Center for Demography and Ecology
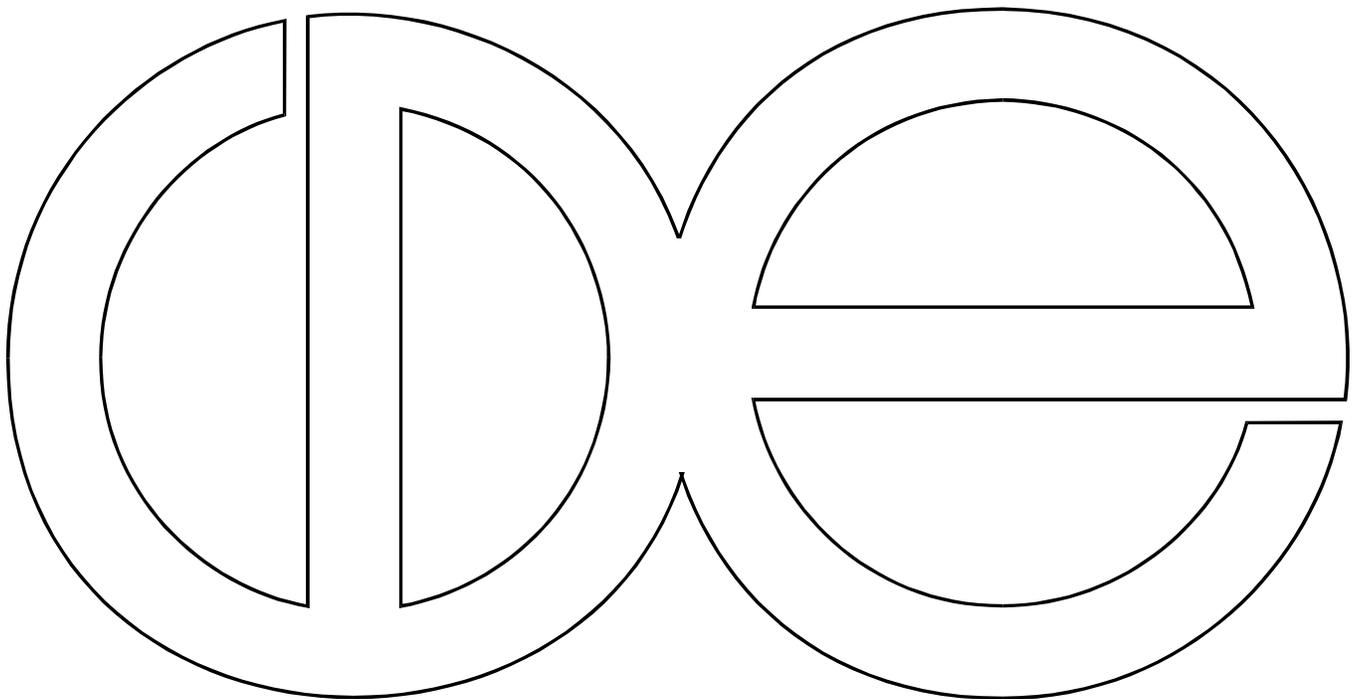
University of Wisconsin-Madison

Survey Measurement of Psychological Well-Being

Kristen W. Springer

Robert M. Hauser

SURVEY MEASUREMENT OF PSYCHOLOGICAL WELL-BEING*

Kristen W. Springer
Robert M. Hauser

Institutional Affiliation: University of Wisconsin-Madison Department of Sociology and Center for Demography of Health and Aging

Running Head: Measurement of Psychological Well-Being

Key Words: psychological well-being, well-being, measurement, survey design

Word Count: 7,892 including table location indicators but excluding notes

Abstract Count: 149

Number of Tables: 7

Number of Figures: 0

1

SURVEY MEASUREMENT OF PSYCHOLOGICAL WELL-BEING

ABSTRACT

This study assesses the measurement properties of Ryff's scale of Psychological Well-Being (PWB) -- a widely-used scale designed to measure six dimensions of PWB. Analyses of self-administered PWB data from three major surveys – Midlife in the United States (MIDUS), National Survey of Families and Households II (NSFH II), and the Wisconsin Longitudinal Study (WLS) yielded very high overlap among subscales. We eliminated several methodological sources of confounding, including question wording, question order, and negative item wording. We also analyzed WLS telephone survey data and found smaller subscale correlations. Self-administered instruments may provide more valid psychological measurements than telephone surveys, and survey mode may affect the structure of factor models. Thus, we believe that mode effects may explain the consistent differences between findings from telephone administration and self-administration of PWB scales. That is, Ryff's scale does not measure six distinct dimensions of PWB, and researchers should use it with caution.

SURVEY MEASUREMENT OF PSYCHOLOGICAL WELL-BEING

Health researchers have long moved past looking at mortality as the only health-related outcome to examine a range of outcomes including morbidity, disability, quality of life, and psychological well-being. Mental health research has often focused on negative health—for example on depression and anxiety. However, there is an increasing desire to examine positive as well as negative aspects of mental health. Much of this research has drawn from the rich well of psychological literature on well-being.

Well-being has been studied extensively by social psychologists (Campbell, 1981; Ryan & Deci, 2001). While the distinct dimensions of well-being have been debated, the general quality of well-being refers to optimal psychological functioning and experience. Two broad psychological traditions have historically been employed to explore well-being. The hedonic view equates well-being with happiness and is often operationalized as the balance between positive and negative affect (Ryan & Deci, 2001; Ryff, 1989a). Scholars from the time of Aristotle have pointed to the limitations of hedonic happiness as the sole criterion for psychological well-being (Ryan & Deci, 2001; Waterman, 1993). These scholars often subscribe to the eudaimonic tradition of well-being which assesses how well people are living in relation to their true self (Waterman, 1993). According to Waterman (1993), eudaimonia occurs when people's lives are congruent with their true direction and values.

MEASURING WELL-BEING

Conceptualizing types and dimensions of well-being is an important first step, but the validity of the concepts remains questionable until they are subject to scientific

examination. Therefore, an important second step toward a theory of psychological well-being is to operationalize measures of this concept. There does not appear to be a standard or widely accepted measure of either hedonic or eudaimonic well-being in the literature, although commonly used instruments include: Bradburn's (1969) Affect Balance, Neugarten's (1961) Life Satisfaction Index, Rosenberg's (1965) self-esteem scale, and a variety of depression instruments. In addition, some scholars have pointed to the multidimensionality of well-being and believe that instruments should encompass both hedonic and eudaimonic well-being (Compton et al., 1996; McGregor & Little, 1998; Ryan & Deci 2001).

CONCEPTUALIZING A MULTIDIMENSIONAL MODEL OF WELL-BEING

Carol Ryff has argued in several publications that previous perspectives on operationalizing well-being are atheoretical and decentralized (Ryff, 1989a; Ryff, 1989b). To address this shortcoming, Ryff developed a new measure of psychological well-being (PWB) that consolidated previous conceptualizations of eudaimonic well-being into a more parsimonious summary. The exact methods used to develop this measure and the specific theoretical foundations underlying each dimension have been thoroughly presented elsewhere and will not be detailed here (Ryff, 1989a; Ryff 1989b). Briefly, Ryff's PWB scale includes the following six components of psychological functioning: a positive attitude toward oneself and one's past life (self-acceptance), high quality, satisfying relationships with others (positive relations with others), a sense of self-determination, independence, and freedom from norms (autonomy), having life goals and a belief that one's life is meaningful (purpose in life), the ability to manage life and one's

surroundings (environmental mastery), and being open to new experiences as well as having continued personal growth (personal growth).

Ryff (1989a) originally validated her instrument on a sample of 321 well-educated, socially-connected, financially-comfortable and physically healthy men and women. In this study Ryff used a 20-item scale for each of the six constructs, with approximately equal numbers of positively and negatively worded items. The internal consistency coefficients were quite high (between .86 and .93) and the test-retest reliability coefficients for a subsample of the participants over a 6-week period were also high (.81-.88).

The intercorrelations of the new measures ranged from .32 to .76. The largest correlations were between self-acceptance and environment mastery (.76), self-acceptance and purpose in life (.72), purpose in life and personal growth (.72), and purpose in life and environmental mastery (.66).[1] As noted in her paper, these high correlations can indicate a problem because: "as the coefficients become stronger, they raise the potential problem of the criteria not being empirically distinct from one another" (Ryff, 1989a p. 1074). However, Ryff provides evidence that these dimensions are distinct. Such evidence includes the findings that dimensions relate differentially to age (an issue that we do not investigate in this paper).

Importantly, some other studies that use Ryff's scale have explored correlations among the subscales. One study conducted by Schmutte & Ryff (1997) with a group of 114 women and 101 men aged 44 to 65 years old using 14-item subscales found scale correlations of .80 between self-acceptance and environmental mastery, .80 between purpose in life and self-acceptance, and .72 between purpose in life and environmental

mastery. The other subscale intercorrelations were much smaller, but all were .34 or greater. The items with high correlations in Schmutte and Ryff's (1997) analysis are also the items that correlated most highly in Ryff (1989a).

In contrast, a study by Clarke et al. (2001) examined PWB in the Canadian Study of Health and Aging and found relatively low scale correlations. The Ryff scale includes the same 18 items used by Ryff and Keyes (1995) but was administered orally in the home using a cue card with six response categories (personal communication with Liz Sykes, 03/24/03). This sample was not young; the average age was 76. The scale intercorrelations in this study were small to modest ranging from -.04 to .39. The correlation between environmental mastery and self-acceptance was .39, the correlation between purpose in life and self-acceptance was -.08 and between purpose in life and environmental mastery was .05.

In a more recent study, Ryff and Keyes (1995) directly tested the multidimensionality of Ryff's PWB model on MIDUS pretest data--a national probability sample of 1,108 men and women. Rather than testing the full scale, Ryff and Keyes selected 3 of the original 20 items in each subscale "to maximize the conceptual breadth of the shortened scales." They also report, "The shortened scales correlated from .70 to .89 with 20-item parent scales. Each scale included both positively and negatively phrased items" (p. 720). Respondents were interviewed by telephone, and PWB items were administered using an unfolding technique—where respondents were first asked if they agreed or disagreed with the statement and then were asked whether their (dis)agreement was strong, moderate, or slight. Ryff and Keyes (1995) estimated confirmatory factor models by weighted least squares in LISREL based on

variance/covariance matrices produced by PRELIS to account for the non-normality of the data (Jöreskog and Sörbom 1988). However, they did not use polychoric correlation matrices; that is, they analyzed all variables as if they were continuous, not ordinal. They estimated several models including (a) all 18 indicators loading on a single common factor, (b) each set of three indicators loading on a factor representing a specific dimension of well-being, (c) models accounting for the possibility that people might respond differently to negatively worded items, and (d) a second-order factor model in which the six factors for specific dimensions of well-being loaded on a general well-being factor. Although none of their models yielded a satisfactory fit by conventional measures (the lowest chi-square was 378.7 with 129 degrees of freedom), the Bayesian Information Criterion (BIC) was consistently a large negative number, indicating satisfactory model fit (Raftery 1995). Ryff and Keyes concluded that a second-order factor model is the best fitting model. Examining the correlations among their latent variables we see some overlap among the subscales. The largest correlation, .85, is between environmental mastery and self-acceptance, suggesting that these factors are largely measuring the same concept.[2] However, Ryff and Keyes note that these two concepts have different age profiles, thus indicating that they may be distinct at different stages of the life course. The other correlations range from .24 to .65: .65 between positive relations and self-acceptance, .65 between environmental mastery and positive relations, and .64 between purpose in life and personal growth. Ryff and Keyes' estimates of correlations among the six latent dimensions of PWB are reproduced in Table 1.

Ryff and Keyes (1995) also compare their scale with prior measures and find that self-acceptance and environmental mastery have moderate correlations (absolute value between .32 and .50) with happiness, life satisfaction, and depression. Overall, the other dimensions are less strongly correlated with other measures of well-being, although all but two relationships are statistically significant at the .05 level. Previous work by Ryff (1989a) has demonstrated even larger associations of PWB dimensions with affect, life satisfaction, and depression. The correlations range from .25 to .73 in absolute value with all relationships significant at the .001 level. This consistency across studies suggests that although Ryff's scale is designed to measure eudaimonic well-being, at least some of the dimensions are also related to hedonic well-being.

In addition to Ryff and Keyes (1995), other scholars have also explored the measurement properties of Ryff's PWB model in diverse samples. In the Clarke et al. (2001) study discussed earlier, the authors used EQS with maximum likelihood estimation to explore the factor structure of PWB. The authors began with a single-factor model but then added additional factors in a consistent manner assessing the model fit at each stage. The authors found that a six factor model fit better than models with fewer factors but the best fitting model was a modified six-factor model that allowed four items (one each from four dimensions) to load on their specified dimension *and* another dimension. The factor correlations in the pure six-factor model ranged (in absolute value) from small to quite substantial (.03 to .67). The authors conclude that their analyses "support the multidimensional structure of the Ryff measure" (p. 86). The

authors also note that results from their modified six-factor model suggests areas for improvement of the 18 item model, though they caution against making such modifications if they would alter the conceptual breadth of the constructs.

In addition to research finding support for a six factor model of PWB, research is beginning to emerge that questions the multidimensionality of PWB. For example, Kafka & Kozma (2002) examined PWB, the Satisfaction with Life Scale (SWLS), and the Memorial University of Newfoundland Scale of Happiness (MUNSH) in a sample of 277 participants ranging from 18-48 years old. Their PWB scale contained the full set of items (20 per subscale) and was administered to university students in a self-report questionnaire. The authors used principal-components analysis with varimax rotation. When the number of factors was not specified, 15 factors were extracted. However, when the authors limited the scale to 6 factors the factors did not correspond to the six dimensions of PWB. In an additional test the authors examined a factor model with SWLS, MUNSH and each dimensions of PWB. They extracted three factors with eigenvalues greater than 1. The first factor had loadings above .6 for four PWB scales: environmental mastery, self-acceptance, purpose in life, and personal growth and accounted for almost one-half the variance. The second factor accounted for 16% of the variance and was primarily the MUNSH and SWLS, though environmental mastery and self-acceptance also had loadings above .40 on this dimension. The final factor had loading of over .80 for autonomy and personal relations. The authors conclude by saying "it would appear that the structure of PWB[3] is limited to face validity" (p. 186).

Van Dierendonck (2004) examined the factorial structure of PWB in two Dutch samples—a group of 233 college students with a mean age of 22 years old and a group of

420 community members with a mean age of 36 years old. The college student sample was administered 14 items per subscale and the community sample received 9 items per subscale. The items were translated into Dutch and back-translated into English when needed to refine item wording. Van Dierendonck compared model fit and factorial structure of the 3, 9 and 14-item subscales using LISREL 8.5 with covariance matrices and maximum likelihood estimation. The author found that across both samples, for all subscale sizes, the best fitting model was a six factor model with a single second order factor. However for the 3-item scale in sample 1, a second order five factor model (with environmental mastery and self-acceptance together) did not fit significantly worse than the second order six factor model. Only the 3 items per subscale version, which had relatively low internal consistency, fit reasonably well and modification indices suggested allowing some items to load on two dimensions. According to Van Dierendonck, "the conclusions from the reliability analyses and the confirmatory factor analyses are ambiguous. To reach an acceptable internal consistency, scales should be longer, whereas an (somewhat) acceptable factorial validity requires the scales to be short" (p. 636). In addition, Van Dierendonck found very high factor correlations among self-acceptance, purpose in life, environmental mastery and personal growth indicating substantial overlap among these dimensions (personal communication with Dirk van Dierendonck, 7/26/04)

Hillson (1997) examined the structure of PWB among two samples of college students who were given self-administered questionnaires containing an 84 item Ryff PWB scale. A series of exploratory and confirmatory factor analyses did not reveal a model with six dimensions, but rather a model with 3 and 4 dimensions. In one sample

one of the dimensions was primarily composed of items from the self-acceptance, environmental mastery, purpose in life and autonomy subscales. In the second sample one factor was primarily composed of self-acceptance, purpose in life, and environmental mastery items.

Ryff's model of PWB has been administered in major studies, for example, the National Survey of Families and Households II (NSFH II), the National Survey of Midlife in the United States (MIDUS), the Wisconsin Longitudinal Study (WLS), and the Canadian Study of Health and Aging (CSHA). In addition, Ryff's 1989 paper, where this model was developed, has been cited in more than 350 research papers. Given the substantive importance and wide-spread use of Ryff's scales, it is important to understand their measurement properties. Also, given mixed evidence about the multidimensionality of PWB reported in the literature, it is somewhat surprising that the factorial structure of PWB has not been examined systematically in any of the large, widely-used U.S. surveys. Finally, the seminal study (Ryff & Keyes 1995) on the measurement of Ryff's model of PWB was conducted on items administered by telephone whereas most large scale studies using this measure are self-administered. It is therefore essential to rigorously assess the role of mode effects in the measurement properties of Ryff's PWB model.

The present project attempts to fill this gap by exploring the measurement properties of PWB using self-administered mail surveys in the WLS, MIDUS, and NSFH II, and telephone data from the WLS. We start by examining the measurement properties of Ryff's model using the WLS mail data. The WLS is a particularly useful sample in which to explore Ryff's PWB model because: (a) the sample is large and 6282 graduates

answered all of the mail questions in 1992-1993, (b) almost all of the graduates were born in 1939, so we have a unique opportunity to look at how the PWB measure works for individuals at midlife, and (c) the WLS administered PWB items both by telephone and mail, thus allowing us to explore mode effects.

In order to test the validity of our results and examine possible confounders, we employ a variety of tests. First, using the WLS mail data, we explore whether measurement artifacts (negative wording and question ordering) could be driving our findings. We then turn to NSFH II and MIDUS to explore the generalizability of the WLS findings and to test whether the WLS results are caused by age truncation, educational truncation, a primarily white sample, item selection, or something geographically distinct about Wisconsin. Finally, we assess mode effects by analyzing the WLS telephone data.

DATA

Items from Ryff's scales of PWB were included in the WLS mail and telephone instruments, MIDUS mail, and NSFH II mail instruments. Before going into each study in detail, an overview is warranted. It is important to point out the differences and similarities in order to fully assess the why the measurement properties of PWB may vary across samples. Table 2 shows which items were asked on each survey, how they were worded, how they were introduced, the order in which they were asked, and what response categories were used. NSFH II and MIDUS contain the same 18 items—with some slight wording differences. The WLS mail instrument contains 6 of these 18 items

in addition to 36 other items. The WLS telephone instrument contains 12 of these 18 items. Also, note that the response categories are not identical across the surveys.

--Table 2 About Here--

WLS

The WLS has followed a random sample of 10,317 men and women who graduated in 1957 from Wisconsin high schools (Sewell, Hauser, Springer, Hauser 2002). Respondents were surveyed in 1957 and then again in 1975. In 1977 the study design was expanded to collect information similar to the 1975 survey for a highly stratified, random subsample of approximately 2000 siblings of the graduates. Between 1992 and 1994 another major wave of data collection was undertaken. This included follow-up interviews with living graduates and with an expanded sample of siblings. Briefly, the WLS now has active samples of 8500 WLS graduates out of 9750 survivors and 5300 of their siblings. We focus on the WLS graduates in this paper, but analyses of the sibling data yield essentially the same findings. WLS participants mirror the racial composition of most Wisconsin high school graduates in 1957 and as such are primarily white and non-Hispanic.

Items from Ryff's psychological well-being model were included in the 1992-1994 telephone interview and mail survey. The mail survey contained seven items for each subscale, yielding a total of 42 items (see Table 2b). In the mail survey all six constructs of Ryff's PWB model included items with reversed scales. The order of the items in the mail survey generally follow the pattern of asking one item from each of the

constructs in the following order: autonomy (aut), environmental mastery (env), personal growth (grow), positive relations (rel), purpose in life (purp), and self-acceptance (acc). Six sets of sequential questions ask items in this order. These six sets are split up by items from the remaining 7th set. For example questions 1-21 covered the following constructs: aut, env, grow, rel, purp, acc, *aut,* aut, env, grow, rel, purp, acc, *env*, aut, env, grow, rel, purp, acc, *grow*, etc. where the italicized items are those in between the set of six constructs. Note that these "splitter" items are in the same order as the six groups of constructs. Therefore, two items from the autonomy construct and two items from the self-acceptance construct are adjacent in the mail survey. Participants in the mail survey were given a 6 point scale ranging from strongly agree to strongly disagree.

All of the items are highly skewed and do not follow a normal distribution. A variety of transformations were attempted to help create a normal distribution, however the significant skew warranted more extensive treatment which will be described in the methods section. To explore the possibility of artificial answers (outliers) we checked for cases where people answered all questions with a six or all questions with a one. Given that many of the items are reverse coded, this seems implausible and would be highly suspicious. We did not find any cases where this occurred. There were 6875 respondents who responded to at least some of the PWB items, and a total of 6282 respondents have complete data for all mail items.

The WLS telephone instrument contains 2 items from each scale for a total of 12 PWB items. These items are different from those asked in the mail questionnaire. The two positive relations items were both negatively worded and the two self-acceptance items were both positively worded. The four other subscales contained one positively

14

and one negatively worded item. The items were ordered randomly (see Table 2b). An

unfolding technique (Groves, 1989) was used during the telephone interview. As in the

MIDUS pretest, participants were first asked whether they agreed or disagreed with the

statement and then asked about the intensity of this belief (strong, moderate, or slight).

There were 6038 respondents with complete data on the PWB telephone items, which

were administered in a random 80 percent of the WLS interviews. As with WLS mail

items we checked to see if anyone answered all 1s or all 6s, but found that no one had

done so. Interestingly, the distribution of the responses was bimodal and skewed, likely

reflecting the use of the unfolding technique.

MIDUS

MIDUS is a multistage probability sample of over 3000 noninstitutionalized

adults between the ages of 25 and 74 years old. Participants were selected based on

random-digit dialing and were administered a telephone interview as well as two mail-

back questionnaires. Data were collected during 1994 and 1995. The PWB scale was

included in one of the mail questionnaires and contained 18 items in what appears to be a

random order (see Table 2b). Response choices ranged from 1 to 7 (agree strongly, agree

somewhat, agree a little, don't know, disagree a little, disagree somewhat, and disagree

strongly). As with the WLS, we looked for outliers and found that one person chose

answer "1" for all items. This individual was removed from the analyses. For the current

project the mid-point "don't know" category was recoded as missing data. There were

2731 cases with complete PWB data. The majority of the items were unimodal and all

were skewed toward positive (or non-negative) responses.

NSFH II

The National Survey of Families and Households began in 1987-1988 with a

national sample of more than 10,000 households. In each household, a randomly selected

adult was interviewed. The five year follow-up was conducted in 1992 to 1994 and

included data collection from 10,000 respondents, 5600 interviews with spouses/partners,

2400 interviews with children, and 3300 interviews with parents. The focus of this

project is on the main respondents. The PWB scale was included in the self-administered

health module completed during an in home interview. The PWB scale contained the

same 18 items as MIDUS arranged in a seemingly random order, though in a different

order than MIDUS (see Table 2b). As with the WLS, we checked for outliers in the data

and found that 12 people answered either all sixes or all ones. These people were

removed from the analyses leaving 9237 NSHF II cases with complete data. The

distributions were generally unimodal and all were skewed toward positive (or non-

negative) responses**.**

METHODS

Our strategy for exploring the structure of PWB was to begin with the WLS mail

data, systematically assessing the model fit and correlations of factors for a series of

models starting with the single factor model. Then, in order to test possible confounders

and explanations for our findings, we ran a series of validity checks including tests for

methodological artifacts, age truncation, instrumentation issues, item selection problems,

and cultural variation. In the latter process we examined NSFH II and MIDUS mail data and WLS graduate telephone data.

In order to explore the structural relationship of the items with their conceptual dimensions we estimated confirmatory factor models using LISREL 8.53 (Jöreskog & Sörbom, 1996a). However, LISREL may produce biased estimates if the variables are ordinal or non-normal. In this case, it is necessary to provide LISREL with polychoric correlation matrices and asymptotic variance/covariance matrices rather than simple covariance or correlation matrices. In order to calculate polychoric correlations, PRELIS 2.53 (Jöreskog & Sörbom, 1996b) assumes that each ordinal or non-normal variable is a crude measurement of an underlying, unobservable, continuous variable. In the case of ordinal data, this unobservable variable has a multivariate normal distribution and polychoric correlations are theoretical correlations of these hypothetical, normally distributed, underlying variables. We used PRELIS to estimate the polychoric correlations for all models. After obtaining the polychoric correlations and the asymptotic variance/covariance matrix, we used weighted least squares estimation in LISREL to obtain parameter estimates and model fit statistics. In addition to examining minimum fit chi-square statistics, we used BIC to assess model fit, accounting for sample size (Raftery 1995).

FINDINGS

WLS Graduate Mail Items

--Table 3 About Here--

In order to have a baseline we started out by running a single factor model—a model with all indicators loading on only one common factor. As shown in Table 3, the fit for Model 3-1 is very poor both by chi-square and BIC standards. We next ran a six factor model allowing the latent variables to correlate (Model 3-2). This model fits very well compared to the single factor model; chi-square is 9036 with 804 degrees of freedom. We next ran the second-order factor model (Model 3-3) and found that it does not fit as well as the six factor model without a second-order factor.

In panel 1 of Table 5 we present the correlations among latent variables in the six factor model. There are very high correlations among latent variables—particularly between self-acceptance & purpose in life (.976), self-acceptance & environmental mastery (.971), and environmental mastery & purpose in life (.958). Personal growth also correlated highly with self-acceptance (.951), purpose in life (.958) and environmental mastery (.908).

Given the large second-order factor loadings and correlations among environmental mastery, purpose in life, and self-acceptance, we wanted to test a model where the variances of the disturbances of environmental mastery, purpose in life, and self-acceptance are set equal to zero. What this effectively does is to constrain each of the factor loadings for these three dimensions to equal 1, stating that these three dimensions are essentially the same as the general well-being factor and equal to each other. As shown by the BIC and chi-square in Table 3, this model (3-4) fits better than

the single factor model, but it does not fit as well as the unconstrained second-order

factor model or the six factor model.

Because personal growth also has a very large factor loading in the second-order

factor model and because it correlates highly with environmental mastery, purpose in life,

and self-acceptance, we decided to also try a model with the variances of the disturbances

of all of these dimensions fixed to zero. As shown in Table 3 this model (3-5) is the

worst fitting of all models yet estimated, with the exception of the single factor model.

Finally, we explored the possibility of a model with the items from environmental

mastery, self-acceptance, and purpose in life loading on one factor (well-being 1) and the

other items loading on first order subscale factors which load on a second-order well-

being factor (well-being 2). These two second-order factors are allowed to correlate. By

including this correlation we allow any departure of equivalence among environmental

mastery, self-acceptance, and purpose in life to be captured by this correlation. Thus, this

model permits us to ask whether our items are actually measuring two underlying aspects

of well-being that are correlated, but not equivalent.

As shown in Table 3 (Model 3-6), this model has a BIC of approximately 2200,

which is smaller than the other constrained models (Models 3-4 and Models 3-5), but

indicates a poor fit relative to the six factor model. Also, it is important to note that the

correlation between the two well-being factors was .99, suggesting that there is

essentially complete overlap between these two factors.

In a final model, we conducted an analysis similar to that for Model 3-6, with the

exception that personal growth was included in the dimensions loading onto a distinct

factor. As shown in the findings for Model 3-7, this model does not fit well. The BIC is

2334, only smaller than that in the single factor model and Model 3-5. In addition, the correlation between the two well-being factors was 1.00 suggesting that these are not two factors at all—but rather one factor.

Given the very high correlations among factors for the subscales, it is important to find out what could be driving these results. First, we explored the possibility that methodological artifacts were responsible for the very high correlations.

--Table 4 About Here--

Testing for methodological effects

Using the six factor model as the baseline (Model 4-1) we explored several possible methodological artifacts.[4] First, following Ryff & Keyes (1995), we introduced a latent variable for negatively worded items. A negatively worded item is one to which someone must answer "strongly disagree" to indicate positive well-being. One example from the autonomy subscale is: "I tend to worry about what other people think of me." To report a high degree of autonomy one would have to report strongly disagree. By including a factor for negative wording, we test whether people answer items differently simply because they are worded negatively. Indeed, some researchers have found that people provide inconsistent answers to negatively and positively worded items (Chapman & Tunmer, 1995; Marsh, 1986; Melnick & Gable, 1990; Pilotte & Gable, 1990). To carry out this test, we allowed all 22 negatively worded items to load on this factor as well as on their corresponding well-being dimensions. As Model 4-2 shows, including negative items vastly improves model fit—resulting in a reduction of 1500 chi-square and

a BIC of 753 compared to 2004 for the six factor model. Clearly, this is the best fitting

model yet. The loadings on the negative wording item ranged from .298 to .497. The

correlations of the negative latent construct with the other PWB dimension constructs

were relatively low—ranging (in absolute value) from .012 to .128.

A second methodological artifact is correlated measurement error between

adjacent items. As explained in the data section, the PWB items were interspersed in a

systematic manner—but probably one invisible to the participant. Nonetheless, we

hypothesized that a response to a particular question might affect response to following,

adjacent questions. To test this, we introduced correlated errors of measurement between

all adjacent questions. The results of this model (Model 4-3) are presented in Table 4 and

indicate a substantial improvement in fit over the six factor model (chi-square = 8141

with 763 degrees of freedom).

Model 4-4 combines the two methodological effects tested in Model 4-2 and

Model 4-3. As Table 4 indicates, this model fits better than all previous models with a

chi-square of 6660 and 735 degrees of freedom resulting in a BIC of 232. There were

three very large modification indices suggesting that the model would be substantially

improved if a few changes were made. Specifically, the model suggested correlating the

measurement error of three pairs of items. Before deciding whether or not to allow the

pairs to correlate, it was essential to determine whether there were plausible reasons for

this. We believe there are for the following reasons. First, each pair is in the same PWB

subscale, suggesting we might have some items within PWB dimensions that overlap

greatly. Second, the manifest content of the paired items is similar. The two

environment mastery questions are: "I am good at juggling my time so that I can fit

everything in that needs to get done" and "I am quite good at managing the many responsibilities of my daily life." The two personal relations items are: "People would describe me as a giving person, willing to share my time with others" and "Most people see me as loving and affectionate." The personal growth items are: "I don't want to try new ways of doing things—my life is fine the way it is" and "I do not enjoy being in new situations that require me to change my old familiar way of doing things."

After allowing these items to correlate we lost three degrees of freedom but had a reduction of 700 chi-square—a substantial improvement in fit (see Model 4-5). This model yields a good fit as indicated by a negative BIC (-449). It is important to look at the correlations among latent variables in this methodologically corrected model. Do the corrections account for the high overlap among Ryff's subscales? As shown in panel 2 of Table 5, we still find very high correlations between purpose in life & self-acceptance (.970), self-acceptance & environmental mastery (.970), and purpose in life & environmental mastery (.958). We can therefore rule out the possibility that three methodological artifacts – negative wording, adjacency effects, and item redundancy – are responsible for the high correlations among the factors for these three latent subscales.

--Table 5 About Here--

Additional Validity Checks: MIDUS and NSFH II

There are several other possible sources of the high latent variable intercorrelations that we estimated in the WLS mail survey. Item selection is one

possibility. Specifically, it is possible that the items included in the WLS mail survey

had more conceptual overlap than other items—for example those used by Ryff and

Keyes (1995). As explained above, NSFH II and MIDUS contain the same 18 items, and

these are also the items used by Ryff and Keyes (1995). The WLS mail instrument, on

the other hand, contains only 6 of these items—one per dimension. Analyzing MIDUS

and NSFH II allows us to test whether the specific choice of items in the WLS accounts

for the high latent variable correlations. Analyzing MIDUS and NSFH II also allows us

to explore the possibility that age truncation, education truncation, limited racial

diversity, or something distinct about Wisconsin are driving our results. That is, both of

these national surveys cover adults of all ages, all levels of completed schooling, and all

race-ethnic groups.

--Table 6 About Here--

In order to compare factor correlations with the WLS, we ran six-factor models

for the NSFH II and MIDUS. The chi-square statistic for NSFH II is 4567.97 with 120

degrees of freedom (BIC = 3,472) and the chi-square statistic for MIDUS is 1471.22 with

120 degrees of freedom (BIC = 522). Table 6 contains the latent variable correlations for

both the NSFH II and MIDUS studies. As panel 1 in Table 6 shows, there are universally

large latent variable correlations using the NSFH II data with the correlation between

personal growth and purpose in life at .981 and that between self-acceptance and

environmental mastery at .933. Panel 2 in Table 6 shows that the MIDUS factor

correlations are more modest, though not small. In addition, the highest correlations are

between purpose in life and self-acceptance (.871), and self-acceptance and environmental mastery (.858). We see from these correlations that there are several very high correlations, and generally the highest correlations across all three mail surveys are among four factors (purpose in life, self-acceptance, environmental mastery and personal growth).

--Table 7 About Here--

Mode of Administration: WLS telephone interview

One final possible test is to explore the method of item administration. In addition to 42 PWB items on the mail survey, the WLS also contains 12 telephone items. These 12 items are among the 18 items included in the NSFH II and MIDUS. Looking again at a six factor model, the chi-square is smaller than for the WLS mail items (chi-square = 578). Table 7 contains the factor correlations for the WLS telephone items. In general the factor correlations are much more modest; the lowest is only .321. However, the largest correlation is greater than 1 for environmental mastery and self-acceptance, and the correlation for self-acceptance and personal growth is greater than .80. The correlation greater than one is troubling at first look; however when the correlation is constrained to be equal to 1 the chi-square change is not significant for one degree of freedom indicating that the estimated correlation is not significantly greater than one.

DISCUSSION

The purpose of the present paper was to explore the measurement properties of Ryff's model of PWB in several widely used large data sets under a variety of conditions. While it is important to understand the measurement characteristics of any scale, there may be a greater urgency to do so with Ryff's scales because they are widely used. Ryff and Keyes (1995) provide empirical support for the multidimensionality of PWB, with the exception of environmental mastery and purpose in life. Clarke et al. (2001) also find some evidence for the multidimensionality of PWB using a sample of older Canadians. However, in this sample they also find substantial overlap of environmental mastery with self acceptance and of purpose in life with personal growth. Finally, some studies using smaller sample sizes and exploratory factor analyses do not find evidence of the multidimensionality of PWB (Kafka & Kozma 2002; Hillson 1999). Our study is the first that we know of to explore the measurement properties of Ryff's PWB model in the WLS, MIDUS, or NSFH II. It is also the first measurement analysis of Ryff's models to use polychoric correlations and weighted least squares estimation. Weighted least squares estimation with polychorics is becoming recognized as one of the best ways to model ordinal, non-normal data (Joreskög & Sorbom 1996a). Using these preferred methods for analyzing three large U.S. self-administered samples, we find very little support for the theoretically proposed multidimensionality of Ryff's PWB scales.

Comparing our WLS telephone estimates, along with those reported by Ryff and Keyes (1995) in the MIDUS pretest, to the findings from self-administered items, helps explain the seemingly discrepant findings in the literature. We found consistently high factor correlations across all self-administered instruments, even in the WLS when we accounted for methodological factors such as negative wording, item placement, and item

redundancy. In addition, examination of NSFH II and MIDUS provides strong evidence that the high factor correlations found in the WLS are not a result of educational truncation, age truncation, item selection, or some distinctive characteristic of the Wisconsin population.

It is worth noting that the NSFH II and MIDUS correlations, though very high, are not as high as those estimated in the WLS. Importantly, the highest latent variable correlations across all samples were consistently those among purpose in life, self acceptance, environmental mastery and personal growth. These four dimensions are those identified previously in the literature as having conceptual overlap. Kafka & Kozma (2002) found these four dimensions largely clustered in one factor, Hillson (1999) found that one factor contained primarily self-acceptance, environmental mastery, and purpose in life subscales, and Van Dierendonck (2004) found high correlations among these subscales (personal communication with Van Dierendonck, 7/26/04).

Because the factor correlations for the self-administered items we analyzed do not support a multidimensional model of PWB, we were originally surprised that our model fit statistics for the WLS seemed to suggest otherwise. Specifically, given the extremely high factor correlations, we expected that models constraining self-acceptance, purpose in life, and environmental mastery to reflect only a single factor would fit better than the unconstrained models. However, on closer reflection this is not too surprising and does not necessarily indicate that the six dimensions are distinct. First, it is important to remember that our sample size is very large, with over six thousand cases. With a sample this large, almost any deviation will produce a statistically significant difference in chi-square, whether or not the difference is substantially meaningful. One of the reasons we

presented BIC statistics in addition to chi-square was because BIC helps control for sample size. However, Weakliem (1999) suggests that BIC might not sufficiently adjust for exceptionally large sample sizes.

Second, the way that the PWB subscales were originally created may make it impossible to find that the scales/constructs are identical statistically, even if the substantive differences between them are truly negligible. The reason for this is that items were assigned to each subscale if they loaded more highly on that subscale than on any other subscale in Ryff's early studies. Therefore, it is possible that statistically distinct subscales were created, that by design cannot be identical even if the substantive dimensions of PWB they are designed to measure are equivalent. In other words, it is possible that the design of the PWB model capitalized on incidental, but persistent differences among items.

Although we found consistent evidence of high factor correlations using the WLS mail data, NSFH II and MIDUS, we also found evidence of a mode difference in the findings. Specifically, the telephone items in the WLS perform much more like the telephone-administered items analyzed by Ryff & Keyes (1995) rather than like the self-administered items from the WLS, NSFH II, and MIDUS. In addition, the Clarke et al. (2001) study of the Canadian elderly was conducted in person, producing findings that are more similar to those obtained by telephone. What accounts for the consistent differences between self-administered, paper and pencil assessments and those conducted in person or by telephone?

Extensive research has been conducted on mode effects and in general, researchers find that people are less prone to social desirability bias on mail surveys than

in telephone or in-person interviews (Schwartz et al. 1991; Dillman 1991; Krysan et al. 1994; Pruchno & Hayden 2000; Moum 1998). In addition, studies have found that items like PWB, which ascertain psychological characteristics, are especially prone to social desirability bias (Pruchno & Hayden 2000; Moum 1998). Three subscales of Ryff's PWB (purpose in life, self-acceptance and personal growth) have been explored using self-administered, telephone, and in person instruments (Pruchno & Hayden 2000). For all three scales, there were significantly more negative reports on the self-administered instrument compared to the telephone or in person. The authors conclude: "these trends tempt the conjecture that persons participating in the self-administered survey experienced greater privacy and had more time to consider their responses than persons responding to interviewer-assisted modes; therefore, their responses were more accurate. Verification of these speculations awaits further study" (p. 21).

Verification, of course, requires a gold standard such as medical reports or some other non-self-report data. While Ryff's model has not yet been tested in this way, it is informative to look at other survey-based measures of psychological health. Fournier & Kovess (1993) compared mail and telephone mental health instruments and found, as others have, that the mail instrument demonstrated higher rates of mental illness. As a validation check, Fournier & Kovess (1993) sub-sampled a group of respondents for an in-depth interview in order to obtain enough information to make a DSM-III diagnosis. The sample size was small, and therefore it was difficult to find significant effects for specific psychiatric disorders, but they did find that that the kappa of the mail method for "any of these diagnoses" was significantly higher than in the telephone method.

The mode effects that we have reviewed pertain to univariate distributions. Could mode effects also account for the differences in factor structure between self-administration and telephone or in-person administration? It may be that mode simply shifts marginal distributions, but not the associations among variables. Alternatively, shifts in marginal distributions and other mode effects may alter the structure of joint distributions. (De Leeuw et al. 1996). De Leeuw et al. (1996) provide relevant evidence by examining a causal model of loneliness and a confirmatory factor model of subjective well-being. The latter model is of greatest interest to the current project because of the model and content—though Ryff's scales were not used. The authors find strong support for the second hypothesis—that univariate mode differences are magnified in multivariate analyses. They find that the substantive interpretation of the results depends on the mode of administration.

One more issue warrants comment—specifically, we must assess the methodological differences between our study and previous work (for example Ryff and Keyes 1995). Though we believe it is more accurate to treat the data as ordinal and employ weighted least squares estimation with polychoric correlations, it is important to determine how much these methodological differences affect our results and conclusions. As explained in the literature review, Ryff & Keyes (1995) used PRELIS to obtain variance/covariance matrices and an asymptotic covariance matrix defining the variables as continuous; they did not estimate polychoric correlations. Employing their model specifications, we reran their models and were essentially able to reproduce their results.[5] When we analyzed the Ryff & Keyes (1995) data using polychorics and weighted least squares estimation we obtained similar results but, as expected, many of the factor

correlations were larger, as was the chi-square. Likewise when we reran the six factor models on the WLS mail and phone data using the same specifications as Ryff & Keyes (1995), we found somewhat smaller factor correlations and substantially reduced chi-square statistics compared to the findings presented in this paper. For example in the six factor model using the mail data from the WLS, the factor correlations were slightly reduced but many were still over .900; the largest was .954 between purpose in life and self-acceptance. The factor correlations for the phone items were also smaller than when using polychorics with weighted least squares estimation. In short, our conclusions about mode effects and the non-dimensionality of Ryff's model hold even when we don't employ polychoric correlations with weighted least squares estimation. The other important difference between the two methods is that the chi-square statistic using polychorics and weighted least squares estimation is much larger than when polychoric correlations are not used—for example the chi-square for the six factor WLS mail model is 9036, compared to 6580 if the the same model is estimated as in Ryff and Keyes (1995).

Extensive research indicates that items such as, and specifically including Ryff's scales of PWB, are more accurately measured using mail instruments rather than telephone items. Research also suggests that structural models of well-being using different modes can produce substantially different results. In combination, this evidence leads us to believe the consistent results from the WLS mail items, NSFH II, and MIDUS, rather than findings from the WLS telephone items or those of Ryff and Keyes (1995). This also helps explain the seemingly discrepant findings in the literature. Studies finding lower factor correlations (c.f. Ryff & Keyes 1995; Clarke et al. 2001) use

telephone or in-person interviews whereas studies finding less support for Ryff's multi-dimensional model of PWB used self-administered instruments (c.f. Kafka & Kozma 2002 and Van Dierendonck 2004).

Given the consistently high latent variable correlations across mail items in the WLS, MIDUS, and NSFH II, combined with extensive research indicating the greater validity of psychological measures on self-administered surveys, we think that Ryff's model of PWB does not measure six distinct dimensions of PWB. We have not pursued external validation techniques such as looking at age or gender differences in the dimensions for several reasons. First, as we discussed, it is possible that procedures used to choose items may make it impossible to find that these dimensions are not distinct. Second, our study adds to a growing body of literature indicating that Ryff's model is not six-dimensional. Third, because WLS, NSFH II and MIDUS samples are being followed up, we plan to conduct external validity checks in these longitudinal data when they become available.

We applaud the attempt to develop multidimensional models of health—both psychological and physical. As students of physical health have begun to move from life expectancy to quality adjusted life expectancy and other multidimensional concepts, so too should students of mental health. Ryff's model is one such attempt, and we can only applaud this line of research. However, our study adds to a growing body of evidence that Ryff's scale does not measure six distinct dimensions of psychological well-being. While Ryff's scale certainly captures some aspects of positive mental health, we strongly caution against analyses that treat the scale components as if they measured six distinct dimensions of psychological well-being.

NOTES

[1] The other correlations were positive relations & self-acceptance (.52), autonomy & self-acceptance (.52), personal growth & self-acceptance (.48), autonomy & positive relations (.32), environmental mastery & positive relations (.45), purpose in life & positive relations (.55), personal growth & positive relations (.57), environmental mastery & autonomy (.53), purpose in life & autonomy (.46), personal growth & autonomy (.39), and personal growth & environmental mastery (.46).

[2] The correlation of 0.85 means that 85 percent of the variance in these two constructs is in common (Jensen 1971).

[3] In the text the authors actually wrote SPWB for Ryff's Scales of Psychological Well-Being.

[4] Model 4-1 is the same model as Model 3-2. This model is included in both tables for ease of model comparison.

[5] We thank Corey Keyes for providing us with their data and code. We reproduced their factor correlations exactly except we obtain a correlation of .38 for purpose in life and autonomy whereas they reported a correlation of .39.

REFERENCES

Bradburn, N.M. (1969). *The Structure of Psychological Well-Being*. Chicago: Aldine.

Campbell. A. (1981). *The Sense of Well-Being in America*. New York: McGraw-Hill.

Chapman, J.W and Tunmer, W.E. (1995). "Development of young children's reading self-concepts: An examination of emerging subcomponents and their relationship with reading achievement." *Journal of Educational Psychology* 87(1): 154-167

Clarke, P.J., Marshall, V.M., Ryff, C.D. and Wheaton, B. (2001). "Measuring psychological well-being in the Canadian Study of Health and Aging." *International Psychogeriatrics* 13(1): 79-90.

Compton, W.C., M.L. Smith, K.A. Cornish, and D.L. Qualls. (1996). "Factor structure of mental health measures." *Journal of Personality and Social Psychology* 71: 406-413.

DeLeeuw, E.D., G.J. Mellenbergh, and J.J. Hox. (1996). "The influence of data collection method on structural models: A comparison of a mail, a telephone, and a face-to-face survey." *Sociological Methods and Research* 24(4): 442-472.

Dillman, D.A. (1991). "The design and administration of mail surveys." *Annual Review of Sociology* 17: 225-249.

Fournier, L. and K. Kovess. (1993). "A comparison of mail and telephone interview strategies of mental health surveys." *The Canadian Journal of Psychiatry* 38(8): 525-533

Groves, R. (1989). *Survey Errors and Survey Costs*. New York: Wiley.

Hillson, J.M. (1997). "An investigation of positive individualism and positive relations with others: Dimensions of positive personality." Ph.D. Dissertation, University of Western Ontario.

Jensen, Arthur R. 1971. "Note on Why Genetic Correlations Are Not Squared." *Psychological Bulletin* 75(3): 223-24.

Jöreskog, K.G., and Dag Sörbom. (1988). *LISREL 7: A Guide to the Program and Applications*. Chicago, Ill: SPSS Inc.

Jöreskog, K. and D. Sörbom. (1996a). *LISREL 8: User's Reference Guide*. Chicago, IL: Scientific Software International, Inc.

Jöreskog, K. and D. Sörbom. (1996b). *PRELIS 2: User's Reference Guide*. Chicago, IL: Scientific Software International, Inc.

Kafka, G.J. and A. Kozma. (2002). "The construct validity of Ryff's scales of psychological well-being (SPWB) and their relationship to measures of subjective well-being." *Social Indicators Research* 57: 171-190.

Krysan, M., H. Shuman, L.J. Scott, and P. Beatty. (1994). "Response rates and response content in mail versus face-to-face surveys." *Public Opinion Quarterly* 58(3): 381-199.

Marsh, Herbert. (1986). "Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon." *Developmental Psychology* 22(1): 37-49.

McGregor, I. and B.R. Little. (1998). "Personal projects, happiness, and meaning: On doing well and being yourself." *Journal of Personality and Social Psychology* 74: 494-512.

Melnick, S.A and R.K. Gable. (1990). "The use of negative item stems: A cautionary note." *Educational Research Quarterly* 14(3): 31-36.

Moum, T. (1998). "Mode of administration and interviewer effects in self-reported symptoms of anxiety and depression." *Social Indicators Research* 45: 279-318.

Neugarten, B.L., R. Havighurst, and S. Tobin. (1961). "The measurement of life satisfaction." *Journal of Gerontology* 16: 134-143.

Pilotte, W.J. and R.K. Gable. (1990). "The impact of positive and negative item stems on the validity of a computer anxiety scale." *Educational and Psychological Measurement* 50(3): 603-610.

Pruchno, R.A. and J.M. Hayden. (2000). "Interview modality: Effects on costs and data quality in a sample of older women." *Journal of Health and Aging* 12(1): 3-24.

Raftery, Adrian E. (1995). "Bayesian model selection in social research." Pp. 111-195 in *Sociological Methodology, Vol. 15* (Peter V. Marsden, ed.) Cambridge, MA: Blackwell Publishers.

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Ryan, R.M. and E.L. Deci. (2001). "On happiness and human potential: A review of research on hedonic and eudaimonic well-being." *Annual Review of Psychology* 52: 141-66.

Ryff, C.D. (1989a). "Happiness is everything, or is it? Explorations on the meaning of psychological well-being." *Journal of Personality and Social Psychology* 57(6): 1069-1081.

Ryff, C.D. (1989b). "Beyond Ponce de Leon and life satisfaction: New directions in the quest of successful aging." *International Journal of Behavioral Development*, 12, 35-55.

Ryff, C.D., Y.H. Lee, M.J. Essex, and P.S. Schmutte. (1994). "My children and me: Midlife evaluations of grown children and self." *Psychology and Aging* 9: 195-205.

Ryff, C.D. and C.L. Keyes. (1995). "The structure of psychological well-being revisited." *Journal of Personality and Social Psychology* 69(4): 719-727.

Sewell, W.H., R.M. Hauser, K.W. Springer, and T.S. Hauser. (2002). "As We Age: A Review of the Wisconsin Longitudinal Study, 1957-2001." *Research in Social Stratification and Mobility, Vol. 19* (Kevin T. Leicht, ed.). Greenwich, Conn.: JAI Press.

Schmutte, P.S. and C.D. Ryff. (1997). "Personality and well-being: Reexamining methods and meanings." *Journal of Personality and Social Psychology* 73(3): 549-559.

Schwartz, N., F. Strack, H.J. Hippler, and G. Bishop. (1991). "The impact of administration model on response effects in survey measurement." *Applied Cognitive Psychology* 5: 193-212.

Van Dierendonck, Dirk. (2004). "The construct validity of Ryff's Scales of Psychological Well-Being and its extension with spiritual well-being." *Personality and Individual Differences* 36: 629-643.

Waterman, A.S. (1993). "Two conceptions of happiness: contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment." *Journal of Personality and Social Psychology* 64: 678-91.

Weakliem, David (1999). "A critique of the Bayesian Information Criterion in Model Selection." *Sociological Methods and Research* 27(3): 359-397.

Table 1: Correlations among Latent Constructs in 3-indicator, 6-factor model for a nationally representative sample over 25 years old (Ryff and Keyes (1995)

	AUT	ENV	GROW	RELAT	PURP	ACC
AUT	1.000					
ENV	0.59	1.000				
GROW	0.51	0.56	1.000			
RELAT	0.24	0.65	0.31	1.000		
PURP	0.39	0.38	0.64	0.30	1.000	
ACC	0.53	0.85	0.53	0.65	0.55	1.000

Note: N=928, aut=autonomy, env=environmental mastery, grow=personal growth, relat=positive relations, purp=purpose in life, acc=self-acceptance

Table 2a: PWB Items in the WLS, MIDUS, and NSFH II

1.   WLS Graduate Mail Survey

Please read the statements below and decide the extent to which each statement describes you	*Circle the number that best describes your agreement or disagreement with each statement*					
	Agree			Disagree		
	Strongly	Moderately	Slightly	Strongly	Moderately	Slightly
	1	2	3	4	5	6

```
  2.  WLS Graduate telephone Interview
"The next section provides several statements that people might use
     to describe themselves.  Please tell us whether you agree or
     disagree with the following statements."

After replying to the question the respondent was then asked:

"Is that strongly, moderately, or slightly ?"

The responses are coded:
01    AGREE STRONGLY
02    AGREE MODERATELY
03    AGREE SLIGHTLY
04    NEITHER AGREE NOR DISAGREE
05    DISAGREE SLIGHTLY
06    DISAGREE MODERATELY
07    DISAGREE STRONGLY

  3.  MIDUS

"Please indicate how strongly you agree or disagree with each of the
following statements."
              Agree                             Disagree
   ────────────────────────────────    ────────────────────────────────

Strongly    Some       A little    Don't    A little    Some       Strongly
            What                   Know                 What

   1          2          3          4          5          6          7

  4.  NSFH II

"Please indicate how much you agree or disagree with each of the
     following statements:

          (circle your answer to each question)"
  _____

 STRONGLY    MODERATELY    SLIGHTLY    SLIGHTLY    MODERATELY    STRONGLY
  AGREE       AGREE         AGREE       DISAGREE    DISAGREE      DISAGREE

    1           2             3           4           5             6
```

Table 2b. Wording and order of PWB items in the WLS, MIDUS, and NSFH II

Items that are asterisked vary somewhat across different studies. The differences are italicized to help with comparisons. The order of the items is indicated by the numbers in the columns and the column numbers correspond to the following studies:

1. WLS graduate mail survey
2. WLS graduate telephone interview
3. MIDUS
4. NSFH II

Item	1	2	3	4
Autonomy Items				
My decisions are not usually influenced by what everyone else is doing.	1			
*I have confidence in my opinions even *if* they are *contrary to the general consensus.*	7			
*I have confidence in my own opinions even *if* they are *different from the way most other people think*.			17	6
I tend to worry about what other people think of me.	8			
I often change my mind about decisions if my friends or family disagree.	15			
I am not afraid to voice my opinions, even when they are in opposition to the opinions of most people.	22			
Being happy with myself is more important to me than having others approve of me.	29			
It's difficult for me to voice my opinions on controversial matters.	36			
I tend to be influenced by people with strong opinions.		1	15	5
*I judge myself by what I think is important, not by what others think is important.		10		
*I judge myself by what I think is important, not by *the values* of what others think is important.			18	18
Environmental Mastery				
I am good at juggling my time so that I can fit everything in that needs to get done.	2			
I often feel overwhelmed by my responsibilities.	9			
*I am *quite* good at managing the *many* responsibilities of *my* daily life.	14			15
*I am good at managing the responsibilities of daily life.			9	
I do not fit very well with the people and community around me.	16			
I have difficulty arranging my life in a way that is satisfying to me.	23			
I have been able to create a lifestyle for myself that is much to my liking.	30			
I generally do a good job of taking care of my personal finances and affairs.	37			
In general, I feel I am in charge of the situation in which I live.		2	8	13
The demands of everyday life often get me down.		6	4	3
Personal Growth				
I am not interested in activities that will expand my horizons.	3			
I have the sense that I have developed a lot as a person over time.	10			
When I think about it, I haven't really improved much as a person over the years.	17			
*I think it is important to have new experiences that challenge how *I* think about *myself* and the world.	21		12	
*I think it is important to have new experiences that challenge how *you* think about *yourself* and the world.				2
I don't want to try new ways of doing things -- my life is fine the way it is.	24			
I do not enjoy being in new situations that require me to change my old familiar ways of doing things.	31			

There is truth to the saying you can't teach an old dog new tricks.	38			
*For me, life has been a continuous process of learning, changing, and *growing*.		7		
*For me, life has been a continuous process of learning, changing, and *growth*.			11	17
I gave up trying to make big improvements or changes in my life a long time ago.		11	14	12
Positive Relations				
I don't have many people who want to listen when I need to talk.	4			
I enjoy personal and mutual conversations with family members and friends.	11			
I often feel lonely because I have few close friends with whom to share my concerns.	18			
It seems to me that most other people have more friends than I do.	25			
People would describe me as a giving person, willing to share my time with others.	28		13	11
Most people see me as loving and affectionate.	32			
I know I can trust my friends, and they know they can trust me.	39			
Maintaining close relationships has been difficult and frustrating for me.		3	6	4
I have not experienced many warm and trusting relationships with others.		8	16	8
Purpose in Life				
I enjoy making plans for the future and working to make them a reality.	5			
My daily activities often seem trivial and unimportant to me.	12			
I am an active person in carrying out the plans I set for myself.	19			
I tend to focus on the present, because the future nearly always brings me problems.	26			
I don't have a good sense of what it is I'm trying to accomplish in life.	33			
I sometimes feel as if I've done all there is to do in life.	35		10	14
I used to set goals for myself, but that now seems like a waste of time.	40			
Some people wander aimlessly through life but I am not one of them.		4	3	16
I live life one day at a time and don't really think about the future.		9	7	7
Self-Acceptance				
I feel like many of the people I know have gotten more out of life than I have.	6			
In general, I feel confident and positive about myself.	13			
When I compare myself to friends and acquaintances, it makes me feel good about who I am.	20			
My attitude about myself is probably not as positive as most people feel about themselves.	27			
I made some mistakes in the past, but I feel that all in all everything has worked out for the best.	34			
The past had its ups and downs, but in general, I wouldn't want to change it.	41			
In many ways, I feel disappointed about my achievements in life.	42		5	10
*When I look at *the story of* my life, I am pleased *with* how things have turned out.		5		
*When I look at *the story of* my life, I am pleased *about* how things have turned out.				9
*When I look at *the story of* my life, I am pleased *with* how things have turned out *so far*.			2	
*I like most *parts* of my personality.			1	1
*I like most aspects of my personality.		12		

Table 3: Indices of Fit Based on Weighted Least Squares—WLS Graduates
Prior to Accounting for Methodological Effects

	Chi-square	DF	BIC
1. Single Factor Model	10865.08	819	3702.55
2. Six Factor Model w/all psi correlated	9035.73	804	2004.39
3. 2nd Order, Six Factor Model	9199.43	813	2089.39
4. 2nd Order Factor Model w/variance of env, purp, acc equal to zero	9342.91	816	2206.62
5. 2nd Order Factor Model w/variance of env, purp, acc, & grow equal to zero	9481.17	817	2336.14
6. 2nd Order Factor Model w/env, purp, acc loading on 1 factor and aut, grow, rela on another factor	9327.29	815	2199.76
7. 2nd Order Factor Model w/env, purp, acc, growth on 1 factor and aut & relations on another factor	9480.34	816	2334.05

Table 4: Indices of Fit Based on Weighted Least Squares—WLS Graduates Accounting for Methodological Effects

	Chi-square	DF	BIC
1. Six Factor Model w/all psi correlated	9035.73	804	2004.39
2. Six Factor Model w/correlated neg factor	7539.77	776	753.31
3. Six Factor Model w/correlated errors for all adjacent questions	8141.15	763	1468.38
4. Six Factor Model w/correlated errors for all adjacent questions & a correlated neg factor	6660.05	735	232.15
5. Six Factor Model w/correlated errors for all adjacent questions & a correlated neg factor and 3 additional error correlations	5953.10	732	-448.56

Table 5: Correlations Among Latent Constructs for the WLS mail survey

1. Correlations among Latent Constructs in WLS mail 6-factor model

	AUT	ENV	GROW	RELAT	PURP	ACC
AUT	1.000					
ENV	0.784	1.000				
GROW	0.815	0.911	1.000			
RELAT	0.735	0.915	0.866	1.000		
PURP	0.793	0.958	0.942	0.895	1.000	
ACC	0.825	0.970	0.951	0.903	0.976	1.000

2. Correlations among Latent Constructs in Methodologically Corrected WLS mail 6-factor model

	AUT	ENV	GROW	RELAT	PURP	ACC
AUT	1.000					
ENV	0.771	1.000				
GROW	0.791	0.915	1.000			
RELAT	0.721	0.925	0.866	1.000		
PURP	0.780	0.958	0.938	0.887	1.000	
ACC	0.808	0.970	0.943	0.900	0.970	1.000

Note: aut=autonomy, env=environmental mastery, grow=personal growth, relat=positive relations, purp=purpose in life, acc=self-acceptance

Table 6: Correlations Among Latent Constructs for NSFH II and MIDUS

1. Correlations among Latent Constructs in NSFH II 6-factor model

	AUT	ENV	GROW	RELAT	PURP	ACC
AUT	1.000					
ENV	0.785	1.000				
GROW	0.765	0.809	1.000			
RELAT	0.647	0.781	0.724	1.000		
PURP	0.757	0.820	0.981	0.727	1.000	
ACC	0.714	0.933	0.803	0.823	0.764	1.000

2. Correlations among Latent Constructs in MIDUS 6-factor model

	AUT	ENV	GROW	RELAT	PURP	ACC
AUT	1.000					
ENV	0.663	1.000				
GROW	0.510	0.749	1.000			
RELAT	0.476	0.677	0.661	1.000		
PURP	0.541	0.756	0.828	0.729	1.000	
ACC	0.581	0.858	0.779	0.777	0.871	1.000

Note: aut=autonomy, env=environmental mastery, grow=personal growth, relat=positive relations, purp=purpose in life, acc=self-acceptance

Table 7: Correlations Among Latent Constructs for WLS graduate telephone items

Correlations among Latent Constructs in WLS telephone 6-factor model

	AUT	ENV	GROW	RELAT	PURP	ACC
AUT	1.000					
ENV	0.633	1.000				
GROW	0.442	0.692	1.000			
RELAT	0.321	0.649	0.554	1.000		
PURP	0.332	0.668	0.608	0.501	1.000	
ACC	0.562	1.032	0.809	0.730	0.713	1.000

Note: aut=autonomy, env=environmental mastery, grow=personal growth, relat=positive relations, purp=purpose in life, acc=self-acceptance