SUPPORTING STATEMENT – PART B

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS

1. Description of the Activity

The research project will develop and validate a measure of Spiritual Fitness that is consistent with the Chairman of the Joint Chiefs of Staff Instruction (CJCSI) definition of spirituality. The CJCSI definition is, "Spiritual Fitness refers to the ability to adhere to beliefs, principles, or values needed to persevere and prevail in accomplishing missions" (CJCSI 3405.01, A-2). Improving Spiritual Fitness could improve resiliency if there is a stressful life event, and potentially reduce the prevalence of posttraumatic stress disorder. This could benefit all Americans. To develop the measure, we will sample from Amazon's Mechanical Turk (mTurk) respondents who reside in the United States. We anticipate that mTurk respondents reflect aspects of the U.S. population more broadly since they are a subset of this population. There will be no active recruitment of any particular segment of the population through mTurk.

To determine burden for just the civilian population, we had the following assumptions. Based on national data, approximately 2% of the U.S. population are Active Duty Service Members (see OMB Statement A). We are not gathering a military subsample. We are identifying respondents from the general population who have prior or current military experience to interpret the responses in the appropriate context. This is to ensure data quality through contextualizing responses. The breakdown of the civilian burden can be seen on Appendix G (tab name "Sample Size and Sampling Method"], provided separately). Appendix D (provided separately) contains the burden compensation details.

According to published research with mTurk samples, there is a higher rate of self-identified atheists, compared to the general US population. Therefore, we anticipate that approximately 15% of the mTurk sample will self-identify as atheist as this is the mid-point based on the range of 6-24% cited in recent literature (Casey, Chandler, Levine, Proctor, & Strolovitch, 2017; Exline, Levay, Freese, & Druckman, 2016; Pargament, Grubbs, & Yali, 2014).

In the demographically representative samplings, collected through Qualtrics, the percentages of self-identified atheists are lower than those reported based on the estimates in mTurk. In the demographically representative sampling for this project, we made the following assumptions. Reliable data on atheists from the Pew Research Center (Lipka, 2016) found that 3% of Americans identified as atheist. According to the military data published in Christianity Today (Smietana, 2015), there were 12,764 atheists among the 1,300,000 members of the military, i.e., <1%. Service Member participants and atheist participants will receive tailored surveys with specific questions that are appropriate for these subsets of the population. These responses will be pooled across the first 10 samples.

Samples 11-14 have branching logic for military and civilian participants and no branching logic for atheists. The demographics and outcome indictors will be held constant across samples, so comparisons are possible. The developed measure, without branching logic, will be validated on two nationally representative samples conducted by Qualtrics. Since both mTurk and Qualtrics have been contracted for sampling, they will deliver the number of responses that we determined are necessary for validating this metric.

To ensure data quality, we estimate approximately 15% of the responses will be removed for cases that do not pass the embedded data validation checks. To account for this, an additional 15% of responses have been incorporated into our sample estimates (see OMB Statement A and Appendix G [tab name "Sample Size and Sampling Method"], provided separately).

2. Procedures for Collection of Information

   a. Statistical methodologies for stratification and sample selection;

All procedures are geared towards measure development, specifically measure reliability and validity. Sample #1 through Sample #14 utilize a convenience sample of mTurk respondents who complete the survey for a flat rate compensation.

Two samples will be collected by Qualtrics. Qualtrics has been contracted to recruit participants for two samples. Qualtrics respondents are compensated under a variable compensation rate determined by Qualtrics. Using Qualtrics, these samples are stratified based on three factors: age, gender, and head of household income.

Appendix G (tab name "Sample Size and Sampling Method", provided separately) denotes the sample size and sampling method for each of the 16 samples.

   b. Estimation procedures;
      a. Maximum likelihood
      b. Least squares
      c. Principle components analysis (PCA)
      d. Principal Axis factoring (PAF)
      e. Bayesian estimation methods

All estimation procedures are geared towards measure development, specifically measure reliability and validity. The estimation methods that this project will use depends on the whether the data meets the underlying assumptions for both the statistical test and the estimation method. We will be using the estimation method and test that is most appropriate. Maximum likelihood estimation will be potentially used for missing data imputation, logistic regression, and confirmatory factor analysis and structural equation modeling. If certain assumptions of the data are not met for analyses that use Maximum likelihood, alternative estimations like unweighted least squares, weighted least squares, etc. will be used. Least squares estimation may also be used in regression analyses for convergent validity and missing data imputation. PCA and PAF estimation methods will be used for data reduction techniques and assessment of underlying structure (respectively). Bayesian estimation methods will be utilized in instances where the data are found to be non-normal or have small sample sizes (Kruschke, 2014).

   c. Degree of accuracy needed for the purpose discussed in the justification;

All test development techniques and statistics are geared towards measure development, specifically reliability and validation. We utilized recommended sample sizes for the various statistical analyses that were selected. Prior effect sizes for power analysis are not available because no measure exists that has captured the broad aspects of Spiritual Fitness used in this project. Conservative benchmarks for analyses that provide stable estimates and adequate statistical power are used.

Using a planned missing data (PMD) design, we selected a minimum sample size of 250 respondents per item pair in each of the test item pools (Samples 1, 3, 4, 6). This provides stable correlation coefficients (Schönbrodt, & Perugini, 2013) that will be used as input data on subsequent analyses (i.e., factor analysis, principle components analysis [PCA], item response theory [IRT] and multi-dimensional IRT [MIRT], and confirmatory factor analyses [CFA]).

We assume that 15% of respondents self-nominate as atheist in mTurk. An additional 2% of respondents will identify as a Service Member. These individuals will receive a tailored test item pool through branching logic because of their specific values and beliefs. We have increased our required sample sizes to maintain adequate power and stability. These percentages, along with a potential 15% who do not meet our embedded data validation checks, were factored into our numbers listed in OMB Statement A. Note that the Statement A participant numbers represent only civilians. For the later samples, we use sample sizes that will meet the minimum sample sizes necessary to provide stable estimates (Appendix G, tab name "Statistical Tests by Sample", provided separately).

If the pooled sample sub-groups are smaller than needed for null hypothesis statistical testing, Bayesian estimation analogues of null hypothesis statistical tests will be utilized. Bayesian estimation analyses perform well with small sample sizes, and the credibility of the estimates improve with additional data (Kruschke, 2014). Bayesian estimation requires information on the prior knowledge about the data as part of the analysis. An uninformative prior (the potential values are dispersed across a large amount of potential credible values) will be used in the early samples. The results of these early samples will then be used as prior information for subsequent Bayesian estimations, since the prior would better reflect what is known about the distribution of credible values (Gelman et al., 2014; Kruschke, 2014).

d. Unusual problems requiring specialized sampling procedures;

An unusual problem with the proposed project is the large number of test items in the pool for the measure development. Traditional measurement development studies require responses on every item for each participant, which creates too great a burden. Alternatively, dividing the items into different groups would make it so there would not be correlations for each item set, which introduces bias into the results of any factor analyses because the results are dependent on the items used.

To reduce the risk of potential bias and reduce participant burden, we divided the test item pool into four pools according to our theoretical framework. A panel of subject matter experts classified items into one of the following Spiritual Fitness categories: 1) vertical spirituality items (connection of self and higher power(s)), 2) horizontal spiritual items (connection of the self and others), 3) mixed vertical items (predominantly vertical with some aspects of horizontal), and 4) mixed horizontal items (predominantly horizonal with some aspects of vertical). Each participant will see a random subset of items from each of the four test item pools for four of the samples.

There will be correlations for each test item pair when the overall correlational matrix for each sample is viewed. This is a variation of PMD (Enders, 2010; Graham, Taylor, Olchowski, & Cumsille, 2006). By reducing the number of items seen by each participant, participant burden is reduced. This will maximize response rates and keep the data quality high.

Across all the responses, there will be enough data to determine which items are more strongly related to Spiritual Fitness outcomes at the sample level. Through data reduction techniques, items that are not strongly related to the outcomes will be removed, further reducing the number of potential items from

the test pool. The resulting variance/covariance matrix will be used as input for subsequent data analyses to maintain the relationships between the items. This way there is no need to impute any missing values since the missing values have been planned a priori. A traditional measurement approach will be utilized once the test item pools have been reduced by removing poor items (e.g., the removal of items with low factor loadings or non-significant relationships to outcome indicators).

      e.   Use of periodic or cyclical data collections to reduce participant burden.

Periodic or cyclical data collections do not apply since we are utilizing the PMD methodology to reduce the number of test items from the pools. We are utilizing PMD in Samples #1, 3, 4, and 6 to limit the number of test items from the four pools. The number of demographic, religious orientation, and validation items in the first 9 samples are held constant so that each survey is approximately 135 items to complete.

3.   Maximization of Response Rates, Non-response, and Reliability

To maximize response rates, we are using PMD methodology (see above discussion on PMD) to provide a shorter survey. We do not anticipate non-responses at the case level, due to mTurk respondents self-selecting into the study until we reach the anticipated sample size. Additionally, Qualtrics has been contacted to obtain 2,000 respondents for the demographically representative samples (1,000 respondents each). Qualtrics guarantees they will meet the required number of completed valid surveys. Survey responses that do not meet completion criteria are removed from the response count and additional responses are obtained until the target number of responses is met. Survey response rates will be requested from Qualtrics.

The specific missing data approach will be determined based on an inspection of the amounts, patterns of missing data, and the underlying assumptions of the statistical analyses (Enders, 2010). For missing values on the non-test item pool items in the surveys, we will potentially utilize one of the following:
- Regression based imputation (to impute values based on relationships of other items as predictors)
- Maximum likelihood (to impute values based on relationships of other items)
- Mean imputation (if minimal data missing)
- Ipsative (for full scale composition)
- Full information maximum likelihood (in the case for CFA)

To ensure that all the measures are reliable, the internal consistency for each measure will be examined at each sample (Sample #1 to Sample #16). This allows increased confidence that our measures are consistent with each sample. To further examine the internal consistency, we will compute the 95% confidence interval around each estimate.

The first 14 samples utilize a convenience sample of U.S. located mTurk respondents, who are a subset of the U.S. population. The final two samples are demographically representative samples by age, gender, and head of household income (not randomly sampled or weighted). Qualtrics acts as a panel aggregator, building samples from multiple sources using Grand Mean certified sample partners. All sample partners redirect members by matching qualifying demographic information from their profiles to a specific survey. Potential respondents build their profile from a standardized list of questions. The panels then use the profiles to select studies that would best fit the case specifications. Qualtrics panel partners randomly select respondents for surveys where respondents are highly likely to qualify. Each

sample from the panel base is proportioned to the general population and then randomized before the survey is deployed. To exclude duplication and ensure validity, Qualtrics checks every IP address and uses a sophisticated digital fingerprinting technology. Each panel has its own confirmation procedures including, but not limited to: TrueSample, Verity, SmartSample, USPS verification, and digital fingerprinting. All panel partners verify respondent address, demographic information, and email address. Potential respondents are sent an email invitation informing them that the survey is for research purposes only, how long the survey is expected to take and what incentives are available. Members may unsubscribe at any time. To avoid self-selection bias, the survey invitation does not include specific details about the contents of the survey. Respondents receive an incentive based on the length of the survey, their specific panelist profile and target acquisition difficulty. Qualtrics replaces respondents who straight-line through surveys or finish in less than 1/3 of the average survey completion length.

Standard guidelines dictate that researchers should avoid nonprobability online panels when one of the research objectives is to accurately estimate population values. There currently is no generally accepted theoretical basis from which to claim that survey results using samples from nonprobability online panels are projectable to the general population. However, because the goal of this research is to develop and validate a measurement construct rather than to generate nationally representative estimates, the research team believes that the proposed methods are fit for the intended research purposes. Since the underlying demographics of the military are constantly changing, we believe that it is optimal to administer the survey to a group that is demographically similar to the general population, rather than a group that is demographically similar to the military. By administering the final survey to a demographically representative group, we can begin to research normative estimates and cutoffs, which will assist in interpreting scores derived from the survey when it is used. Findings will not be described as nationally representative.

4. Tests of Procedures

Below is a list of potential statistical tests that may be utilized in the project. Which specific tests are utilized will depend on the nature of the data and whether the underlying statistical test assumptions are met. If the data does not meet the statistical assumptions of the tests below, other options will be investigated. See Appendix G (tab name "Statistical Tests By Sample", provided separately) for a list of tests for each sample.

a. Correlations to assess the relationships of the test items with outcome indicators, input for other statistical analyses (e.g., factor analysis, PCA, and CFA), and convergent and discriminant validity during the construct validation samples (Samples #1-14).
b. Principle Components Analysis (PCA) to reduce the number of test items into the smallest number of components.
c. Factor analyses to assess the factor structure for the pools (Samples 1-7) and the overall Spiritual Fitness measure factor structures for replication of the underlying structure (Samples #11-16).
d. Item response theory (IRT)/multidimensional item response theory (MIRT)/Bayesian item response theory will be utilized to assess how the test items perform. Depending on the nature of the data and whether certain statistical assumptions are met, different variations of IRT will be utilized (i.e., MIRT or Bayesian IRT).
e. Confirmatory factor analyses (CFA) will be utilized to assess whether the factor structure is replicated.

      f.     Multiple regression may be utilized as a missing data imputation strategy. It will also be used to assess the relationship of the Spiritual Fitness measure with other correlates of Spiritual Fitness to explain variance in the validation outcomes.

      g.     Latent profile analysis to determine underlying profiles of Spiritual Fitness.

      h.     Structural equation modeling will be utilized to better understand the relationships between the validated Spiritual Fitness measure and outcomes related to spiritual performance.

      i.     Bayesian estimation methods will be utilized in instances where that are small samples sizes or when information on the credibility of the estimates would be useful.

5.     Statistical Consultation and Information Analysis

      a.     Names and telephone numbers of the individuals consulted on the statistical aspects of the design:

           Patricia Deuster, Ph.D., MPH, FACSM (301) 295-3020
           Josh Kazman, M.S. (301) 295-9251
           Harold Koenig, M.D. (919) 681-6633
           Ralph Hood, Ph.D. (423) 425-4274
           Christopher Silver, Ph.D. (423) 425-5720

      b.     Name and organization of persons who will actually collect and analyze the collected information:

           Eric R. Schuler, Ph.D. (301) 319-6995
           Kathleen G. Charters, Ph.D., RN (301) 295-9250
           Ian Gutierrez, Ph.D. (301) 295-1362
           Josh Kazman, M.S. (301) 295-9251

References:

Casey, L. S., Chandler, J., Levine, A. S., Proctor, A., & Strolovitch, D. Z. (2017). Intertemporal differences among MTurk workers: Time-based sample variations and implications for online data collection. SAGE Open, 7(2), 2158244017712774.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Exline, J.J., Pargament, K.I., Grubbs, J.B., & Yali, A.M. (2014). The religious and spiritual struggles scale: Development and initial validation. Psychology of Religion and Spirituality, 6, 208-222.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis (Vol. 2)*. Boca Raton, FL: CRC Press.

Graham, J. W., Taylor, B. J., Olchowski, A. E., &amp; Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, 11(4), 323.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. Sage Open, 6(1), 2158244016636433.

Lipka, M. (2016, June). 10 facts about atheists. PEW Research Center. Retrieved from: http://www.pewresearch.org/fact-tank/2016/06/01/10-facts-about-atheists/.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5), 609-612.

Smietana, B. (2015, April 23). Atheists outnumber Southern Baptists in U.S. military. *Christianity Today*. Retrieved from: http://www.christianitytoday.com/news/2015/april/atheists-outnumber-southern-baptists-in-us-military.html.