TERENCE J. BOSTIC, DORIS McGARTLAND RUBIO and MARK HOOD

A VALIDATION OF THE SUBJECTIVE VITALITY SCALE USING STRUCTURAL EQUATION MODELING

(Accepted 13 June, 2000)

ABSTRACT. Ryan and Frederick's (1997) measure of vitality, the subjective feeling of being alive and alert, was developed in the context of a single factor analysis. The present investigation employed structural equation modeling (SEM) to assess construct validity and utility of the new measure. A large sample (N = 526) was collected in two waves, allowing the investigators to further develop the model proposed by Ryan and Frederick, and to then validate it on a second data set. The final model is presented, and the process of achieving that model is discussed, as are the relative strengths of SEM in test development.

Subjective vitality was recently defined and measured by Ryan and Frederick (1997) as the subjective experience of being full of energy and alive. Those high in subjective vitality report being alert, energized, and vital. The concept is called by many names in different cultures. The Taoist culture of ancient China refer to ch'i and jing as the feeling of being full of internal energy (Liao, 1990), while the ancient Japanese refer to Ki as the energy and power one can call upon to mobilize mental and physical health (Ryan and Frederick, 1997). McNair, Lorr, and Droppleman (1971), used the similar term "vigor" to describe a positive mood state in their measure, the Profile of Mood States (POMS). Ryan and Frederick (1997) further describe vitality as "energy that is perceived to emanate from the self" (p. 535). It is a feeling of energy from an internal source and not from specific threats in the environment. Vitality differs from mania in that vitality is defined as feeling alive and energized, not driven or compelled. Indeed, the construct of vitality appears to be well recognized, if differently named by various cultures.

Ryan and Frederick (1997) designed a measure of this construct as they noticed that while the feeling of being alive and energized is familiar and notable, it is not often investigated in the scientific literature. The authors wanted to measure a sense of spirit and enthusiasm as a hypothesized reflection of physical and psychological well-being. In their initial investigation, Ryan and colleagues found the scale to have Cronbach's alpha = 0.84. The questions are rated on a seven point scale from "not at all true" to "very true." The seven items in the scale are from an original pool of 19 items. Of those 19 items, three were eliminated due to poor variability or poor content. The remaining items loaded on two factors from an oblique rotation. The first factor indicated the items related to vitality, and had an eigenvalue = 6.77, and a coefficient alpha = 0.84. The second factor was related to having goals and purpose, and those items were dropped from the measure. According to the authors, the seven items in the final scale "were seen as reflecting, from a content perspective, an adequate definition of a phenomenological sense of aliveness and vitality, and were thus summed ..." (p. 540) (Appendix). The seven-item scale was shown to have positive correlations with measures of self-actualization, self-esteem, and satisfaction with life (correlations ranging from 0.42 to 0.76) and negative correlations with measures of psychopathology, depression, negative affect, and anxiety (correlations ranging from –0.25 to –0.60) (Ryan and Frederick, 1997). Currently, there is no published test-retest data on populations that were not subject to experimental manipulations.

As the literature has recently put forward an innovative scale to measure this construct, affirming the construct validity is crucial before using it as the benchmark for further investigations. Researchers, such as the original authors of the subjective vitality measures, often rely on two separate techniques to determine the soundness of their instrument: Cronbach's alpha and factor analysis. Cronbach's alpha is a measure of internal consistency; it is the mean of inter-item correlation of all the items in the measure (Nunnally and Bernstien, 1994). It does not provide any data on how well the items measure a construct, as it is only a statistic reflecting how similarly participants responded to all the items as a whole. Furthermore, coefficient alpha tends to underestimate the measure's true reliability (Miller, 1995). Factor analysis provides a better understanding of which variables form a "relatively coherent subset, independent of others" (Tabachnik and Fidel, 1996: p. 635). The analysis is similar to Cronbach's alpha in that it determines which variables are correlated with one group, and not correlated with other groups.

Variables that are shown to be related to their own group and not to other groups are thought to define factors, that is, latent constructs. Factor analysis, provides not only which variables "hang together" as a whole, but how well each item contributes to the description of the proposed construct. Ryan and Frederick (1997) completed both of these analyses competently and completely on large data sets, and were able to duplicate their findings. However, there are some inherent limitations in these techniques.

Cronbach's alpha and factor analysis are data driven techniques similar in that they use larger data sets to define factors by eliminating dissimilar items (Tabachnik and Fidel, 1996). Both of these techniques have some inherit limitations. Cronbach himself pointed out that test length will effect the measure named in his honor, causing some difficulty in understanding alpha:

> Conceptually, it seems as if 'homogeneity' or 'internal consistency' of a test should be independent of it's length. A gallon of homogenized milk is no more homogenous than a quart. Alpha increases as the test is lengthened (Cronbach, 1951: p. 323).

Cronbach's alpha also presupposes that all items are tau-equivalents, that is, they all have equal loadings on a single common factor with their unique variances composed entirely of error, an often overlooked and infrequently met condition (Miller, 1995).

Exploratory factor analysis (EFA) relies on the assumption that some set of variables (usually fewer than the total number of observed variables) comprises a good measure of an underlying construct. EFA allows the analysis program to use the data to determine what factor structure explains most of the variance in the data. Again, some limitations are worth noting with this technique, as described by Bollen (1989). First, investigators are not allowed to constrain any of the factor loadings to zero, so each factor is assumed to contribute to some total variance. Second, EFA does not allow for correlated error among measurement items, although such error is likely to exist. The number of factors and which items contribute to factors is often based on somewhat arbitrary eigenvalues. Additionally, EFA does not allow for an estimation of model fit, unlike present SEM techniques. It is quite possible that one could find significant factor loadings for each factor, but still have a poor overall model fit, indicating a poorly fitting model.

Structural equation modeling represents a qualitative improvement over both aforementioned techniques. SEM is a more top-down approach, in that it assumes the existence of latent constructs that are measured by items. SEM provides data on how well each item measures a construct in the form of path coefficients, similar to factor analysis, and assesses the unique error in each item, unlike other techniques. As error exists in all of our measurements, understanding both how well and how poorly each item measures (or in factor analysis, defines) a construct is crucial. Finally, SEM allows for a global assessment of how well all the items in a scale measure the construct, offering several alternative statistics to determine goodness of fit for a model. Different statistics become more salient depending on the nature of the analyses or nature of the data set. Factor analysis only has eigenvalues and percent of variance accounted for by the factors.

Clear advantages to the use of SEM are evidenced by an *a priori* approach with analyses that yield accurate information about the data. The present investigation uses SEM to determine the utility of the newly proposed subjective vitality scale. Making the scale as powerful and parsimonious as possible can only serve to benefit future researchers and practioners. Two separate samples were collected for cross validation of the model and to decrease the likelihood that models resulting from modification would be capitalizing on chance (MacCallum, 1986). This use of cross validation is widely accepted to increase the likelihood that a particular model holds in the population (Schumacker and Lomax, 1996).

METHOD

Participants

The sample for this study came from a large, private school in the Midwest. An effort was made to recruit an equal number of men and women, resulting in a final sample of 526 participants. This number was achieved through collecting two samples, allowing researchers to develop a structural equation model on one data set, and cross validate on another sample, a popular approach in the current literature (Hoyle, 1995; Schumaker and Lomax, 1996). The first sample (N = 263) was collected on students in an introductory psychology class. They completed several questionnaires in a classroom as part of their class requirements. Demographically, this sample contained 92 men and 172 women (34% and 66%, respectively), was 72% first year students, had a mean age of 18.83 (SD = 2.48), and contained 44 (16.3%) ethnic minority students.

The second sample (N = 268) was collected approximately one month later by approaching students in several other psychology classes, and offering them extra course credit for filling out the survey at home and returning it at the following class. Demographically, this sample contained 64 men and 204 women (24% and 76% respectively), was 63% first year students, had a mean age of 19.04 (SD = 2.51), and contained 46 (17.2%) ethnic minority students.

Measures

All participants filled out Ryan and Frederick's (1997) vitality questionnaire as part of a larger packet of measures. The measure contains seven times which participants endorse on a seven point likert scale, ranging from "not at all" to "very true." Of the seven items, one is negatively worded, and thus reversed scored.

RESULTS

To ensure that the present data were similar to the original data from Ryan and Frederick, Cronbach's alpha was computed on both halves of the data set. Cronbach's alpha for the first data set was 0.80, and 0.89 for the second, approximating the original published values of 0.84 and 0.86. Principal Components factor analysis indicated a one factor solution for both sets of data. All of the items had high factor loadings (i.e. loadings > 0.60). With evidence that the data had similar internal consistency to Ryan and Frederick's (1997), modeling was undertaken.

Initial data analyses revealed that both samples possessed adequate univariate and multivaraite distributional properties for modeling. The first model tested was identical to that proposed by Ryan and Frederick (1997), in that all seven items were included (Table I). The first model obtained a $\chi^2(14) = 50.51$, $p < 0.01$, suggesting that improvements could be made in the model. An investigation of validity coefficients revealed that item

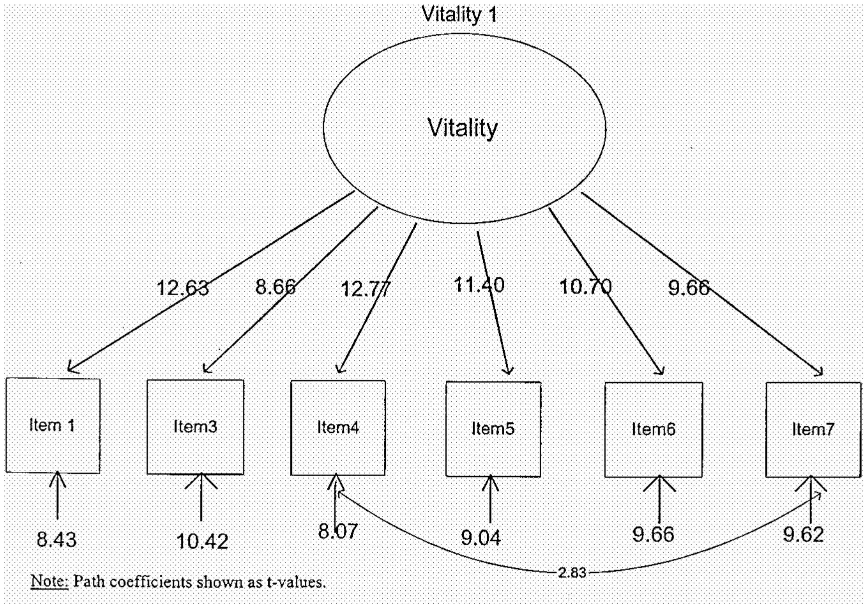

Figure 1. Vitality Model, First Data Set

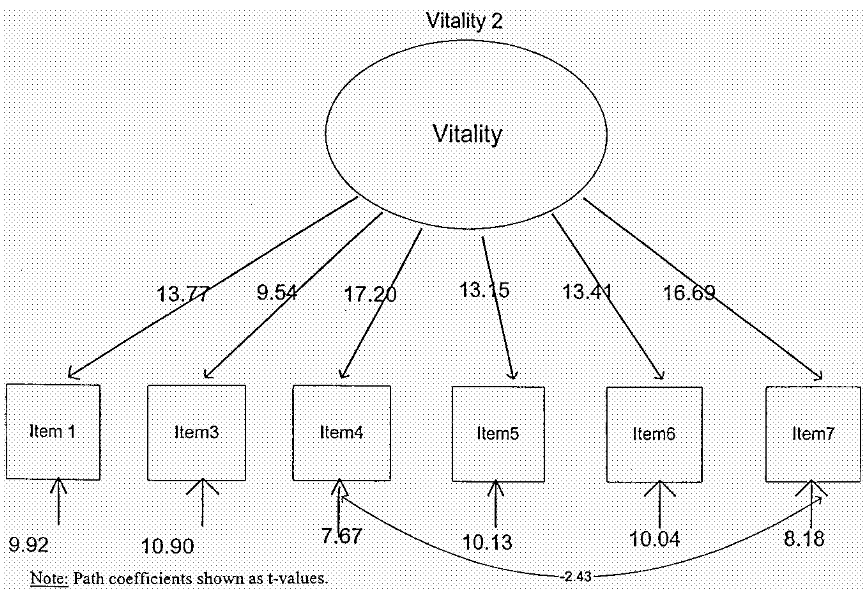

Figure 2. Vitality Model, Second Data Set

TABLE I

Comparison of Three Models on the First Data Set

	Model 1	Model 2	Model 3
χ^2 (df)	(14) = 50.51, $p < 0.01$	(9) = 19.44, p = 0.02	(8) = 10.21, p = 0.25
GFI	0.95	0.98	0.99
AGFI	0.90	0.95	0.97
NFI	0.91	0.96	0.98
RMSEA	0.10	0.06	0.03

Note: χ^2 (df) = Chi square, with degrees of freedom; GFI = Goodness of Fit Index; AGFI = Adjusted Goodness of Fit Index; NFI=Normed Fit Index; RMSEA = Root Mean Square Error of Approximation.

two was not functioning well as a measure of vitality ($R^2 = 0.14$). Removing this item from the model and re-estimating proved a better fitting model, $\chi^2(9) = 19.44$, p = 0.02. Interestingly, this was the one negatively worded item, and there is recent evidence suggesting the lack of equivalence between positively and reversed worded items (Chang, 1996), that negatively worded questionnaires provide different factor structures than positively worded counterparts (Schriesheim and Eisenbach, 1996), and that positively worded items provide greater internal consistency (Schriesheim, Eisenbach, Hill and Kenneth, 1991). Neither the coefficient alpha nor the factor analysis indicated any problems with this item. In fact, the coefficient alpha would have slightly increased by 0.01 with the inclusion of this item.

Finally, an inspection of the modification indices suggested, among many suggestions, the addition of an error covariance between items four and seven; these items are similarly worded and both contain the word “energy.” As these two items were similarly worded and both asked about a specific aspect of vitality, it is likely that they have similar error in their measurement. Allowing the error in these two questions to covary provided the best fitting model, both in terms of χ^2 and all fit indices (see Table I). As recognition and inclusion of correlated error has been shown to affect both goodness of fit measurement and parameter estimates (Reddy, 1993), this final analysis likely provides the most realistic assessment of the present model.

TABLE II

Comparison of Three Models on the Second Data Set

	Model 1	Model 2	Model 3
χ^2 (df)	(14) = 55.30, $p < 0.001$	(9) = 25.27, $p < 0.01$	(8) = 19.95, $p = 0.01$
GFI	0.94	0.97	0.97
AGFI	0.88	0.92	0.93
NFI	0.95	0.97	0.98
RMSEA	0.11	0.08	0.08

Note: χ^2 (df) = Chi square, with degrees of freedom; GFI = Goodness of Fit Index; AGFI = Adjusted Goodness of Fit Index; NFI = Normed Fit Index; RMSEA = Root Mean Square Error of Approximation.

The final model from the first data set was then verified on the second data set to confirm its utility and parsimony. As seen in Table II, this model computed a $\chi^2 = 19.95$, $p = 0.01$, suggesting the overall fit was not as good as it was for the final model in the first data set. Model modifications were then made in reverse from the procedure in the first data set, such that the error covariance between items four and seven was removed in one step, and then item two was re-entered in the model. As can be seen in Table II, each of these modifications provided worse fitting models, in terms of χ^2 and fit indices. No other sensible modifications were either suggested by modification indices or by inspections of the path coefficients, suggesting that the first model of the second data set provided the best fitting, more parsimonious solution. It is important to note that the second data set has a negative correlation between the two error variances. Negative correlations are a result of the model being overfitted. Given that the parameter estimates in the second data set are considerably higher than the first wave of data, it is not surprising that the model overfits the data for the second wave.

To better understand the difference in model fit, the data sets were further explored. T-tests for independent samples were computed for all items, and showed significant differences for two items between the two samples: for items 1 ($p < 0.05$) and 3 ($p < 0.01$), indicating some difference in the two samples. Factor analyses were computed for both data sets using principal components. Data set one showed a factor structure with all items loading on one factor with an Eigenvalue = 3.39 accounting for 48% of the variance. Data set two showed all items loading on one factor, with and Eigenvalue = 4.32, accounting for 62% of the variance.

DISCUSSION

Vitality is a complex construct to measure. Ryan and Frederick (1997) provide a brief measure to assess one's level of vitality. The results of this study showed that the measure does in fact measure one construct, vitality. However, one negatively worded item was removed from the model in order for the model to fit the data. This suggests that this item does not perform well in a unidimensional measure of vitality. The researchers recommend that instead of adding other negatively worded items, that this item be removed from the scale. Previous research indicates that when a balance of negatively and positively worded items comprise a scale; two factors emerge (for further discussion see Chang, 1995 or Schriesheim and Eisenbach, 1996). This would prohibit a single factor model for the measure of vitality.

The data sets showed some similarity, as would be expected for data sets collected at slightly different times using different procedures. However, the final model, with item two removed and an error covariance added proved the best fitting model for two separate data sets, providing strong support for this form of the vitality questionnaire as the best measure of the latent construct of vitality. This approach to model testing by validating a model on an additional sample is a preferred method for model assessment (Hoyle, 1996; Schumaker and Lomax, 1996).

Making model changes, as done with the first data set, allows the researcher to capitalize on chance, that is, to produce a model that fits the data based on idiosyncrasies of that particular data set. Thus the four suggestions of MacCallum (1986) were followed. The initial model closely corresponded to the final model. The search for a better fitting model continued even after a statistically plausible model was found (proceeding from model two to model three in the first data set). Valid restrictions were placed on permissible modifications, and a large data set was used. An excellent fit was achieved on our first data set and adequate fit on the second data set. No sensible modifications would have given a better fit in the second data set, suggesting that the final solution was indeed the best fitting model.

One limitation of this study was its reliance on a college population. However, when comparing the results to Ryan and Frederick's (1997) findings, this measure shows similar properties when collected on a college sample and on a general community sample. This study's strength comes from its use of modeling, large sample size, and use of cross validation techniques. Interestingly, other diagnostic criteria such as factor analysis or coefficient alpha did not indicate the inferior performance of the second item as highlighted in the results. It was only through the use of modeling that this item was found to be deficient.

In summary, this investigation provides the most efficient and valid instrument to measure vitality (a six item measure instead of seven, Appendix). This new measure retained only the positively worded items, creating a measure that has more desirable psychometric properties, and utilizing the inclusion of correlated error which resulted in better goodness of fit indices. Accounting for correlated error can only be done using Structural Equation Modeling. Researchers who calculate a composite score for the measure assume the error is equally distributed for each of the items. This could result in an artificial inflation of the summed score. The diminutive magnitude of the correlated error found herein suggests that the inflation would not be influential in other analyses. Furthermore, the improved measure (i.e. the deletion of item two) is highly correlated with the old measure (0.98) suggesting the comparability between the two measures. The performance of both measures should be identical in practical applications of the scale.

Finally, the present investigation utilized two large samples to allow for a confirmation of findings. Hopefully, future investigators will use this new measure, providing more modeling data on the measure's properties under a wide range of conditions and populations.

ACKNOWLEDGMENT

Correspondence concerning this article should be addressed to Terence J. Bostic, Department of Psychology, Saint Louis University, Shannon Hall, 221 North Grand Boulevard, St. Louis, MO 63139, USA.

APPENDIX

RYAN AND FREDERICK'S (1997) SUBJECTIVE VITALITY MEASURE

1. I feel alive and vital
2. I don't feel very energetic[1]
3. Sometimes I am so alive I just want to burst
4. I have energy and spirit[2]
5. I look forward to each new day
6. I nearly always feel awake and alert
7. I feel energized[2]

Note: (1) This item was reversed scored in all calculations, both by the present investigators and by Ryan and Frederick (1997). This was also the item that was removed in the final model. (2) These two items had the error covariance added in the final model. Reproduced with permission of the authors.

REFERENCES

Bollen, K.: 1989, Structural Equations with Latent Variables (Willey & Sons, New York).

Cahng, L.: 1995, 'Connotatively consistent and reversed connotatively inconsistent items are not fully equivalent: Generalizability study', Education and Psychological Measurement 55, pp. 991–997.

Cronbach, L.J.: 1951, 'Coefficient alpha and the internal structure of tests', Psychometrica 16, pp. 297–334.

Hoyle, R.: 1995, Structural Equation Modeling; Concepts, Issues, and Application (Sage, Thousand Oaks, CA).

Liao, W.: 1990, T'ai Chi Classics (Shambahala Publications, Boston, MA).

MacCallum, R.: 1986, 'Specification searches in covariance structure modeling', Psychological Bulletin 100, pp. 107–120.

McNair, D.M., M. Lorr and L.F. Droppleman: 1971, Profile of Mood States Manual (Educational and Industrial Testing Service, San Diego).

Miller, B.: 1995, 'Coefficient alpha: A basic introduction from the perspective if classical test theory and structural equation modeling', Structural Equation Modeling 2, pp. 255–273.

Nunnally, J. and I.H. Bernstein: 1994, Psychometric Theory, 3rd ed. (McGraw-Hill, New York).

Ryan, R.M. and C. Frederick: 1997, 'On energy, personality and health: Subjective vitality as a dynamic reflection of well-being', Journal of Personality 65, pp. 529–565.

Reddy, S.: 1991, 'Effects of ignoring correlated measurement error in structural equation models', Education and Psychological Measurement 52, pp. 549–570.

Schreisheim, C.A. and R.J. Eisenbach: 1996, 'The exploratory and confirmatory factor-analytic investigation of item wording effects on the obtained factor structures of survey questionnaire measures', Journal of Management 2, 1177–1193.

Schriesheim, C.A., R.J. Eisenbach and K.D. Hill: 1991, 'The effects of negation and polar opposites on questionnaire reliability and validity: An experimental investigation', Education and Psychological Measurement 51, pp. 67–78.

Schumaker, R.E. and R.G. Lomax: 1996, A Beginner's Guide to Structural Equation Modeling (Lawrence Erlbaum Associates, Mahwah, NJ).

Tabachnik, B.G. and L.S. Fidel: 1996, Using Multivariate Statistics, 3rd ed. (HarperCollins, New York).

Department of Psychology
College of Arts & Sciences
Saint Louis University
St. Louis, MO 63103
USA