JOURNAL OF PERSONALITY ASSESSMENT, *86*(3), 263–272

STATISTICAL DEVELOPMENT AND APPLICATIONS

An Illustration of Issues in Factor Extraction and Identification of Dimensionality in Psychological Assessment Data

Michael F. Steger
Department of Psychology
University of Minnesota, Twin Cities

Structural validity has long been regarded as critical to psychological measurement. However, in practical application, issues central to structural validity are often neglected. The purpose of this study was to illustrate the importance of several crucial choices that face researchers attempting to evaluate the structure of data from a given scale. In this study, I compared the structural solutions derived via principal components analysis and principal axis factoring using eigenvalues, scree plots, and traditional parallel analyses with data from the Purpose in Life Test (Crumbaugh & Maholick, 1964). I discuss the importance of structural validity for overall construct validity and the importance of carefully considering factor analytic methodology. I provide recommendations for uses of factor analysis.

The underlying structure of psychological measures has long been regarded as being central to construct validity, and factorial or structural validity has even been equated with construct validity (Nunnally, 1978). Recent efforts have been made to draw attention to some of the issues involved in establishing the structural validity of psychological measures (e.g., Thompson & Daniel, 1996). Determining the number of factors in a set of data presents researchers with several important decision points. Choice of extraction method and the decision about how many factors to retain are considered among the most critical in scale development (Cudeck, 2000; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Glorfeld, 1995; Tabachnick & Fidell, 2001; Zwick & Velicer, 1986). In this study, I used data from a meaning in life measure to examine the impact of two common methods of factor extraction (principal components analysis [PCA] and principal axis factor analysis [PAF]) on the composition of subscales. First, I compared the number of factors identified using the eigenvalues greater than one heuristic ($K > 1$), scree plot examination, and the traditional method of parallel analysis (PA) as well as the guidelines developed by Guadagnoli and Velicer (1988). I then calculated scores for the subscales suggested by these analyses and used their correlations with measures of optimism, self-esteem, and life satisfaction to demonstrate that creating hybrid scales can lead to potentially misleading and confusing results. Finally, I discuss these results in light of the role of structural validity in the overall construct validation of psychological measures.

Structural validity is important to the measurement of a construct because high structural validity assures assessors that the scores generated by their chosen instruments in a particular sample reflect the theorized structure of that instrument including number, organization, and cohesiveness of any subscales. Guilford (1948) argued that factorial description is a precise and efficient way to create unitary tests of a given construct, which allows the combination of unitary tests to study complex psychological phenomena. Indeed, it is often in test batteries that psychological measures are used. If the factor structure of a measure cannot be replicated from sample to sample, its usefulness is diminished, perhaps to the point where its construct validity is called into question. Cronbach and Meehl (1955), in their watershed monograph on construct validity, cautioned that subgroups of items in tests (extraneous factors), “may set an upper limit to construct validity by showing that irrelevant elements influence scores” (p. 288). Cronbach and Meehl used the example of a response set bias, but their maxim holds true for the inclusion of extraneous psychological constructs as well. In addition to multiple item content areas, an inability to replicate the factor structure of a measure can indicate poor item discrimination, items that are interpreted differently by different samples (due variation by age, culture, gender, etc.), or items that are overly affected by momentary mood sources.

It is critical to psychological research that the factor structure of scores from an instrument is replicable across samples. When the same structure is supported in multiple samples, one can be confident in interpreting scale scores based on that structure. On the other hand, if factor replicability has been a problem, researchers may turn to post hoc means to identify the dimensionality of a scale. One typical post hoc approach researchers have used is to perform a factor analysis of their data, create scales based on those results, and report the observed relations with those scales. This approach can create a confusing marketplace of results in which a disjointed array of relations with differing subscales created from the same items multiply without effectively advancing the field.

The coping literature provides a good example of this phenomenon. Researchers have been interested in how people cope with adversity and in particular, whether certain coping styles are more or less effective in achieving adaptive outcomes. One prominent coping measure that has received extensive attention is the Ways of Coping Scale (Folkman & Lazarus, 1985). This scale is based on Folkman and Lazarus's theoretical model of stress and coping and was designed to assess eight dimensions: problem-focused coping, wishful thinking, distancing, emphasizing the positive, self-blame, tension reduction, self-isolation, and seeking social support. Most of the subscales were designed to assess emotion-focused coping processes, whereas the first subscale taps problem-focused coping, and the eighth is a mixed-focus coping style. However, the original and revised versions have been factor analyzed in numerous studies using several extractions and rotations and have produced a variety of structural solutions (Aldwin & Revenson, 1987). Although some authors have used the original eight scales, it appears more common to use the superordinate problem-focused versus emotion-focused dichotomy (e.g., Sanders-Dewey, Mullins, Chaney, 2001). Additional solutions abound. For instance, Aldwin and Revenson (1987) found and formed subscales from eight factors, but they did not match the eight reported by Folkman and Lazarus. Two-factor (Smyth & Williams, 1991), five-factor (Regan, Lorig, & Thoresen, 1988), and even one-factor (Watson, Willson, & Sinha, 1998) uses of the scale have been reported. Adding to the confusion, or possibly because of this confusion, many researchers have taken to selecting their own subsets of items and either forming still more subscales (e.g., Gueritault-Chalvin, Kalichman, Demi, & Peterson, 2000) or simply reporting the findings for each individual item they used (e.g., Weaver, Turner, & O'Dell, 2000). Such a scenario obviously presents a challenge to researchers interested in using the scale or in providing a review or meta-analysis of the field. This phenomenon is not restricted to coping research. The use of hybrid measures also has emerged in meaning in life research.

In this article, I explore these issues of structural validity in another area in which the factor structure of scales has been a concern: meaning in life research. The most often used meaning in life measure is the Purpose in Life Test (PIL; Crumbaugh & Maholick, 1964). Meaning measures overall, and the PIL in particular, have received periodic criticism for including items that appear to assess distinct constructs with which they should theoretically correlate (Dyck, 1987; Frazier, Oishi, & Steger, 2003; Garfield, 1973; Yalom, 1980). This article was stimulated by the fact that the factor structures of meaning in life scales have created confusion and debate about how best to interpret the proposed scales (e.g., Harris & Standard, 2001). As has been the case in coping research, this has led to post hoc revision and hybridization of scales (e.g., Debats, 1998; Debats, van der Lubbe, & Wezeman, 1993; Harlow, Newcomb, & Bentler, 1986).

Low structural validity may be a pervasive problem for the two most often used measures of meaning in life, the PIL and Life Regard Index (LRI; Battista & Almond, 1973). A contributing factor may be the failure to use common factor analytic procedures in scale construction (e.g., Battista & Almond, 1973; Crumbaugh & Maholick, 1964). Another is the diversity of types of factor extractions researchers have used in their reports. The most common method used in meaning research appears to have been PCA (see, e.g., Chamberlain & Zika, 1988). This procedure is generally more suited for data reduction than for identification of latent factors (Floyd & Widaman, 1995), as PCA introduces spurious common variance into solutions (Comrey, 1988) due to assumptions of perfect measurement (Finch & West, 1997). Less often used are maximum likelihood factor analysis (e.g., Harlow, Newcomb, & Bentler, 1987) and confirmatory factor analysis techniques (see, e.g., Debats et al., 1993, although no fit indexes were reported). There often seems to be little consideration for the fact that different extractions may be more or less appropriate, and there seems to be an overreliance on the eigenvalues > 1.0 criterion as the indicator of the number of factors despite the fact that this approach has been almost unanimously criticized (e.g., Cudeck, 2000; Fabrigar et al., 1999; Tabachnick & Fidell, 2001; Thompson & Daniel, 1996; Zwick & Velicer, 1986). Meaning in life measurement provides an apt example for illustrating consequences of the many decision points in using factor analysis to identify the dimensionality of data.

In this study, I tested the factor structures of PIL scores from a single sample of undergraduate students using both PCA and PAF with orthogonal rotation. When different means of extraction lead to similar solutions, one can be confident that the factors are well identified and robust (Glorfeld, 1995; Tabachnick & Fidell, 2001). Although it is not a preferred method, the $K > 1$ heuristic is the most frequently used method of determining the number of factors (Fabrigar et al., 1999; Henson & Roberts, 2001). Because scree plot examination is the second most often used method

(Fabrigar et al., 1999; Henson & Roberts, 2001), I also included it in this study. In addition, I performed PA (Horn, 1965) to provide a contrast with these two common methods. In its most typical uses, PA generates eigenvalues from correlation matrices of random data sets with the same number of variables and participants as the real data set of interest. One then can be relatively confident that the number of eigenvalues from real data exceeding those from random data represents the number of legitimate latent factors. A more accurate approach is to generate random permutations of the raw data of interest. These approaches can lead to differing results. However, the random permutations of raw data approach is more accurate because the random data approach assumes that the real variables have normal distributions, which is not always the case. Even when this condition is met, the random data approach only approximates actual findings (B. P. O'Connor, personal communication, March 8, 2005). Finally, I also applied Guadagnoli and Velicer's (1988) guidelines to determine the most plausible number of dimensions.

Using statistical syntax developed by O'Connor (2000), one can use PA in its traditional manner, which is to analyze the randomized data according to a principal components model or in an alternative manner, which is to analyze the randomized data according to a principal factors model. Most research to date has been conducted on PA following a principal components model (Velicer, Eaton, & Fava, 2000). The principal components model uses values of 1.0 on the diagonal of the correlation matrix to estimate initial eigenvalues. Using unities on the diagonal directs analyses toward explaining the full amount of variance in each variable. In contrast, the principal factors model uses communality estimates on the diagonal of the correlation matrix to estimate initial eigenvalues.[1] Communalities indicate the proportion of variance in a variable that can be explained by the underlying factors, which in turn are defined by the variables included in the analysis. Using communality estimates on the diagonal directs analyses toward explaining just that portion of the variance in each variable that is common to the other variables in the matrix. Research has not yet fully explored applications of a principal factors model to PA, so it is not clear to what extent support for the use of the traditional, principal-components-based PA will generalize to this alternative method. Therefore, I only used the traditional PA approach for extraction decisions. However, I include consideration of the results suggested by this alternative model in the Discussion section.

Traditional PA was demonstrated to be the most accurate estimate of the number of components in a known data set among five common methods examined (Zwick & Velicer, 1986). In this study, I examined the random permutations of the raw data from the PIL using traditional PA to identify the dimensionality underlying the scores.

[1]O'Connor's (2000) program used squared multiple correlations as estimates of communalities in the principal factors model.

METHOD

Participants and Procedure

A sample of undergraduate psychology students was solicited from introductory courses at a large, Midwestern university. A total of 148 participants completed a packet of questionnaires containing the PIL. Participants were 21.4 years of age (SD = 5.4), mostly female (64%), and mostly White (76%) followed by Asian (10%), African American (3%), Native American (3%), Asian American (2%), and Hispanic (1%), with 6% endorsing "Other." Data were collected as part of a larger investigation reported elsewhere (Steger, Frazier, Oishi, & Kaler, 2006). None of the analyses from this study have been previously conducted.

Measures

PIL. The 20-item PIL (Crumbaugh & Maholick, 1964) is the most widely used meaning-in-life scale despite the concerns described previously regarding confounding with other variables and problems with its factorial structure. Nonetheless, the scale has generally demonstrated good convergent validity with measures of well-being and distress and good internal consistency (e.g., $\alpha = .91$; Zika & Chamberlain, 1992). The PIL provides participants with unique anchors for each item, some of which are bipolar, some of which are unipolar, and some of which span unique continua (i.e., "If I could choose, I would prefer never to have been born … live nine more lives just like this one"). Internal consistency was good in this sample ($\alpha = .89$).

I included three measures of well-being in the study to test whether different factor identification methods produced subscales that differed in their relations with related variables.

Optimism. The Life Orientation Test (LOT; Scheier & Carver, 1985) is a 12-item measure of optimism. Respondents indicated the extent to which they agreed or disagreed with 4 positively worded, 4 negatively worded, and 4 filler items on a 5-point scale ranging from 0 (*disagree strongly*) to 4 (*agree strongly*). The eight substantive items are summed, giving total scale scores that ranged from 8 to 40. The LOT is commonly used, and evidence of its reliability and validity can be found in Scheier and Carver. The alpha coefficient for the LOT in this sample was .83.

Self-esteem. Packets also included the Rosenberg Self-Esteem Test (RSET; Rosenberg, 1965), a 10-item measure of self-regard. The RSET is widely used and accepted and has been shown to be reliable and valid in a large body of studies.

Items are rated on a scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*), giving a scale range of 10 to 40. The alpha coefficient for the RSET in this sample was .84.

Life satisfaction. The Satisfaction With Life Scale (SWLS; Diener, Emmons, Larsen, & Griffin, 1985) is a widely used and well-validated measure of life satisfaction. Five items (e.g., "In most ways my life is close to the ideal") are rated from 1 (*absolutely untrue*) to 7 (*absolutely true*) for scale totals ranging from 5 to 35. The alpha coefficient for the SWLS in this sample was .85.

RESULTS

Data Analysis Strategy

The PIL has items that are reverse scored. I reversed these items prior to all analyses. I first analyzed the data using the $K > 1$ heuristic in conjunction with PCA and PAF. I performed PCA with varimax rotation using SPSS Version 11. I used orthogonal rotation for the sake of simplicity in presentation and varimax specifically because it is the most widely used rotation (Fabrigar et al., 1999). I acknowledge, however, that oblique rotations are often preferred due to the fact that they allow factors and components to be correlated (Fabrigar et al., 1999). Following PCA, I subjected the data to PAF with the same rotation. I chose PAF because it will generally produce similar results to maximum likelihood extraction (which was the case in these studies when I performed duplicate analyses) and is less sensitive to any non-normality in the data (Finch & West, 1997). Next, I examined scree plots, and I used traditional PA using permutations of the raw data as additional indicators of the number of factors. I attempted to use PAF with varimax rotation to extract the number of factors indicated by the $K > 1$ heuristic, scree plot analysis, traditional PA, and Gudagnoli and Velicer's (1988) criteria, and I created scales based on these results. I compared these solutions with the a priori conception of the PIL as unitary (Crumbaugh & Maholick, 1964) as well as with three well-being measures.

Comparison of Principal Components and Principal Axis Factor Analyses Using the Eigenvalues > 1 Criterion

Table 1 presents the eigenvalues and the percent of variance explained by each component with an eigenvalue over 1. The $K > 1$ heuristic indicates that these PIL scores have five factors. This rule has been shown to overestimate the number of factors (Zwick & Velicer, 1986) though. Despite this, it is the most often used approach, perhaps because it is the default option on many statistical programs. Because the purpose of this article is to critically examine these typical approaches, I attempted to produce orthogonally rotated solutions for the number of factors indicated by the eigenvalues > 1 criterion using both PCA and PAF.

Although I extracted five components using PCA with no problems, I reached the maximum number of iterations allowed by SPSS (9,999) without being able to extract the five-factor solution from the PIL using PAF due to commonalities exceeding 1.0. This is most likely indicative of a weak five-factor structure to the PIL scores. When measures consistently produce weak score structures, replicability can be problematic. Thus, it was not possible to determine whether PCA and PAF produce substantially different five-factor solutions in this sample. To get around this problem, I used unweighted least squares (ULS) extraction of factors. ULS derives communalities after estimating a solution and can be seen as a special case of PAF (Tabachnick & Fidell, 2001); therefore, it is suitable for investigating whether different extractions result in substantially different subscales. Based on the varimax-rotated component and factor matrices for the PIL scores, I created five subscales and attempted to describe their content.

TABLE 1
Eigenvalues and Amount of Variance Explained for the PIL As Well As the Eigenvalues Derived by a Traditional PA Run According to a Components Model and an Alternative PA Run According to a Factors Model

Real Data Eigenvalues From Principal Components Analysis	*Random Permutation Eigenvalues From a Traditional PA (Components Model)*	*% Variance*	*Real Data Eigenvalues From PAF With Communalities on the Diagonal*[a]	*Random Permutation Eigenvalues From an Alternative PA According to a Common Factors Model*
<u>6.72</u>	1.85	33.59	<u>6.20</u>	<u>1.02</u>
1.54	1.67	7.69	<u>0.95</u>	<u>0.83</u>
1.37	1.56	6.83	<u>0.80</u>	<u>0.71</u>
1.20	1.45	6.02	<u>0.62</u>	<u>0.60</u>
1.14	1.37	5.69	<u>0.55</u>	<u>0.51</u>
0.99	1.29	4.95	0.28	0.43

Note. $N = 148$. PA were based on random permutations of the raw data. The 95th percentile random eigenvalues from 1,000 datasets generated by random permutations of the raw data are presented under the Random column. Eigenvalues for which the real data exceeds the random permutations of the data are underlined. PIL = Purpose in Life Test; PA = parallel analysisl; PAF = principal axis factor analysis.

[a]Because these eigenvalues were obtained from a matrix with initial communalities on the diagonal, they can only be obtained in SPSS or SAS by using O'Connor's (2000) syntax.

Solution interpretability is another essential element of understanding the structure of scale data. The five factors identified were fairly easy to interpret. The most difficult factor to interpret was the third factor concerning boredom and difficulty finding a mission as well as feeling unprepared for death. Despite difficulties with this factor, the preponderance of factors was interpretable. As can be seen in Table 2, the content of the extractions, in this case, was nearly identical. The only difference between PCA and ULS was that ULS extracted the Zest/Appreciation for Life factor as the second and the World Confusing/Suicidal Ideation factor as the fifth one. I reversed this order for ease of comparison in Table 2. The magnitudes of the loadings within factors also differed in some cases (see Table 3 for the rotated factor matrix from the ULS extraction), and this is reflected in different orders of items on the first and third factors. Thus, in this data, aside from the fact that PAF could not extract the number of factors indicated by the $K > 1$ heuristic, the results are very similar across extractions. Of course, this does not sidestep the issue of whether the item content of the PIL is confounded with other psychological constructs. It does, however, indicate that this five-factor solution is somewhat plausible across extraction methods, at least when using the $K > 1$ heuristic.

Most factorial investigations of meaning measures have stopped here, and using these analyses as the basis, one could conclude that the PIL possesses a multifactor structure that is at odds with its a priori, single-factor structure by these criteria alone. Thus, the scores derived in this sample would not be considered structurally valid, thereby placing an upper limit on the PIL's construct validity (cf. Cronbach & Meehl, 1958). However, other methods do support the single-factor structure. Figure 1 presents the scree plot from these data and from a traditional PA. Interpretation of the scree plot focuses on identifying visually compelling "bends" in a line graph created by connecting eigenvalues. The preferred method is to use a straight edge to extend the slope of the right-hand side of the scree plot and observe where the left-hand side of the plot appears to climb significantly above that slope. This scree plot for the genuine data is clearly suggestive of a single, predominant factor, accounting for 33.6% of the total variance (Table 1). One rationale for examining scree plots is that they may hedge against overextraction because minor factors will not appear very compelling (Velicer et al., 2000). Scree plot examination performed fairly well in identifying dimensionality (Zwick & Velicer, 1986) and has been suggested as one part of a multiple method approach (Velicer et al., 2000), preferably in conjunction with PA (Fabrigar et al., 1999).

PAs

PA proceeds under the assumption that factor analysis of random data from *N* number of participants and *j* number of items will generate more than zero numbers of eigenvalues greater than one. Most common uses of PA produces eigenvalues based on factor analysis of a correlation matrix derived from random data that matches a given, real data set in number of participants and number of items. In this study, I generated random permutations of the actual PIL data set, which is a more accurate approach. The eigenvalues from these random permutations are compared to those derived from the real data. A scree plot is constructed using these two sets of eigenvalues. The point before which the two scree plots cross each other indicates the actual number of factors in the real data set. This controls, in a sense, for the tendency for data sets of the specified size to randomly indicate some structure. The eigenvalues generated from random data approximate a normal distribution (Glorfeld, 1995), and most applications have used the mean eigenvalues. Glorfeld demonstrated that using the eigenvalues at the 95th percentile of the distribution leads to less overextraction than mean eigenvalues, with few problems of underextraction. In this study, I report the eigenvalues at the 95th percentile of those generated from 1,000 randomly permuted data sets.

TABLE 2
Subscales Derived From PCA Versus PAF of the PIL Items and Their Apparent Content

PCA				*ULS*			
Subscale Name	*Items*	*% Var.*	*% Cum. Var.*	*Subscale Name*	*Items*	*% Var.*	*% Cum. Var.*
Enthusiastic Pursuit of Meaningful Life Purpose and Goals	13, 3, 1, 8, 4, 20, 10(–)	16.96	16.96	Enthusiastic Pursuit of Meaningful Life Purpose and Goals	20, 4, 3, 8, 1, 13, 10(–)	14.09	14.09
External Locus of Control	18(–), 14(–)	12.62	29.58	External Locus of Control	18(–), 14(–)	10.69	24.78
Life Boring, Tasks Painful, Afraid of Death	2(–), 5(–), 15(–), 19(–), 17(–)	12.60	42.17	Life Boring, Tasks Painful, Afraid of Dseath	5(–), 2(–), 19(–), 17(–), 15(–)	10.62	35.40
World Confusing, Suicidal Ideation	12, 16	10.14	52.32	World Confusing, Suicidal Ideation	12, 16	6.91	42.31
Zest, Appreciation for Life	6, 9, 11, 7(–)	7.50	59.82	Zest, Appreciation for Life	6, 9, 11, 7(–)	6.33	48.64

Note. $N = 148$. Items that are reverse scored are indicated by an en dash (–). PCA = principal component analysis; PAF = principal axis factor analysis; PIL = Purpose in Life Test; ULS = unweighted least square; Var. = variance explained by a factor; Cum. Var. = cumulative variance explained by each factor and their preceding factors.

TABLE 3
Rotated Factor Matrix of Unweighted Least Squares Extraction With Varimax Rotation of Five Factors From PIL Scores

Factor	*1*	*2*	*3*	*4*	*5*
PIL20	.715	.071	.241	.351	–.068
PIL3	.663	.060	.200	.222	–.001
PIL4	.607	.368	.168	.200	–.104
PIL8	.508	.507	.058	.062	.043
PIL10R	.448	.365	.258	.244	.142
PIL1	.440	.362	.243	–.027	.063
PIL13	.400	.204	.092	–.004	.116
PIL6	.167	.653	.051	.235	.161
PIL9	.298	.596	.312	.214	–.048
PIL11	.329	.475	.316	.376	–.064
PIL7R	.070	.282	.158	.054	.103
PIL5R	.229	.309	.741	–.084	.030
PIL2R	.214	.142	.626	.100	.118
PIL17R	.411	.029	.483	.271	.085
PIL19R	.334	.123	.478	.240	.144
PIL15R	.031	.089	.394	.144	–.042
PIL12	.231	.073	.205	.648	–.033
PIL16	.119	.262	.069	.505	.130
PIL18R	.097	.101	.007	.114	.986
PIL14R	.015	.348	.175	–.056	.366

Note. $N = 148$. Factor number is designated in the first row. Factors 2 and 5 in this table were switched in Table 2 to allow ease of comparison with the principal components analysis. The highest loading factor scores for each factor are underlined. Nine items were reverse scored, as indicated in Table 2. PIL = Purpose in Life Test.

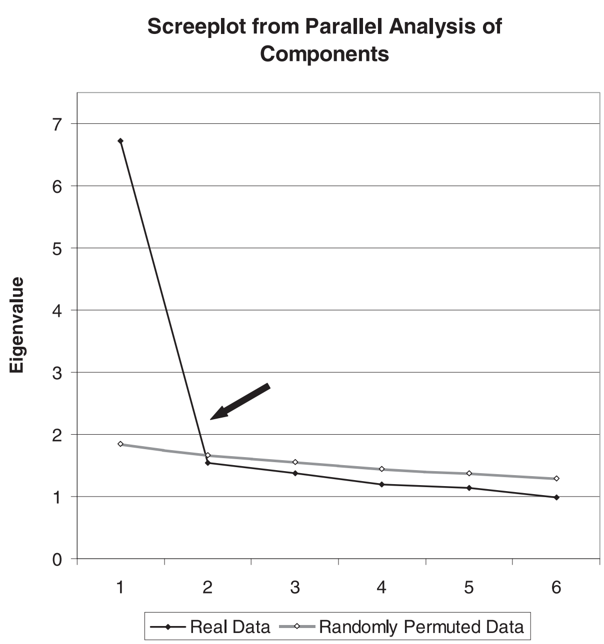

FIGURE 1 Plotting of eigenvalues from factor analysis of the real data for the Purpose in Life Test versus the 95th percentile of those derived from parallel analysis. Eigenvalues were derived from random permutations of the raw data using SPSS Version 11 and syntax developed by O'Connor (2000). The point at which the scree plot for the real data achieved separation from that for the random permutations of the data is signified by an arrow. Simple examination of the scree plot in this case would also focus on the steep falloff after the first factor.

I subjected the data for the PIL to PA of random permutations of the raw data, using the SPSS Version 11 statistical package and syntax developed by O'Connor (2000). The columns on the left in Table 1 provide the results, which clearly indicate one factor for the PIL, matching its a priori structure. The eigenvalues are also shown in a scree plot (see Figure 1). The point at which the two scree plots of eigenvalues cross over is indicated with an arrow. This solution is shown in Table 4, and it appears that the dominant factor of the PIL assesses a general positivity toward life construct that is inclusive of, but not restricted to, meaning in life.

Using guidelines developed by Guadagnoli and Velicer (1988; see also Velicer & Fava, 1998), further support for a single-factor interpretation is indicated. Guadagnoli and Velicer argued that components indicated by four or more variables loading higher than .60 can be interpreted regardless of sample size (Velicer & Fava, 1998, reduced this criterion to three or more large loadings). In this case, only one factor is indicated by four or more loadings greater than .60. When factors are indicated by variables with lower loadings, a sample size of at least 150 is recommended (this sample size was 148) in which case caution should be exercised when interpreting factors indicated by fewer than 10 to 12 variables with loadings of .40 or higher. Again, only the first factor, or a single-factor solution, met these criteria.

However, underextraction can be a serious problem under more circumstances than overextraction (Fava & Velicer, 1992); therefore, it seemed prudent to revisit the five-factor solution suggested by the $K > 1$ heuristic. In further support of this decision is the fact that the amount of total variance accounted for by the single-factor solution was very small (33.6%) in comparison to the average amount of variance explained in a survey of factor analysis studies (52.03% of the total variance; Henson & Roberts, 2001), whereas the first five factors from the solution previously generated from the initial PCA of the PIL items accounted for 59.8% (the first five factors accounted for 48.6% when extracted using ULS; see Table 2). The most obvious difference in core content between the one- and five-factor solution is that the locus of control subscale from Table 2 does not appear in the single-factor solution from Table 4 due to unsatisfactory factor loadings.

Correlations for Factor-Analytically Derived Scales

I derived scale scores for the PIL as originally conceived by Crumbaugh and Maholick (1964) as well as for the 16 items that loaded on the single-factor solution and each of the five factors indicated by the $K > 1$ heuristic. I computed correlations among these scale scores to provide an estimate of the distinctiveness of the factors (see Table 5). The correlations among the original scale, the single-factor version, and the first of the five factors identified in this study are extremely high, which suggests that they measure the same construct. It is clear from these correlations that the first factor (P1) of the

TABLE 4
One-Factor Solution (As Suggested by Parallel Analysis of Components) for the PIL

Item No.	*Factor Loading*	*Item Description*
11	.71	Often wonder why exist
20	.71	Discovered no mission or purpose
10	.70	If died would feel life has been worthwhile
9	.69	Life empty, filled with despair
4	.69	Personal existence utterly meaningless
8	.61	Made no progress in achieving life goals
3	.61	Have no goals
17	.60	Great ability to find mission in life
19	.60	Daily tasks source of pleasure/satisfaction
5	.60	Every day constantly new
1	.57	Usually completely bored
2	.55	Life always exciting
6	.54	Prefer never to have been born
12	.48	World completely confuses me
16	.42	Thought seriously of suicide as way out
13	.41	Very irresponsible
15	.30	Prepared and unafraid regarding death
7	.29	Do exciting things after retiring
14	.29	Absolutely free to make all life choices
18	.25	In control of life

Note. N = 148. Data were analyzed using principal axis factor analysis. Nine items were reverse scored, as indicated in Table 2. PIL = Purpose of Life Test.

TABLE 5
Correlation Matrix of the Original PIL, the Single-Factor Solution PIL, and the Five PIL Factors, Along With Well-Being Measures to Provide Evidence of Convergent Validity

	PIL	*faPIL*	*P1*	*P2*	*P3*	*P4*	*P5*
PIL							
faPIL	.97**						
P1	.86**	.90**					
P2	.46**	.30**	.25*				
P3	.80**	.76**	.57**	.24*			
P4	.62**	.64**	.42**	.16	.37**		
P5	.81**	.79**	.64**	.29**	.51**	.48**	
LOT	.68[a]**	.69**	.60**	.16	.52**	.58**	.57**
RSET	.64[a]**	.66**	.63**	.25*	.35**	.57**	.53**
SWLS	.69[a]**	.68**	.62**	.34**	.48**	.51**	.54**

Note. N = 148. PIL = Purpose in Life Test orginial score; faPIL = single-factor solution PIL score; P1–P5 = scores for five factors identified using unweighted least squares with varimax rotation, as in Table 2; LOT = Life Orientation Test; RSET = Rosenberg Self-Esteem Test; SWLS = Satisfaction Wsith Life Scale.
[a]Correlation taken from Steger, Frazier, Oishi, & Kaler (2006).
*$p < .005$. **$p < .001$.

five-factor solution appears to capture the central meaning in life variance. The correlations among the scales derived in this study are similar in magnitude across the five factors—with the notable exceptions of P2, which had markedly lower correlations with all of the scales, and P4, which had somewhat lower correlations. P2 contains half of the items that were eliminated from the PIL in the single-factor solution, and the pattern of correlations suggests that it captures variance that is not central to what the rest of the PIL measures. P4 also appeared to capture some unique variance, but its items were retained in the single-factor solution, raising questions about whether that solution only assesses meaning in life. Both the original PIL configuration and the single-factor solution correlated extremely highly with the well-being measures, but P2 appeared distinct. Taken together, the correlation matrix suggests the possibility that the PIL contains a Meaning/Life Appreciation factor, an External Locus of Control factor, and a Suicide/Depression factor. However, given the fact that the first factor is so dominant and that two of the other factors are comprised entirely of negatively worded items as well as that the two most different factors (P2 and P4) consist of only two items, nearly all the analyses I performed in this study converge in indicating that these PIL scores reveal a single, dominant factor structure along with a handful of items that do not load on the primary factor and could be removed in revision.

DISCUSSION

Using the most pervasive factor extraction techniques, PCA and PAF, in conjunction with the $K > 1$ heuristic, a mixed picture of the factor structure of the PIL emerged in which the $K > 1$ heuristic indicated a multifactor solution. In contrast, scree plot examination and PA as well as additional criteria (i.e., Guadagnoli & Velicer, 1988) supported the a priori structure. Many writers feel that method of extraction does not have a great impact (e.g., Clark & Watson, 1995), and aside from the fact the five-factor solution was not attainable using PAF due to problems with excessively high communalities, there were practically no differences between PCA and ULS extractions of the five-factor solution indicated by the $K > 1$ heuristic. It is likely that the means used to determine dimensionality does have more impact than extraction method. For instance, I retained one factor for the PIL because that solution matched the PA criteria, the scree plot, Guadagnoli and Velicer's (1988) criteria, and the a priori structure of the scale, and the trailing factors indicated by the $K > 1$ heuristic appeared likely to reflect substantial method variance (e.g., factors composed entirely of negatively worded items). However, the amount of total variance accounted for by the first factor was low, and the five-factor solution indicated by the $K > 1$ heuristic was somewhat interpretable. The measurement literature further suggested that the multifactor solutions should seriously be considered due to problems associated with underextraction (Fava & Velicer, 1992). An analysis of the correlation matrix suggests the possibility that multiple sources of true score variance (e.g., suicidality) could have been present but generally supported the dominance of the primary factor. Thus, whereas the suboptimal approach, $K < 1$, led to an apparent overextraction of five factors, the better methods of identifying dimensionality indicated a single factor underlying these PIL scores. Lingering reasons for concern over the psychometric properties of the PIL include the low variance explained, the interpretability of the five-factor solution, and the chance that negatively worded items introduce method variance. Such factors may provide some insight into why unique solutions for the PIL have emerged and replicability has been a problem across samples and researchers. These findings reinforce the importance of using multiple methods, as different approaches led to different solutions (i.e., $K > 1$ vs. scree plot, PA, and Guadagnoli & Velicer's, 1988, critera).

The importance of structural validity is an often-overlooked element of the basic psychometric properties of measures. Much attention is given to convergent validity, probably due to the ease with which such evidence can be gathered. In addition to relating as hypothesized with other measures, data obtained with a psychological measure should have a replicable internal factor structure that reflects the presence of theoretically important facets of a construct. In my review of the literature, I identified occasions within the coping and meaning in life literatures in which inconsistency in the factorial structure of the scale has led to uncertainty in some cases and a proliferation of idiosyncratic subscales in others. A number of possible reasons for this state of affairs could exist. For example, in the meaning in life literature, factor analytic techniques have not been used in the development of the most-often used scales, detracting from optimal item selection. In this study, I examined two other possible contributing elements: the potential for different methods of factor extraction to lead to differing results and problems in identifying the number of factors to retain related to the use of suboptimal decision rules such as reliance on eigenvalues greater than one heuristic.

It is possible that misunderstanding of, or unfamiliarity with, some of the principles of factor analysis have led to a relative neglect of structural validity as one important component of overall construct validity. For instance, as this review of meaning in life assessment revealed, there has been a disproportionate dependence on PCA to indicate the latent structure of scale scores. Researchers have rarely commented on whether they used an appropriate technique for their purposes. Nor have admonitions regarding the use of the eigenvalues greater than one heuristic been widely heeded, as this was the nearly exclusive criteria for factor identification in the articles I reviewed.

PA: Principal Components Versus Factors

In addition to traditional PA computed according to a components model, the PA syntax I used in this investigation (O'Connor, 2000) gives researchers the option to generate initial eigenvalues from randomized data matrices produced according to a common factors model. Initial eigenvalues are simultaneously generated in their real data set according to the same model. Given that one purpose of this study was to examine the effects of choice of extraction, a question might be raised why results produced by this alternative model were not reported here. One reason for this is that as noted earlier, most previous work on PA has occurred within a components model (Velicer et al., 2000), and it is not clear that those findings can generalize to PA computed according to a common factors model.

In addition, when computing initial eigenvalues, the major statistical programs (i.e., SPSS, SAS) automatically insert values of 1.0 along the diagonals of the correlation matrices used in both components analysis and common factors analysis (e.g., PAF). Thus, initial eigenvalues are determined by a components model (i.e., a matrix with unities on the diagonal) regardless of whether one is interested in ultimately producing a components analysis or a common factors analysis. In contrast, O'Connor's (2000) syntax uses communalities along the diagonal for PA according to a common factors model. The eigenvalues generated by this syntax are therefore not comparable to those generated from the real data by standard SPSS or SAS commands, creating somewhat of a disconnect that may prove confusing for some users. In other words, the initial eigenvalues generated under the components model by SPSS and SAS (which includes both PCA and PAF) cannot be compared to the eigenvalues obtained from O'Connor's PA syntax under a common factors model.

An additional concern emerges concerning the logic of conducting PA using a common factors model. Random, or randomly permuted, data sets are generated, and instead of 1.0, communalities are placed on the diagonal. Communalities reference the amount of variance in each variable that is explained by the other variables. As sample sizes approach infinity, the amount of variance in one randomly created variable explained by all the other randomly created variables (i.e., communalities) approaches zero. Eigenvalues generated based on matrices that have values approaching zero along the diagonal will also shrink. Real eigenvalues will be compared to very small eigenvalues generated from data sets in which the near-zero communalities indicate the variables have no shared variance. Decisions based on the logic of traditional PA in which one extracts all factors in the real data set that have eigenvalues exceeding those in the random data would lead to extracting too many factors because small eigenvalues in the real data set will stand out against random eigenvalues that shrink toward zero as the sample size increases. In addition, the risk that factors with eigenvalues less than one would be extracted increases.

As an illustration of this, I performed PA of principal factors in the same way that I did PA of components. Recall that PA of components indicated a single factor that was in line with both the scree plot analysis and Guadagnoli and Velicer's (1988) criteria. The results of PA of principal factors indicated five factors according to a strict interpretation of the crossover rule (see the last two columns of Table 1). Because these eigenvalues were generated by PA using communality estimates rather than 1.0s on the diagonal of the matrix, they differ from those obtained through the earlier PCA. Of these, four factors had eigenvalues less than 1.0. The magnitude of an eigenvalue does not necessarily matter in PA; rather, one should generally retain all of the factors for which the real eigenvalues exceed the random ones, as PA indicates how many variables are detected at greater than chance level. However, given that the major risk of the $K > 1$ heuristic is overextraction and that PA is intended to control for overextraction, extracting four factors based on eigenvalues less than zero defies the reason for using PA in the first place. In addition, in cases in which additional factors are only slightly above chance levels, as is the case in this PA of factors, it is recommended that information from multiple procedures be used to determine the most likely dimensionality of the data (B. P. O'Connor, personal communication, March 8, 2005). Because of these issues, at present not enough is known about PA of common factors model to recommend its general use.

Limitations

There are several limitations to this study, not the least of which is the fact that the sample sizes were too small to perform confirmatory factor analyses. These techniques provide a rigorous assessment of the factor structure of measures and are widely available in several user-friendly statistical packages. An additional limitation includes the relative homogeneity of the sample both in terms of culture and in age (cf. Clark & Watson, 1995). Finally, one limitation that affects all uses of factor, and parallel, analysis is that psychological data—even those collected using Likert-type, polytomous items (e.g., items rated from 1 to 10)—frequently violate the central assumption the data are continuous rather than ordinal.

Recommendations

Several conclusions can be drawn from the results of this study. Because using scores that do not reflect the actual structure of data in a sample diminishes the certainty a researcher can have regarding those results, it behooves investigators to seek out and employ measures with strong records of structural validity. In addition, researchers should attend to the demonstrated consequences that choice of extraction and method of factor identification can have on the end result, particularly if they intend to create subscales based on these results. These results raise some obvious suggestions for evaluating the dimensionality of psychological measurement data and using factor analysis that has echoed those of several other reviews (e.g., Clark & Watson, 1995; Comrey, 1988; Cudeck, 2000; Fabrigar et al., 1999; Finch & West, 1997; Floyd & Widaman, 1995; Henson & Roberts, 2001; Tabachnick & Fidell, 2001; Thompson & Daniel, 1996; Velicer et al., 2000). First, the choice of extraction method should be made deliberately rather than by default. Most authoritative sources appear to favor principal axis or common factor analysis because of the specific modeling of error variance, a frequent companion to measurement data. Second, multiple methods should be used to identify the correct number of factors to extract. PA is woefully underutilized despite several different approaches that can be used within the framework of familiar statistical packages (e.g., the syntax by O'Connor [2000] for SPSS). Whenever possible, random permutations of the raw data should be used, as this approach will produce the most accurate results. Scree plot analysis and interpretability provide important accompanying information that can assist in determining whether any factors should be eliminated, particularly in cases where the eigenvalues from the actual data are only barely higher than those generated by PA. Third, Gaudagnoli and Velicer's (1988) criteria for sample size and factor saturation levels needed to interpret factor patterns should be followed. Velicer and Fava (1998) provided an updated account for both factor and components analysis and also provided guidelines for planning studies with successful factor analysis in mind. Fourth, replication of scale score structures across samples is vital to establishing structural validity. Failure to replicate signifies poor measurement of a construct. Finally, structural validity should be given a place of prominence among those developing, revising, and utilizing psychological measures.

ACKNOWLEDGMENTS

This research was supported by a Harrison Gough research grant awarded by the University of Minnesota. Thanks to Patricia Frazier and an anonymous reviewer for their input on earlier versions of this article. Grateful acknowledgment to Gregory Meyer and Brian O'Connor for substantially stimulating, informing, and improving the use and discussion of PA in this work, particularly in identifying the inherent challenges to performing PA under a principal factors model.

REFERENCES

Aldwin, C. M., & Revenson, T. A. (1987). Does coping help? A reexamination of the relation between coping and mental health. *Journal of Personality & Social Psychology, 53,* 337–348.

Battista, J., & Almond, R. (1973). The development of meaning in life. *Psychiatry, 36,* 409–427.

Chamberlain, K., & Zika, S. (1988). Measuring meaning in life: An examination of three scales. *Personality and Individual Differences, 9,* 589–596.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7,* 309–319.

Comrey, A. L. (1988). Factor-analytic methods of scale development in personality and clinical psychology. *Journal of Consulting and Clinical Psychology, 56,* 754–761.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Crumbaugh, J. C., & Maholick, L. T. (1964). An experimental study in existentialism: The psychometric approach to Frankl's concept of noogenic neurosis. *Journal of Clinical Psychology, 20,* 200–207.

Cudeck, R. (2000). Exploratory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 265–296). San Diego, CA: Academic.

Debats, D. L. (1998). Measurement of personal meaning: The psychometric properties of the Life Regard Index. In P. T. P. Wong & P. S. Fry (Eds.), *The human quest for meaning: A handbook of psychological research and clinical application* (pp. 237–259). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Debats, D. L., van der Lubbe, P. M., & Wezeman, F. R. A. (1993). On the psychometric properties of the Life Regard Index (LRI): A measure of meaningful life. *Personality and Individual Differences, 14,* 337–345.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49,* 71–75.

Dyck, M. J. (1987). Assessing logotherapeutic constructs: Conceptual and psychometric status of the Purpose in Life and Seeking of Noetic Goals Tests. *Clinical Psychology Review, 7,* 439–447.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction in factor and components analysis. *Multivariate Behavioral Research, 27,* 301–322.

Finch, J. F., & West, S. G. (1997). The investigation of personality structure: Statistical models. *Journal of Research in Personality, 31,* 439–485.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7,* 286–299.

Folkman, S., & Lazarus, R. S. (1985). If it changes it must be a process: A study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology, 48,* 150–170.

Frazier, P., Oishi, S., & Steger, M. (2003). Assessing optimal human functioning. In W. B. Walsh (Ed.), *Counseling psychology and optimal human functioning* (pp. 251–278). Newbury Park, CA: Sage.

Garfield, C. (1973). A psychometric and clinical investigation of Frankl's concept of existential vacuum and anomie. *Psychiatry, 36,* 396–408.

Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55,* 377–393.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103,* 265–275.

Gueritault-Chalvin, V., Kalichman, S. C., Demi, A., & Peterson, J. L. (2000). Work-related stress and occupational burnout in AIDS caregivers: Test of a coping model with nurses providing AIDS care. *AIDS Care, 12,* 149–161.

Guilford, J. P. (1948). Factor analysis in a test-development program. *Psychological Review, 6,* 79–94.

Harlow, L. L., Newcomb, M. D., & Bentler, P. M. (1986). Depression, self-derogation, substance use, and suicide ideation: Lack of purpose in life as a mediational factor. *Journal of Clinical Psychology, 42,* 5–21.

Harlow, L. L., Newcomb, M. D., & Bentler, P. M. (1987). Purpose in Life Test assessment using latent variable methods. *British Journal of Clinical Psychology, 26,* 235–236.

Harris, A. H. S., & Standard, S. (2001). Psychometric properties of the Life Regard Index–Revised: A validation study of a measure of personal meaning. *Psychological Reports, 89,* 759–773.

Henson, R. K., & Roberts, J. K. (2001, February). *A meta-analytic review of exploratory factor analysis reporting practices in published research.* Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED449227).

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

O'Connor, B. P. (2000). *SPSS, SAS, and MATLAB programs for determining the number of components using parallel analysis and Velicer's MAP test.* Retrieved July 17, 2004, from http://flash.lakeheadu.ca/~boconno2/nfactors.html

Regan, C. A., Lorig, K., & Thoresen, C. E. (1988). Arthritis Appraisal and Ways of Coping: Scale development. *Arthritis Care & Research, 1,* 139–150.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Sanders-Dewey, N. E. J., Mullins, L. L., & Chaney, J. M. (2001). Coping style, perceived uncertainty in illness, and distress in individuals with Parkinson's disease and their caregivers. *Rehabilitation Psychology, 46,* 363–381.

Scheier, M. F, & Carver, C. S (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology, 4,* 219–247.

Smyth, K. A., & Williams, P. D. (1991). Patterns of coping in Black working women. *Behavioral Medicine, 17,* 40–46.

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The Meaning in Life Questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53,* 80–93.

Tabachnick, B. G., & Fidell, L. S. (2001). Principal components and factor analysis. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (pp 582 – 652). Boston: Allyn & Bacon.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56,* 197–208.

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helms (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41–71). New York: Kluwer Academic.

Velicer, W. F., & Fava, J. L. (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods, 3,* 231–251.

Watson, D. C., Willson, L. R., & Sinha, B. K. (1998). Assessing the dimensional structure of coping: A cross-cultural comparison. *International Journal of Stress Management, 5,* 77–81.

Weaver, G. D., Turner, N. H., & O'Dell, K. J. (2000). Depressive symptoms, stress, and coping among women recovering from addiction. *Journal of Substance Abuse Treatment, 18,* 161–167.

Yalom, I. D. (1980). *Existential psychotherapy.* New York: Basic.

Zika, S., & Chamberlain, K. (1992). On the relation between meaning in life and psychological well-being. *British Journal of Psychology, 83,* 133–145.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.

Michael F. Steger
Department of Psychology
N218 Elliott Hall
University of Minnesota
Minneapolis, MN 55455
Email: steg0043@umn.edu

Received September 30, 2004
Revised December 20, 2005