Social Indicators Research (2006) 79: 171–188
DOI 10.1007/s11205-005-4098-6

BRIAN E. PERRON

A CRITICAL EXAMINATION OF THE ENVIRONMENTAL MASTERY SCALE*

(Accepted 16 October 2005)

ABSTRACT. The Psychological well being Inventory (PWBI) is a multidimensional instrument that has enjoyed widespread use in a variety of research initiatives, from small-scale studies to national surveys. Recent empirical investigation of the measure has raised questions about its validity. This study examines the factorial validity of the Environmental Mastery Scale of the PWBI, a construct that receives much attention in mental health research. The results of a confirmatory factor analysis did not support the unidimensional factor structure of the measure. Correlated uniqueness models were also examined, which did not reveal evidence of a method effect. There are problems with the measure at the item level, which also raises significant questions about the underlying theory of its parent measure.

KEY WORDS: environmental mastery, psychological well-being inventory, factorial validity, confirmatory factor analysis

INTRODUCTION

The concept of environmental mastery was first introduced by Phillips (1961) who described it as an instinct that progresses through five stages: isolation, dependency, autonomy, cooperation, and independence. However, the concept did not receive significant attention until it was re-introduced by Ryff (1989) in the Psychological Well-being

*This research uses data from the Wisconsin Longitudinal Study (WLS) of the University of Wisconsin-Madison. Since 1991, the WLS has been supported principally by the National Institute on Aging (AG-9775 and AG-21079), with additional support from the Vilas Estate Trust, the National Science Foundation, the Spencer Foundation, and the Graduate School of the University of Wisconsin-Madison. A public use file of data from the Wisconsin Longitudinal Study is available from the Wisconsin Longitudinal Study, University of Wisconsin-Madison, 1180 Observatory Drive, Madison, Wisconsin 53706 and at http://www.ssc.wisc.edu/~wls/data/. The opinions expressed herein are those of the authors.

Inventory (PWBI). According to Ryff's formulation, environmental mastery is defined as the "capacity to manage effectively one's life and surrounding world" (Ryff and Keyes, 1995, p. 720). Environmental mastery is considered an important psychological resource and is gaining increased attention in the health and social science research.

The measurement of environmental mastery has been facilitated by the environmental mastery scale, which is part of a multidimensional instrument called the PWBI. In a review of the literature by Windle and Woods (2004), the environmental mastery scale (hereafter referred to as the EMS) was found to mediate the potential negative impact of community relocation for older women, contribute to the absence of mood disorder in rheumatoid arthritis sufferers, and predict globally fatigue and fatigue-related distress among people multiple sclerosis. Windle and Woods (2004) also reported on a mediation model that demonstrated that "environmental mastery is the key to experiencing life satisfaction in the midst of adversity" (p. 595) for older adults living in the community. September et al. (2001) found that environmental mastery was a predictor of confidence and imposter feelings among Canadian undergraduate students. Seltzer et al. (2004) examined environmental mastery among parents of children with mental health problems and developmental disabilities (Seltzer et al. 2004). They found that higher levels of environmental mastery were associated with accommodative coping.

Additional associations among environmental mastery and other variables can be found in the extensive research that has utilized the PWBI. In fact, according to the Social Science Index, the PWBI has been cited in almost 400 research articles. It has also been included in national surveys, including Midlife in the United States (MIDUS), National Survey of families and Households II (NSFH-II), Wisconsin Longitudinal Survey (WLS), and the Canadian Study of Health and Aging (see Springer and Hauser, forthcoming).

To date, there are only a handful of studies that have tested the factor structure of the PWBI (see Ryff, 1989), but no studies have looked specifically at the reliability and validity of the EMS or its other scales. Moreover, of the studies that have been conducted, none have been able validate the second-order factor structure since the initial validation study by Ryff and Keyes (1995). This raises questions about the reliability and validity about the first-order factors (or scales) and whether they can be used as sound measures of their

respective constructs. To help fill this gap in knowledge, this study examines the factorial validity of the environmental mastery factor.

BACKGROUND

Psychometric Properties

As indicated, the EMS is one of the six scales of the PWBI. The PWBI is composed of six self-report scales[1] that measure environmental mastery, self-acceptance, positive relations, psychological growth, and purpose in life. Ryff (1989) hypothesizes that these factors are unified by a second-order factor called well-being. The full version of the PWBI contains 20 items per scale, thus producing a 120-item inventory. Each scale includes both positively and negatively phrased items, and responses are made on a six-point Likert type scale (1 = strongly disagree, and 6 = strongly agree). Positively phrased items are reverse scored and then summed to produce scale. A composite well-being score is computed by summing the six scale scores.

Based on a sample of 321 men and women using multiple age groups, Ryff (1989) reported that the internal consistency (α) of the subscales ranged from 0.86 and 0.93, and test–retest reliability ranged from 0.81 and 0.88. The EMS had the lowest reliability estimates (internal consistency and test–retest) relative to the other five scales. Preliminary evidence for validity was established by finding correlations that were significant and in the predicted direction with other measures of positive functioning (e.g., life satisfaction, affect balance, self esteem) and negative functioning (e.g., depression, chance, powerful others). Regarding the inter-factor correlations, environmental mastery correlated highly with self-acceptance ($r = 0.76$) and purpose in life ($r = 0.72$), which suggests overlap in the constructs.

Factorial validity of the PWBI was reported in a later study by Ryff and Keyes (1995). This study was part of a larger survey study called the Midlife in the United States Pretest, which included a national probability sample of 1108 adults aged 25 and older. Due to resource constraints only 18 total items (three items per factor) from the inventory were tested. Ryff and Keyes (1995) did not report on how the 18 items were selected – that is, whether they were selected randomly or purposively. The administration of the PWBI in the

MIDUS was based on telephone interviews using an unfolding technique. The unfolding technique is where respondents are first asked whether they agreed or disagreed with the item, and then asked to what extent (i.e., strongly, moderately, slightly).

Using CFA Ryff and Keyes (1995) tested five different models and found that the six-factor model with a second-order well-being factor had the best fit with the data. While none of the models had a non-significant chi-square value, Ryff and Keyes (1995) argued that this was most likely due to the large sample size. The evidence for a good fit was based on the Bayesian information criterion (BIC), which had a value of −166.04. On the BIC, positive values represent a poor model fit. They also reported that the internal consistency (α) of the subscales ranged from 0.33 to 0.56. Ryff and Keyes (1995) did not regard these "low to moderate" (p. 721) estimates as problematic because each factor was measured with only three items. Similar to the earlier study by Ryff (1989), the EMS in this study exhibited a high correlation with self-acceptance ($r = 0.85$). Ryff and Keyes (1995) argued that it is not likely these are overlapping constructs as they exhibit different age profiles and are well grounded theoretically.

Since this initial validation study (Ryff and Keyes, 1995), there have been no studies that have successfully replicated the PWBI's factor structure, a surprising finding given its widespread use. Springer and Hauser (2004) conducted an exhaustive study of the factorial structure of the PWBI using multiple data sources, including the NSFH-II, MIDUS, and the WLS. These surveys incorporated different versions of the PWBI and different modes of administration. Using CFA, Springer and Hauser were unable to establish a good fit with either the first- or second-order six-factor structure models among any of the data sets. They also reported that the EMS correlated highly with the other factors. In the WLS, the EMS correlated strongly with self-acceptance ($r = 0.971$), purpose in life ($r = 0.958$), and personal growth ($r = 0.908$). In the NSFH-II, the EMS correlated strongly with self-acceptance ($r = 0.933$). And in the MIDUS, the EMS was again correlated strongly with self-acceptance ($r = 0.858$).

Kafka and Kozma (2002) examined the dimensionality of the PWBI among a sample of undergraduate psychology students. A principal component analysis with varimax rotation extracted 15 factors, each with an eigenvalue greater than 1. When the analysis was restricted to a six-factor solution, the resultant factor structure

differed significantly from the six-factor structure hypothesized by Ryff (1989). In particular, the items that purportedly measure environmental mastery were dispersed among three different factors. However, these data should be interpreted with caution, as a principal components analysis is more aptly suited for data reduction rather than analysis of factor structure.

Van Dierendonck (2004) also tested the factorial structure of the 3-item, 9-item, and 14-item versions of the PWBI among college students and professionals of a diverse occupational background. Using CFA, Van Dierendonck (2004) found evidence of the six-factor second-order model using the 3-item version. However, the internal consistency estimates were too low for the inventory to be deemed reliable. The estimates ranged from $\alpha = 0.17$ to 0.58 (EMS, $\alpha = 0.51$ to 0.58).

Method Effects

While method effects are often understood in terms of the format of the measure (e.g., self-report, interview, observation), it can also include the source of scores, such as the respondent (Kline, 1998). More specifically, according to Marsh (1996), "When psychosocial rating scales contain positively and negatively worded items, factor analyses of responses reveal apparently distinct factors reflecting the positive and negative items, respectively" (p. 810). This raises the question as to whether the factor structure is substantively meaningful or an artifact of the response style associated with the measure. As the PWBI contains multiple psychosocial scales, all of which are based on positively and negatively worded items, it is necessary to consider the presence of a method effect.

In the initial validation study, Ryff and Keyes (1995) tested two models in order to discern a method effect, each "suggesting that respondents answer questions to portray a positive self-image" (p. 722). The first model was a single-factor negative item artifact model. This model posited a single dimension of well-being with negatively worded items loading on the artifactual dimension. The second model was a two-factor solution composed of negative and positive artifacts. This model indicated that respondents agreed with the positively phrased items and disagreed with negatively phrased items.

These models did not lead to a good fit with the data, and the fit relative to the second-order model was poor. However, these two models are not definitive evidence for determining the presence or absence of a method effect. More specifically, this is a multidimensional measure, so it is possible that a method effect might be more pronounced on certain latent factors and not others. Testing for method effects among only certain factors was not possible because in this study each factor was measured with only three items. This analysis also assumes that the models are invariant across populations. Marsh (1996) found evidence for a more pronounced method effect among persons with a lower level of literacy. Moreover, this initial validation study utilized a telephone interview with an unfolding technique, whereas the most common usage of the inventory is self-administration. It is not known whether the mode of administration can influence the degree of a method effect.

In the comprehensive work of Springer and Hauser (2004), they tested for the presence of a method effect by allowing all the scale items to load on their corresponding well-being dimensions. Additionally, all the negatively phrased items were also assigned to load on another latent variable representing a negative factor. This resulted in a significantly reduced chi-square value in relation to the second-order factor model. However, neither model exhibited an overall good fit with the data. A limitation of this analysis is the assumption that the method effect is present in each of the latent factors. Additionally, there is an assumption of model invariance across sub-groups. It is unlikely that the 6000 respondents in their sample are responding to the same underlying causal mechanisms.

Models

Based on this review of the literature, there is surprisingly little evidence for the reliability and validity of the EMS despite its widespread use. The available evidence on the psychometric properties of its parent scale the PWBI is suggestive of its limitations. The evidence of a method effect in this measure remains ambiguous. To address these gaps in knowledge, this study evaluates four *a priori* models derived from the literature, which are depicted graphically in Figure 1.

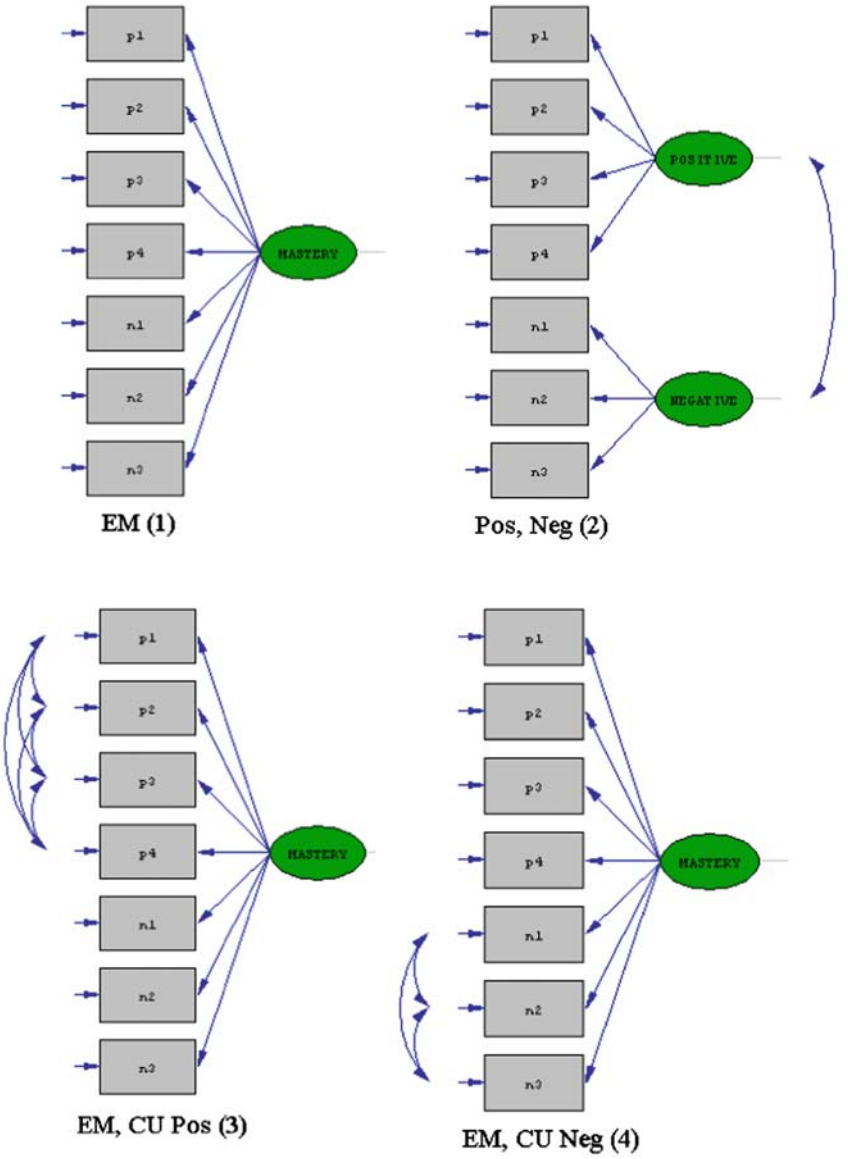

Figure 1. Confirmatory factor analysis models.

The first model tests the hypothesis that the EMS is a unidimensional measure of environmental mastery. Thus, all items are restricted to load on a single environmental mastery factor (EM). This model is isomorphic with the factor structure of the PWBI subscales as described by Ryff (1989) and Ryff and Keyes (1995). It is

also the same factor structure in prior studies that have utilized the EMS.

The other three models are derived from strategies used to test for method effects in the Rosenberg Self-esteem Scale (see Marsh, 1996; Dunbar et al., 2000; Wang et al., 2001; Motl and DiStefano, 2002; Horan et al., 2003). More specifically, the second model is a two-factor structure (Pos, Neg). All positively worded items are hypothesized to load on the positive factor, and all negatively worded items are expected to load on the negative factor. This model suggests that respondents show a tendency to agree with positively worded items and disagree with negatively worded items. Model 3 (EM, CU Pos) posits a single environmental mastery factor with correlated uniqueness (i.e., measurement error) among the positively worded items. Model 4 (EM, CU Neg) posits a single environmental mastery factor with correlated uniqueness among the negatively worded items.

METHOD

Sample

The data for this study are from the Wisconsin Longitudinal Survey (WLS), which is part of a long-term study by the University of Wisconsin-Madison (Hauser et al., 1992/93). The WLS is based on a random sample of men and women who graduated from high school in Wisconsin during 1957. In 1992 and 1993, a mail survey was sent to 8493 study participants in the original cohort. The response rate for this survey was 81% ($n = 6875$). The approximate age of the respondents at the time the mail survey was completed was 54, and the proportion of females in the sample was slightly higher than males (53.56% and 46.44%, respectively). According to the WLS study description, only a ''there are only a handful of African-American, Hispanic, or Asian persons in the sample'' (Hauser et al., 1992/93).

Because environmental mastery is considered an important psychological resource, there is much appeal for its use in clinical research. For example, as previously indicated, Windle and Woods (2004) found environmental mastery to play a key role in to experiencing life satisfaction in the midst of adversity, in addition to being

significantly associated with other clinical factors. Establishing the psychometric properties among a clinical population contributes to the practical utility of the measure. Thus, the factorial validity of the measure is tested among adults with depression, which further contributes to the homogeneity of the sample. The underlying theoretical framework of this measure is also clearly representative of this population (see Ryff, 1989; Ryff and Keyes, 1995).

The mail survey of the WLS also included the Center for Epidemiological Studies Depression Scale (CES-D), which was used to identify this sample. The CES-D is a widely used and psychometrically sound depression-screening instrument (Himmelfarb and Murrell, 1983; Callahan and Wolinsky, 1994; Lewinsohn et al., 1997; Radloff and Teri, 1986), which is composed of 20 depression indicator items. Respondents use a four-point rating system to identify the number of days they have experienced each depression indicator. Responses are summed to produce an overall depression score, which has a theoretical range of 0–80 on this scale. A score of 15–21 indicates a mild to moderate depression, and a score of 22 or greater indicates probable major depression (Measurement Excellence and Training Resource Information Center, n.d.). This study used the cutoff score of 22 for identifying subjects with depression. Respondents who provided responses on all 20 items ($n = 6349$) and had a score of 22 or greater were included in the study. These criteria produced an effective sample size of 602 subjects (60.8% female, 39.2% male).

Environmental Mastery Subscale

The WLS used a seven-item version of the EMS, which are summarized in Table I. Four of the items are positively phrased, and three of the items are negatively phrased. The response format maintains the original six-point Likert-type format (1 = strongly agree, 6 = strongly disagree). The four positive items were reverse scored before performing any analyses.

Analytic Strategy

SAS version 8.0 was used for drawing the subsample from the WLS dataset. Confirmatory factor analysis was performed using LISREL version 8.54s and the SIMPLIS command language. Because these

TABLE I

Environmental mastery subscale (EMS)

Item
P1. I am good at juggling my time so that I can fit everything in that needs to get done.*
P2. I am quite good at managing the many responsibilities of my daily life.*
P3. I have been able to create a lifestyle for myself that is much to my liking.*
P4. I generally do a good job of taking care of my personal finances and affairs.*
N1. I often feel overwhelmed by my responsibilities.
N2. I do not fit very well with the people and community around me.
N3. I have difficulty arranging my life in a way that is satisfying to me.

Note: *Reverse-scored item.

data were at the ordinal level and non-normal, fully weighted least squares estimation was performed with polychoric correlations and an asymptotic covariance matrix. The polychoric correlations and covariance matrix were computed using PRELIS version 2.54s and are reported in Table II. After listwise deletion, the sample size was reduced to $n = 585$.

The adequacy of the model fit was determined by multiple fit indices. This includes the chi-square (χ^2) test statistic, which is the most commonly used fit statistic. A non-statistically significant value indicates that the sample covariance matrix and the model-implied covariance matrix are similar (Schumacker and Lomax, 2004). Because this statistic tends to be inflated with large samples, a chi-square value to degrees of freedom ratio ($\chi^2/\mathrm{df}) < 2.0$ is also an

TABLE II

Polychoric correlation matrix

Item	P1	P2	P3	P4	N1	N2	N3
P1	–						
P2	0.661	–					
P3	0.301	0.431	–				
P4	0.330	0.507	0.336	–			
N1	0.089	0.067	0.069	0.022	–		
N2	0.119	0.239	0.336	0.170	0.116	–	
N3	0.212	0.315	0.425	0.255	0.276	0.314	–

indication of a good fit. Other indices, including the normed fit index (NFI), root-mean-square error of approximation (RMSEA), comparative fit index (CFI), and incremental fit index (IFI) are also considered. The NFI is a rescaled chi-square value (0 = no fit, 1.0 = perfect fit) that is based on a comparison of the proposed model with a model that contains no relations, or a *null* model (Raykov and Marcoulides, 2000). A value >0.95 on the NFI suggests a good model fit. The RMSEA includes an adjustment for degrees of freedom, and a value <0.05 suggests and acceptable model fit. Values >0.95 on the CFI and IFI also suggest a good model fit.

RESULTS

The fit statistics for all four models are summarized in Table III. The single factor model had a very poor fit with the data, as evidenced by a highly significant chi-square value ($p < 0.0001$), a large chi-square to degrees of freedom ration ($\chi^2/df = 7.434$), and an RMSEA value >0.05. Relative to the other models, this single-factor model exhibited the largest ratio of chi-square to degrees of freedom and the lowest NFI, CFI, and IFI values. All factor loadings are significant, ranging from 0.134 to 0.917 (see Table IV). The factor loadings for the negatively phrased items were significantly lower than those for the positively phrased items.

The second model (Pos, Neg) is a two-factor solution, where positively phrased items were assigned to load on the positive factor, and negatively phrased items were assigned to load on the negative

TABLE III

Fit indices for environmental mastery subscale

Model	χ^2	df	χ^2/df	NFI	RMSEA	CFI	IFI
EM (1)	104.07	14	7.434	0.888	0.105	0.901	0.902
Pos, Neg (2)	81.967	13	6.305	0.912	0.095	0.924	0.925
EM, CU Positive (3)	22.304	8	2.788	0.976	0.055	0.984	0.985
EM, CU Negative (4)	62.855	11	5.714	0.932	0.090	0.943	0.944

Note: χ^2 = minimum fit function chi square. df = degrees of freedom. RMSEA = root mean square error of approximation. NFI = normed fit index. CFI = comparative fit index. IFI = incremental fit index.

TABLE IV

Parameter estimates for models of environmental mastery

Corr Error	Model												
	EM (1)		Pos, Neg (2)			EM, CU Pos (3)					EM, CU Neg (4)			
	EM	Error	Pos	Neg	Error	EM	Error	Corr Error			EM	Error	Corr Error	
Factor loadings								P1	P2	P3			N1	N2
P1	0.728	0.471	0.745	–	0.445	0.270	0.927				0.734	0.462		
P2	0.917	0.160	0.900	–	0.191	0.439	0.807	0.553			0.905	0.181		
P3	0.610	0.628	0.570	–	0.676	0.601	0.639	0.132	0.150		0.608	0.631		
P4	0.570	0.676	0.542	–	0.707	0.359	0.872	0.241	0.347	0.101	0.567	0.678		
N1	0.134	0.982	–	0.357	0.873	0.300	0.911				0.094	0.991		
N2	0.365	0.867	–	0.356	0.874	0.478	0.771				0.330	0.891	0.072	
N3	0.532	0.717	–	0.776	0.402	0.757	0.427				0.449	0.798	0.251	0.093

factor. There was a slight decreased in the chi-square value and its ratio with degrees of freedom (χ^2/df = s6.305), but this can be attributed to the addition of another parameter estimate. The other indices also suggest a poor model fit, with an RMSEA value > 0.05 and values < 0.95 on NFI, CFI, and IFI. The correlation between the positive and negative factors were not significant. Like the first model, the positively phrased items in this model had higher factor loadings than the negatively phrased items.

The third model (EM, CU Positive) allowed the error variances on the positively phrased items to correlate, which resulted in a significant improvement in the model fit relative to the other models. All the correlations among the error variances were significant, ranging from $r = 0.101$ to 0.553. This model had a significant chi-square value, and the RMSEA value was slightly above the 0.05 cutoff value (0.055). However, the chi-square to degrees of freedom ratio reduced to 2.788. The NIF, CFI, and IFI values were all in an acceptable range. The factor loadings for the individual items were significant but did not show a notable improvement relative to the first two models. Items P1 and N1 had especially low factor loadings (0.27, 0.30, respectively), which makes the fit of the model ambiguous.

The fourth model (EM, CU Neg) posited a single environmental mastery factor with correlated uniqueness among the negatively phrased items. This model had a better fit than the single factor model without correlated uniqueness. However, this model had an overall poor fit with the data. The chi-square statistic was highly significant, the chi-square to degrees of freedom ratio was large (χ^2/df = 5.714), and the RMSEA value also exceeded 0.05. The other fit indices (i.e., NFI, CFI, and IFI) had values < 0.95. The inter-correlations among the error variances were all significant but relatively weak, ranging from $r = 0.072$ to 0.251.

The reliability estimates of the positively worded were stronger than the negatively phrased items in all models except the third (EM, CU Pos; see Table IV). Thus, an *a posteriori* analysis was conducted to determine whether the negatively phrased items were causing the model to have a poor fit with the data. This model was evaluated without correlated uniqueness and produced a relatively poor fit. A number of fit indices were within the acceptable range (RMSEA = 0.077, NFI = 0.988, CFI = 0.990, IFI = 0.990). However,

the chi-square value was significant, and the chi-square to degrees of freedom ration was very large ($\chi^2/df = 4.5$).

DISCUSSION

This study examined the dimensionality of the EMS, the most commonly used measure of the environmental mastery construct. In this study, there was no unambiguous evidence to suggest that the EMS is a reliable and valid unidimensional measure of environmental mastery. The single-factor model resulted in the poorest fit with the data relative to the models that included specifications for a method effect. The single-factor model with correlated uniqueness among the positive items exhibited the best fit, but the fit indices were marginal and some of the factor loadings among the measured variables were weak. Similarly, Springer and Hauser (2004) found a better fit with the data by including a method effect, but the overall model did not exhibit have a good fit.

This study was based on a relatively homogenous and narrow sample – i.e., white, non-Hispanic, educated adults aged approximately 54 years with depression – which suggests the limitations in the generalizability in the findings. However, it is important to note that the underlying theoretical framework of this measure is clearly representative of this population (see Ryff, 1989; Ryff and Keyes, 1995). In other words, if the measurement model does not contain specification errors, it should be expected to have a good fit with the data among this population.

In a confirmatory factor analysis of the PWBI, Ryff and Keyes (1995) argued that the highly significant chi-square values for the models were most likely due to a large sample size ($N = 1108$). The current analysis is based on a sample size that is nearly half of Ryff and Keyes (1995). Because the chi-square values for these models were also highly significant, the chi-square global fit test should not be overlooked due to sample size. The chi-square global fit test provides a stringent test and important evidence regarding model fit. Thus, the non-significant chi-square values should motivate an exhaustive search for specification errors.

Van Dierendonck (2005) suggests using subscales of the PWBI that are composed of six, seven, or eight items. This suggestion is not

warranted based on the results of this study, as the seven-item version of the EMS did not result in a good fit with the single-factor model. Moreover, eliminating the items with the weakness reliability coefficients failed to produce a good model fit. Moreover, Bollen and Lennox (1991) argue that "equally reliable effect indicators of a *unidimensional* concept are interchangeable" (p. 308; italics used by original authors). Thus, any combination of at least four items from the 20 items on the full version should be able to reproduce its single-factor structure.

From the results of this study, the following question can be raised: Do the scale items adequately tap the environmental mastery construct, or is the underlying theory inadequately specified? An analysis of the phrasing of the EMS items suggests there are problems at item level, which could influence the hypothesized dimensionality of the scale. The hypothesized structure of the PWBI implies that the scale items must be effect indicators of their corresponding latent variables (see Ryff 1989; Ryff and Keyes, 1995). Effect indicators depend on the latent variable, where as cause indicators influence the latent variable (Bollen and Lennox, 1991). In the EMS, item N1 reads, "I often feel overwhelmed by my responsibilities." This item appears to be bi-directional indicator rather than a unidirectional effect indicator. More specifically, a reduction in environmental mastery could cause somebody to feel overwhelmed, which implies that it is an effect indicator. Feeling overwhelmed by responsibilities could also result in a reduced sense of environmental mastery. This implies that it is a cause indicator. Other items on the EMS suggest the presence or absence of a specific skill set that is correlated with environmental mastery (e.g., P1, P4, and N3). As these skills are specified as effect indicators, a change in environmental mastery causes a change in the skills. Most likely, it is a change in the skills that cause a change in environmental mastery, which means they are cause indicators.

Another problem at the item level is that some item stems are not neutral. Item N1 reads: "I often feel overwhelmed by my responsibilities." Variability on this item reflects the extent to which the respondent feels overwhelmed. An endorsement of feeling overwhelmed is not likely made until an emotional threshold is crossed, which might be an extreme state for the respondent. Compare this with item P2, which reads: "I am quite good at managing the many

responsibilities of my daily life.'' This reflects a less extreme state, which allows a greater range of variability to occur. A larger problem is that the two items do necessarily move together (after reverse-scoring P2) with changes in the latent variable. That is, a respondent could exhibit a wide range of values on P2 while maintaining a stable score on N1. Interestingly, item N1 exhibited the lowest factor loading across in all the models and had consistently high error variances. Moreover, this item had very low correlations with the other scale items, ranging from $r = 0.022$ to 0.276 (see Table II).

Item P4 also shows some potential problems due to its double-barreled phrasing: ''I generally do a good job of taking care of my personal finances *and* affairs'' (italics added for emphasis). In this case, 'affairs' can be interpreted as being related or unrelated to finances. A different interpretation could yield significantly different responses. Moreover, it is unlikely that personal finances and affairs unrelated to finances are highly correlated.

Problems at the item level undoubtedly prevent the scale from adequately tapping the underlying latent variable. However, the item-level problems may not fully account for the inability of the measure to produce a single-factor solution. It is important to also consider the underlying theory of well-being that informed the formulation of the indicator items, as a clear, well-specified theory is a necessary (but not sufficient) basis for scale construction.

The EMS and its parent measure (the PWBI) have good face validity. However, the accumulated research suggests that there is much theoretical and empirical work to be done on the PWBI before it or its scales can be used as reliable and valid measures. Prior research that has relied on any of the versions of the EMS needs to be critically examined, such as the mediational model by Windle and Woods (2004). Given that the average of the items of the EMS tends to correlate highly with other constructs (see Springer and Hauser, 2004), additional research is needed to determine whether the research that utilizes the environmental mastery construct is accurate in their theoretical specification. Future research also needs to avoid using Cronbach's alpha (α) as a proxy for the psychometric properties of this measure, as this study shows that it has little correspondence with the validity of the measure (see also Raykov, 2004).

NOTE

[1] 'Scale' is used interchangeably with 'factor' in this report.

REFERENCES

Bollen, K. and R. Lennox: 1991, 'Conventional wisdom on measurement: A structural equation perspective', Psychological Bulletin 110(2), pp. 305–314.

Callahan, C.M. and F.D. Wolinsky: 1994, 'The effect of gender and race on the measurement properties of the CES-D in older adults', Medical Care 32, pp. 341–356.

Dunbar, M., G. Ford, K. Hunt and G. Der: 2000, 'Question method effects in the assessment of global self-esteem', European Journal of Psychological Assessment 16, pp. 13–19.

Hauser, R.M., W.H. Sewell, and T.S.: Hauser: 1992/93. Wisconsin longitudinal study. [Machine-readable data file]. Madison, WI: University of Wisconsin Madison, WLS. Retrieved on January 14, 2005, from http://www.ssc.wisc.edu/~wls/documentation/.

Himmelfarb, S. and S.A. Murrell: 1983, 'Reliability and validity of five mental health scales in older persons', Journal of Gerontology 38, pp. 333–339.

Horan, P.M., C. DiStefano and R.W. Motl: 2003, 'Method effects in self-esteem scales: Methodological artifact or response style?', Structural Equation Modeling 10, pp. 435–455.

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, pp. 171–190.

Kline, R.B.: 1998, Principles and Practice of Structural Equation Modeling (Guilford, New York, NY).

Lewinsohn, P.M., J.R. Seeley, R.E. Roberts and N.B. Allen: 1997, 'Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults', Psychology & Aging 12, pp. 277–287.

Marsh, H.W.: 1996, 'Positive and negative global self-esteem: A substantively meaningful distinction or artifactors?', Journal of Personality & Social Psychology 70, pp. 810–819.

Measurement Excellence and Training Resource Information Center: n.d. Center for Epidemiological Studies Depression Scale. Retrieved January 14, 2005, from http://www.measurementexperts.org/instrument/instrument_reviews.asp?detail=12.

Motl, R.W. and C. DiStefano: 2002, 'Longitudinal invariance of self-esteem and method effects associated with negatively worded items', Structural Equation Modeling 9, pp. 562–578.

Phillips, R.H.: 1961, 'Environmental mastery and personality development', Archives of General Psychiatry 5, pp. 146–150.

Radloff, L.S. and L. Teri: 1986, 'Use of the center for epidemiological studies-depression scale with older adults', Clinical Gerontologist 5, pp. 119–136.

Raykov, T. and G.A. Marcoulides: 2000, A First Course in Structural Equation Modeling (Lawrence Erlbaum, Mahwah, NJ).
Raykov, T.: 2004, 'Behavioral scale reliability and measurement invariance evaluation using latent variable modeling', Behavior Therapy 35, pp. 299–331.
Ryff, C.D.: 1989, 'Happiness is everything, or is it? Explorations on the meaning of psychological well-being', Journal of Personality & Social Psychology 57, pp. 1069–1081.
Ryff, C.D. and C.L.M. Keyes: 1995, 'The structure of psychological well-being revisited', Journal of Personality & Social Psychology 69, pp. 719–727.
Schumacker, R.E. and R.G. Lomax: 2004, A Beginner's Guide to Structual Equation Modeling, 2nd edition (Lawrence Erlbaum, Mahwah, NJ).
Seltzer, M.M., J.S. Greenberg, F.J. Floyd and J. Hong: 2004, 'Accommodative coping and well-being of midlife parents of children with mental health problems or developmental disabilities', American Journal of Orthopsychiatry 74, pp. 187–195.
September, A.N., M. McCarrey, A. Baranowsky, C. Parent and D. Schindler: 2001, 'The relation between well-being, imposter feelings, and gender role orientation among Canadian university students', Journal of Social Psychology 141, pp. 218–232.
Springer, K.W. and R.M. Hauser: forthcoming, 'An assessment of the construct validity of Ryff's scales of psychological well-being: method, mode and, measurement effects', Social Science Research.
Van Dierendonck, D.: 2004, 'The construct validity of Ryff's Scales of Psychological Well-being and its extension with spiritual well-being', Personality & Individual Differences 36, pp. 629–643.
Wang, J., H.A. Siegal, R.S. Falck and R.G. Carlson: 2001, 'Factorial structure of Rosenberg's Self-Esteem Scale among crack-cocaine drug users', Structural Equation Modeling 8, pp. 275–286.
Windle, G. and R.T. Woods: 2004, 'Variations in subjective wellbeing: The mediating role of a psychological resource', Ageing & Society 24, pp. 583–602.

Brian E. Perron

GWB School of Social Work
Washington University in St. Louis
One Brookings Drive
St. Louis, MO 63130
USA
E-mail: bperron@gwbmail.wustl.edu