | | | | |
| 3 MF3 | 0.79* | 0.79* | | | | | | | | | |
| Physical fitness | | | | | | | | | | | |
| 4 PF1 | 0.40* | 0.39* | 0.37* | | | | | | | | |
| 5 PF2 | 0.35* | 0.33* | 0.33* | 0.57* | | | | | | | |
| 6 PF3 | 0.43* | 0.45* | 0.43* | 0.78* | 0.69* | | | | | | |
| Social fitness | | | | | | | | | | | |
| 7 SCF1 | 0.33* | 0.36* | 0.35* | 0.23* | 0.18* | 0.25* | | | | | |
| 8 SCF2 | 0.38* | 0.42* | 0.42* | 0.28* | 0.26* | 0.31* | 0.56* | | | | |
| 9 SCF3 | 0.42* | 0.42* | 0.45* | 0.27* | 0.25* | 0.29* | 0.46* | 0.66* | | | |
| Spiritual fitness | | | | | | | | | | | |
| 10 SPF1 | 0.38* | 0.42* | 0.38* | 0.30* | 0.24* | 0.34* | 0.24* | 0.28* | 0.27* | | |
| 11 SPF2 | 0.39* | 0.44* | 0.39* | 0.31* | 0.24* | 0.34* | 0.24* | 0.28* | 0.28* | 0.89* | |
| 12 SFP3 | 0.42* | 0.45* | 0.41* | 0.30* | 0.24* | 0.34* | 0.25* | 0.29* | 0.28* | 0.81* | 0.83* |

*Note*. Analysis included nonmissing data ($N = 8{,}204$ to 8,596). All variance inflation factor scores across items were below 6 in the context of a supplemental analysis, indicating no issue with multicollinearity.
*$p < .05$.

### *Data analysis*

We first examined the distributional properties of each of the 12 observed indicators in the hypothesized model. Tests of multivariate normality were conducted and indicated significant nonnormality; however, these tests are highly sensitive to sample size (Kline, 2011). Consequently, we examined the skew index and kurtosis index values associated with each observed indicator (see Table 1; Kline, 2011). All skew index values were less than 2 (average $= -1.22$), and all kurtosis index values were less than 8 ($M = 4.47$). This indicated that the distributions of our measures were not necessarily problematic (Curran, West & Finch, 1996; Kline, 2011). A correlation matrix was also estimated for all observed indicators to assess interitem associations. All univariate and bivariate analyses were conducted in Stata 13.0 (StataCorp, 2013).

After a descriptive and bivariate examination of our observed indicators, our analysis consisted of two key components: (a) confirmatory factor analysis and tests of alternative factor structures, and (b) measurement invariances tests with respect to pay grade (five groups), gender (two groups), marital status (two groups), and deployment in the past 12 months (two groups). We used structural equation modeling in Mplus 7.11 (Muthén & Muthén, 2012) to conduct these analyses. Before analysis, we randomly partitioned the full sample ($N = 8{,}730$) into a development sample ($n = 4{,}365$) and a validation sample ($n = 4{,}365$). We used the development sample for initial model-building, tests of alternative factor structures, and measurement invariance tests. We used the validation sample to reanalyze the data and confirm results. Supplemental analyses indicated that the development and validation samples did not significantly differ across sociodemographic and substantive variables in the analysis.

We used the following model fit criteria to evaluate the acceptability of all analyzed models: root mean square error of approximation and its upper-bound 90% confidence interval $\leq .08$ (Browne & Cudeck, 1993), Tucker-Lewis Index (TLI) $\geq .95$, and comparative fit index (CFI) $\geq .95$ (Hu & Bentler, 1999). Because our samples were large, and chi-square difference tests are highly sensitive to sample size, we followed the admonition of Cheung and Rensvold (2002) and determined that model changes/constraints were statistically negligible if the change in the comparative fit index was smaller than or equal to $-0.01$ (i.e., $\Delta\text{CFI} \leq -0.01$). Although our data were ordinal, items with more than 10 response options cannot be specified as ordinal in Mplus. Thus, we used a maximum likelihood estimator. As a robustness check, we reanalyzed our final model with maximum likelihood estimation with robust standard errors and assessed any notable differences. Missing data (less than 4.2% across all indicators) were handled with full-information maximum likelihood estimation.

In terms of factor structure, we analyzed a first-order factor model in which four fitness constructs were specified for mental fitness, physical fitness, social fitness, and spiritual fitness. We used a jigsaw piecewise technique, whereby we assessed model fit and measurement parameters one construct at a time (Bollen, 2000). Because we expected the presence of an overall fitness construct, we then tested a second-order factor model in which the four first-order factors loaded onto a higher order construct. Because model difference tests indicated that both factor structures were statistically indistinguishable, we selected the second-order factor model and subjected it to the measurement invariance tests.

We adapted guidelines outlined by Chen, Sousa, and West (2005) to inform the measurement invariance testing process with our second-order factor model. Specifically, we assessed configural invariance (equivalent factor structure), first-order metric invariance (equivalent first-order factor loadings), second-order metric invariance (equivalent second-order factor loadings), and first-order scalar invariance (equivalent observed indicator intercepts) across all groups within a particular grouping (i.e., pay grade, gender, marital status, deployment). To obtain model identification and metric calibration, first- and second-order factor means and variances were fixed to 0 and 1, respectively. Preliminary calculations indicated that all analytical models were overidentified and sufficiently powered (N. Bowen & Guo, 2012; MacCallum, Browne, & Sugawara, 1996). No adjustments to the model were made that were not specified in the hypothesized model (e.g., no error covariances were specified).

## Results

### *Factor structure*

Table 5 displays the model fit indices associated with the first-order and second-order factor models. Results indicated that the second-order factor structure was statistically indistinguishable from the first-order factor structure (i.e., $\Delta\text{CFI} = -.001$), confirming our hypothesis that first-order fitness constructs can be conceptualized as part of a larger construct representing total or overall fitness. Model fit indices for the second-order factor model were $\chi^2(50) = 533.552$, $p < .001$, root mean square error of approximation (RMSEA) $= .047$ [upper-bound 90% CI: .051], TLI $= .982$, and CFI $= .986$, indicating acceptable model fit on the basis of our prespecified cutoff criteria.

### Measurement invariance tests

Table 5 also displays the model fit indices and ΔCFI associated with all measurement invariances tests. Beginning with invariance tests by pay grade, results indicated that configural (ΔCFI = –.003), first-order metric (ΔCFI = –.007), second-order metric (ΔCFI = –.001), and first-order scalar invariance (ΔCFI = –.008) could be specified without significantly worsening model fit. Model fit indices for the fully constrained measurement model by pay grade were $\chi^2(362) = 1539.655$, $p < .001$, RMSEA = .061 [upper-bound 90% CI: .065], TLI = .970, and CFI = .967, indicating acceptable model fit based on our prespecified cutoff criteria.

Results from invariance tests by gender indicated that configural (ΔCFI = .000), first-order metric (ΔCFI = –.003), second-order metric (ΔCFI = .002), and first-order scalar invariance (ΔCFI = –.002) could also be specified without significantly worsening model fit. Model fit indices for the fully constrained measurement model by gender were $\chi^2(128) = 735.096$, $p < .001$, RMSEA = .047 [upper-bound 90% CI: .050], TLI = .982, and CFI = .983, indicating acceptable model fit based on our prespecified cutoff criteria.

In terms of marital status, results from invariance tests indicated that configural (ΔCFI = –.001), first-order metric (ΔCFI = –.001), second-order metric (ΔCFI = –.001), and first-order scalar invariance (ΔCFI = –.004) could be specified without significantly worsening model fit. Model fit indices for the fully constrained measurement model by marital status were $\chi^2(128) = 865.840$, $p < .001$, RMSEA = .052 [upper-bound 90% CI: .055], TLI = .979, and CFI = .979, indicating acceptable model fit based on our prespecified cutoff criteria.

Finally, results from invariance tests by deployment experience indicated that configural (ΔCFI = .000), first-order metric (ΔCFI = .000), second-order metric (ΔCFI = .000), and first-order scalar invariance (ΔCFI = .000) could be specified without significantly worsening model fit. Model fit indices for the fully constrained measurement model by deployment experience were $\chi^2(128) = 629.870$, $p < .001$, RMSEA = .043 [upper-bound 90% CI: .046], TLI = .985, and CFI = .986, indicating acceptable model fit based on our prespecified cutoff criteria. All measurement invariance test results were confirmed from analyses conducted with the validation sample (results not shown in tables). Validation sample results are available upon request.

### Final model

Figure 2 displays the final second-order factor model with the full sample. All estimated measurement parameters were significant at the $p < .001$ level. Standardized first-

**Table 5.** Model-building and measurement invariance tests with development sample (N = 4,365).

| Model | *N* | Parameters | $\chi^2$ | *df* | *p* | RMSEA | Upper bound | TLI | CFI | ΔCFI | Comparison |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Factor structure[a] | | | | | | | | | | | |
| Model A: First-order factor structure | 4,344 | 42 | 506.890 | 48 | <.001 | 0.047 | 0.051 | 0.982 | 0.987 | | |
| Model B: Second-order factor structure | 4,344 | 40 | 533.552 | 50 | <.001 | 0.047 | 0.051 | 0.982 | 0.986 | −0.001 | Model A |
| Measurement invariance tests | | | | | | | | | | | |
| Pay grade (five groups)[b] | | | | | | | | | | | |
| Model 1A: Configural invariance | 4,301 | 200 | 844.569 | 250 | <.001 | 0.053 | 0.057 | 0.978 | 0.983 | −0.003 | Model B |
| Model 2A: Invariance of first-order factor loadings | 4,301 | 152 | 1132.682 | 298 | <.001 | 0.057 | 0.061 | 0.974 | 0.976 | −0.007 | Model 1A |
| Model 3A: Invariance of second-order factor loadings | 4,301 | 136 | 1196.294 | 314 | <.001 | 0.057 | 0.061 | 0.974 | 0.975 | −0.001 | Model 2A |
| Model 4A: Invariance of observed indicator intercepts | 4,301 | 88 | 1539.655 | 362 | <.001 | 0.061 | 0.065 | 0.970 | 0.967 | −0.008 | Model 3A |
| Gender (two groups)[c] | | | | | | | | | | | |
| Model 1B: Configural invariance | 4,323 | 80 | 607.181 | 100 | <.001 | 0.048 | 0.052 | 0.981 | 0.986 | 0.000 | Model B |
| Model 2B: Invariance of first-order factor loadings | 4,323 | 68 | 630.456 | 112 | <.001 | 0.046 | 0.050 | 0.985 | 0.983 | −0.003 | Model 1B |
| Model 3B: Invariance of second-order factor loadings | 4,323 | 64 | 645.385 | 116 | <.001 | 0.046 | 0.049 | 0.983 | 0.985 | 0.002 | Model 2B |
| Model 4B: Invariance of observed indicator intercepts | 4,323 | 52 | 735.096 | 128 | <.001 | 0.047 | 0.050 | 0.982 | 0.983 | −0.002 | Model 3B |
| Marital status (two groups)[d] | | | | | | | | | | | |
| Model 1C: Configural invariance | 4,344 | 80 | 633.525 | 100 | <.001 | 0.050 | 0.053 | 0.980 | 0.985 | −0.001 | Model B |
| Model 2C: Invariance of first-order factor loadings | 4,344 | 68 | 683.503 | 112 | <.001 | 0.048 | 0.052 | 0.981 | 0.984 | −0.001 | Model 1C |
| Model 3C: Invariance of second-order factor loadings | 4,344 | 64 | 701.190 | 116 | <.001 | 0.048 | 0.052 | 0.981 | 0.983 | −0.001 | Model 2C |
| Model 4C: Invariance of observed indicator intercepts | 4,344 | 52 | 865.840 | 128 | <.001 | 0.052 | 0.055 | 0.979 | 0.979 | −0.004 | Model 3C |
| Deployment (two groups)[e] | | | | | | | | | | | |
| Model 1D: Configural invariance | 4,320 | 80 | 586.849 | 100 | <.001 | 0.047 | 0.051 | 0.982 | 0.986 | 0.000 | Model B |
| Model 2D: Invariance of first-order factor loadings | 4,320 | 68 | 612.332 | 112 | <.001 | 0.045 | 0.049 | 0.983 | 0.986 | 0.000 | Model 1D |
| Model 3D: Invariance of second-order factor loadings | 4,320 | 64 | 621.225 | 116 | <.001 | 0.045 | 0.048 | 0.984 | 0.986 | 0.000 | Model 2D |
| Model 4D: Invariance of observed indicator intercepts | 4,320 | 52 | 629.870 | 128 | <.001 | 0.043 | 0.046 | 0.985 | 0.986 | 0.000 | Model 3D |

*Note.* [a]21 cases are omitted due to missing values on all variables. [b]Pay grade subgroups: E1-E4 (N = 1,354), E5-E6 (N = 1,427), E7-E9 (N = 589), O1-O3 (N = 476), O4+ (N = 455). [c]Gender subgroups: Female (N = 916), Male (N = 3407). [d]Marital status subgroups: Married (N = 2,563), Not Married (N = 1,781). [e]Deployment subgroups: Not Deployed in Past 12 Months (N = 3,343), Deployed in Past 12 Months (N = 977). Invariance tests were conducted as outlined in Chen, Sousa, and West (2005). For the purposes of model identification and metric calibration, first-and second order factor intercepts were fixed to 0 across all groups and first- and second-order variances were fixed to 1 across all groups from the onset of measurement invariance tests. The number of cases in each set of invariances tests varies due to missing values associated with the grouping variable.

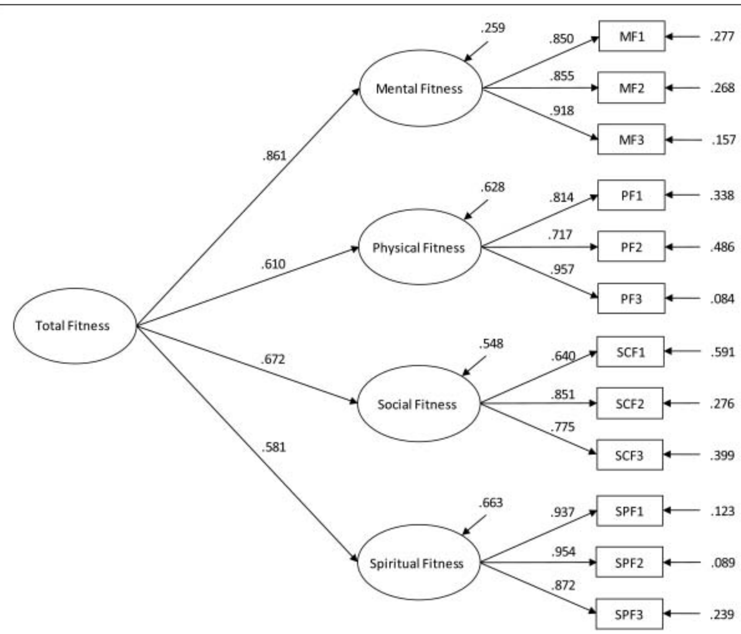
order factor loadings ranged from .640 to .957. Standardized second-order factor loadings ranged from .581 to .861. As mentioned previously, the final model was reanalyzed with maximum likelihood estimation with robust standard errors as a robustness check. The results were identical to those estimated with maximum likelihood. Refer to Figure 2 for more details regarding the final model.

## Discussion and implications

Our results indicate that the four components of airman fitness (mental, physical, social, and spiritual) can be conceptualized as pieces of a total or comprehensive fitness construct. Our results also indicate that this comprehensive measure of airman fitness reliably captures the same phenomena for (a) members of all pay grades, (b) men and women, (c) those who are married and those who are not, and (d) those who have been deployed in the past 12 months and those who have not. Thus, our results suggest that the instrument is particularly robust for active-duty members, and shows significant promise as a Comprehensive Airman Fitness metric derived from an existing community-based Air Force sponsored survey.

These results have important implications for monitoring and evaluating the implementation of the Comprehensive Airman Fitness. In discussing the intersections across the domains of total force fitness (e.g., "physical fitness is enhanced by psychological factors"), Jonas and colleagues (2010) noted the need to move beyond a "siloed approach to components of fitness" (p. 12). Although the authors did not go so far as to suggest a total fitness score, they did suggest the need for "a comprehensive set of measures of success and [their use] in an integrated fashion for continual process improvements" (p. 12).

As the Comprehensive Airman Fitness specialist and facilitator at the installation level, as well as the Community Action Information Board Executive Director, Integrated Delivery System Chair, and Caring for People Coordinator (Air Force Instruction 90–506, April 2, 2014, p. 7) the community support coordinator could consult with unit commanders to administer either the Support and Resiliency Inventory in full or a shortened version of the assessment tool, which includes the 12-item measure of fitness, to profile the fitness of their respective units. It would be especially appealing to develop a one-page summary report that displays a total fitness score, as well as the four-component fitness scores. These scores could be further displayed for demographic subgroups, such as by pay grade, gender, marital status, and deployment during the past 12 months. As




*Note:* Final model fit indices: $\chi^2(50) = 860.605$, $p < .001$, RMSEA = .043 [90% CI: .041 - .046], TLI = .985, CFI = .989. Maximum Likelihood estimator was used for the analysis. All estimated parameters are standardized and significant at the $p < .001$ level. Fifty cases were omitted from the analysis because they were missing values on all observed indicators.

**Figure 2.** Final second-order factor model with the full sample ($N = 8,680$).

suggested by Jonas and colleagues (2010), a highly informative graphic could be developed for visualizing "strengths and gaps in fitness," as well as "areas for improvement" (p. 12). In time, data from a representative sample of Air Force active duty members could be used to develop comparison norms, including norms for multiple subgroups like pay grade, gender, and job functions.

Assuming a web-based delivery system for administrating the measure, whether the full or an abbreviated version of the current Support and Resiliency Inventory, respondents could be offered the opportunity to view a copy of their fitness profile, including the ability to print or email themselves a copy of their results. The inventory's online delivery system already has these features, including the ability to be administered on a handheld device. Air Force members interested in examining their own fitness and developing strategies for promoting it could also use the 12-item fitness measure as a short, self-assessment tool. Evidence-based tips and strategies for promoting fitness could be added to the online delivery system for informing the development of a personal fitness plan. Heyman and colleagues' (2015) development of evidence-based action sheets for active duty Air Force members, which are focused on dealing with a variety of relationship issues with spouses or significant romantic partners, is an excellent model to follow. In time and after further validations of the measure, which are subsequently discussed, it would be possible to develop a cell phone or other hand-held digital application of the measure, which would promote the ability of airmen to monitor their own fitness.

## Limitations and implications for future research

In the present investigation, we used a nonprobability sample. Non-probability samples are limited in their ability to produce fully generalizable findings. To the extent possible, future investigations should incorporate representative samples of Air Force members to increase the external validity of estimated parameters. The current sample was also restricted to only active duty Air Force members serving in the regular component and excluded those serving in the Air National Guard and the Air Force Reserve Component, or as Air Force civilian employees. Although recently published research indicates that the Comprehensive Airman Fitness measure is invariant for active duty Air Force members, members of the Air National Guard/Air Force Reserve, and Air Force civilian employees (Bowen, Jensen, & Martin, 2016), further research should examine the psychometric performance of the Comprehensive Airman Fitness instrument among members of other population subgroups, such as the family members of Air Force members, who are included in the broad definition of "airman" as specified in the Air Force Instruction 90-506.

Further research is also needed to acquire additional evidence of the criterion-related and construct validity of the Comprehensive Airman Fitness instrument (DeVellis, 2012). First, future studies should examine temporally neutral correlations that are expected to exist between the instrument and other related variables (i.e., criterion-related validity). Such variables may include health-related outcomes (e.g., good sleep quality and social participation in unit and community-based events and activities) and other instruments that purport to measure one or more features of individual fitness.

Second, studies should examine temporally neutral correlations that are not expected to exist between the Comprehensive Airman Fitness instrument and other related variables (i.e., discriminant validity). Although examples of discriminant validity are more difficult to identify in the context of the broad and integrative nature of the fitness concept, such examples may include relatively stable personality traits (e.g., extraversion vs. introversion, conscientiousness vs. spontaneous).

Third, researchers should analyze directional associations between the Comprehensive Airman Fitness instrument and theoretically relevant constructs (i.e., construct validity). For example, Land (2010), in discussing the Department of Defense's flexibility in the way in which the various service components implement Comprehensive Airman Fitness, notes that the Air Force established its model on a human performance framework. Consequently, we would expect the total fitness measure to predict successful role performance in meeting duty and personal responsibilities. Bowen and colleagues (2016), in the same analysis referenced above, found support for this expectation in reporting a strong and positive association between the current Comprehensive Airman Fitness measure and a measure of resiliency, which included three items related to the level of success in meeting the challenges of military life, performing assigned duties, and meeting overall responsibilities associated with personal and family roles among active duty personnel, members of the Air National Guard/Air Force Reserve, and Air Force civilian employees,. However, the cross-sectional nature of the study design restricts the ability to rule out competing explanations for this relationship (e.g., common methods variance). Longitudinal designs are needed to determine the temporal order between the current measure and hypothesized outcomes—a necessary condition to bolster causal inference.

Future studies should explore the extent to which the Comprehensive Airman Fitness instrument successfully distinguishes between members of groups for which differences in scores are expected (i.e., known-groups validity). For example, known-groups validation could be explored by examining how the Comprehensive Airman Fitness instrument scores individuals differentially (significant mean differences) based on factors like rank and duty position that affect the degree to which a service member has control over their work day, or the inherent demands in marriages among young service members, especially in the context of early family formation and the demands of childcare. Beyond job control and family demands, problem behavior status warrants examination (e.g., cited for driving under the influence, established perpetrator of family maltreatment, early return from a deployment for violations of military policy or for problem behavior) as well as the degree to which service members and their families have experienced trauma associated with military duty and service life. Taken together, these tests would help reveal the extent to which the Comprehensive Airman Fitness instrument measures what it purports to measure (DeVellis, 2012).

## References


Air Force Instruction 90–501. (2013, October 15). *Community Action Information Board (CAIB) and Integrated Delivery System (IDS)*. Washington, DC: Department of the Air Force.

Air Force Instruction 90-506. (2014, April 2). *Comprehensive Airman Fitness (CAF)*. Washington, DC: Department of the Air Force.

Bollen, K. A. (2000). Modeling strategies: In search of the holy grail. *Structural Equation Modeling: A Multidisciplinary Journal*, *7*(1), 74–81.

Bowen, G. L., Jensen, T. M., & Martin, J. A. (2016). A measure of comprehensive airman fitness: Construct validation and invariance across Air Force service components. *Military Behavioral Health*, *4*, 149–158.

Bowen, G. L., Jensen, T. M., Martin, J. A., & Mancini, J. A. (2016). The willingness of military members to seek help: The role of social involvement and social responsibility. *American Journal of Community Psychology*, *57*, 203–215.

Bowen, G. L., Mancini, J. A., Martin, J. A., Ware, W. B., & Nelson, J. P. (2003). Promoting the adaptation of military families: An empirical test of a community practice model. *Family Relations*, *52*, 33–44.

Bowen, G. L., & Martin, J. A. (2011). *The Support & Resiliency Inventory* (Version 3, SRI-M). Chapel Hill, NC: Bowen & Colleagues.

Bowen, G. L., & Martin, J. A. (2013a). *Marine Corps Fitness Improvement Tool (MCFIT)*. Charlotte, NC: Flying Bridge Technologies.

Bowen, G. L., & Martin, J. A. (2013b). *Support and Resiliency Inventory (SRI-M): Six-month utilization report (January–June 2013)*. Charlotte, NC: Flying Bridge Technologies.

Bowen, G. L., Martin, J. A., Liston, B. J., & Nelson, J. P. (2009). Building community capacity in the U.S. Air Force: The Community Readiness Consultant Model. In A. R. Roberts (Ed.), *Social workers' desk reference* (2nd ed., pp. 912–917). New York, NY: Oxford University Press.

Bowen, N., & Guo, S. (2012). *Structural equation modeling*. New York, NY: Oxford University Press.

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

Casey, G. W., Jr. (2011). Comprehensive Soldier Fitness: A vision for psychological resilience in the U.S. Army. *American Psychologist*, *66*, 1–3.

Cantril, H. (1965). *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 471–492.

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

Dawes, J. G. (2002). Five-point vs. eleven-point scales: Does it make a difference to data characteristics? *Australasian Journal of Market Research*, *10*(1), 39–47.

Department of the Air Force, Office of the Chief of Staff. (2012, January 12). *Talking points for individual and group discussion during resiliency stand down*. Washington, DC: Author.

Department of Defense, Office of the Deputy Assistant Secretary of Defense (Military Community and Family Policy). (2012). *2012 Demographics: Profile of the military community*. Washington, DC: Author.

DeVellis, R. (2012). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage.

Fraser, M. W., & Galinsky, M. J. (2010). Steps in intervention research: Designing and developing social programs. *Research on Social Work Practice*, *20*, 459–466.

Gonzalez, G. C., Singh, R., Schell, T. L., & Weinick, R. M. (2014). *An evaluation of the implementation and perceived utility of the airman resilience training program*. Santa Monica, CA: RAND Corporation.

Hamaoka, D., Bates, M. J., McCarroll, J. E., Brim, W. L., Lunasco, T. K., & Rhodes, J. E. (2014). An introduction to military service. In S. J. Cozza, M. N. Goldenberg, & R. J. Ursano (Eds.), *Care of military service members, veterans, and their families* (pp. 3–21). Washington, DC: American Psychiatric Publishing.

Heyman, R. E., Smith Slep, A. M., Sabathne, C., Eckardt Erlander, A. C., Hsu, T. T., Snyder, D. K., … Sonnek, S. M. (2015). Development of a multilevel prevention program for improved relationship functioning in active duty military members. *Military Medicine*, *180*, 690–696.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Huebner, A. J., Mancini, J. A., Bowen, G. L., & Orthner, D. K. (2009). Shadowed by war: Building community capacity to support military families. *Family Relations*, *58*, 216–228.

Institute for Alternative Futures. (2009, December 30). *Total fitness for the 21st century: Conference report*. Alexandria, VA: Author.

Jonas, W. B., O'Connor, F. G., Deuster, P., Peck, J., Shake, C., & Frost, S. S. (2010). Why total force fitness. *Military Medicine*, *175*(8), 6–13.

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

Land, B. C. (2010). Current Department of Defense guidance for total force fitness. *Military Medicine*, *175*(August Supplement), 3–5.

Lozano, L. M., Garcia-Cueto, E., & Muniz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, *4*(2), 73–79.

Lundquist, J., & Xu, Z. (2014). Reinstitutionalizing families: Life course policy and marriage in the military. *Journal of Marriage and Family*, *76*, 1063–1081.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149.

Meadows, S. O., Miller, L. L., & Robson, S. (2015). *Airman and family resilience: Lessons from the scientific literature* (RR-106-AF), Santa Monica, CA: RAND Corporation.

Mullen, E. J. (2004). Outcome measurement: A social work framework for health and mental health policy and practice. *Social Work in Mental Health*, *2*(2–3), 77–93.

Mullen, M. (2010). On total force fitness in war and peace. *Military Medicine*, *175*(8), 1–2.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Authors.

Orthner, D., Bowen, G., & Mancini, D. (2003, December). *The community readiness unit service guide for Air Force Space Command Family Support Centers*. Colorado Springs, CO: U.S. Air Force Space Command Family Matters.

Peterson, C., Park, N., & Castro, C. A. (2011). Assessment for the US Army Comprehensive Soldier Fitness program: The Global Assessment Tool. *American Psychologist*, *66*(1), 10–18.

Redmond, S. A., Wilcox, S. L., Campbell, S., Kim, A., Finney, K., Barr, K., & Hassan, A. M. (2015). A brief introduction to the military workplace culture. *Work: A Journal of Prevention, Assessment and Rehabilitation*, *50*, 9–20.

StataCorp. (2013). *Stata statistical software: Release 13*. College Station, TX: Author.