Educational and Psychological
Measurement

http://epm.sagepub.com/

New Procedure for Extension Analysis in Exploratory Factor Analysis

Richard L. Gorsuch

Educational and Psychological Measurement 1997 57: 725
DOI: 10.1177/0013164497057005001

The online version of this article can be found at:
http://epm.sagepub.com/content/57/5/725

Published by:
SAGE
http://www.sagepublications.com

Additional services and information for *Educational and Psychological Measurement* can be found at:

Email Alerts: http://epm.sagepub.com/cgi/alerts

Subscriptions: http://epm.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://epm.sagepub.com/content/57/5/725.refs.html

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

NEW PROCEDURE FOR EXTENSION ANALYSIS IN EXPLORATORY FACTOR ANALYSIS

RICHARD L. GORSUCH
Fuller Theological Seminary

Exploratory common factors have been correlated with variables external to the factor analysis by either extension analysis estimates or by correlating the external variables with estimated factor scores. In item analysis, a set of such correlations with possible scales is often computed for final item selection. The purposes of the present article are to document a problem common to both methods of estimating such correlations and to propose a new method of extension analysis that avoids this problem with less restriction assumptions. Solutions for a common data set by all three methods of extension analysis are presented that show the expected difficulties and improvements in the results. Because the new extension procedure gives the correlations without using estimated factor scores, use of this extension procedure eliminates factor indeterminacy from the multiple ways to estimate common factor scores. This method of evaluating possible scales was shown to have advantages over item-remainder correlational analysis.

In exploratory common factor analysis, the term *extension analysis* refers to computing the relationship of common factors to variables that were not included in the factor analysis. Extension analysis assumes that scores for both the factored and the nonfactored variables are available for the same cases and begins with a common factor analysis of the v variables being factored, which are called *core variables*. The term *variables* includes items, combinations of items, and other scores (see Gorsuch, 1983; Thompson & Daniel, 1996, for general issues with factor analysis and Gorsuch, 1997, for issues in item factor analysis).

Extension analysis expands the common factor model by considering additional e variables not being factored for which the relationships with the

---

The author appreciates the contributions of Ira H. Bernstein and Peter H. Schonemann to an earlier version of this article, as well as those of members of the Society for Multivariate Experimental Society and Stanley Mulaik.

Educational and Psychological Measurement, Vol. 57 No. 5, October 1997 725-740
© 1997 Sage Publications, Inc.

from the SAGE Social Science Collections. All Rights Reserved. Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

factors are desired; these are called *extension variables*. The correlations of the v core variables and e extension variables are assumed to be available, but the extension variables are excluded from the factoring, and so the factor structure for the e extension variables is needed. By definition, factors found only in the extension variables are not of interest.

As the primary references (Dwyer, 1937; Horn, 1973; McDonald, 1978; Mosier, 1938) note, extension analysis is desired when the investigator wishes to define the factors without the factors being influenced by the variables in the extension analysis. For example:

1. Factoring items is problematic due to the low reliability of item scores (Bernstein & Teng, 1989; Cattell, 1986; Gorsuch, 1983). For this reason, some recommend factoring "parcels" of items (i.e., mini scales; Cattell, 1950; Comrey, 1973). After factoring the parcels, extension analysis is used to find the correlations of the items that went into the parcels with the factors that result.
2. Items may be excluded from a factor analysis, and yet the correlations of these extension items with the factors are needed. For example, the factor analysis may include proven items but not new, experimental items. This exclusion would be desired to prevent "long-shot" items from influencing the factor solution and yet to be available just in case they happened to work well.
3. Variables might be included in extension analysis because of experimental dependence occurring when the variables were collected, thus producing correlated error.
4. It may be desirable to compute correlations of nonlinear transformations of a factored core variable with a factor.
5. Ipsative items—such as ranking values or coding nominal variables for inclusion in regression analyses—have methodologically imposed covariations and so cannot be included as core variables. Relating ipsative items or variables to factors is possible by extension analysis. For example, it may be desirable to check for ethnic differences on the factors.
6. Higher order factors need to be interpreted by their relationships to the original variables factored for the primary factors. Although the factor pattern and the correlations of the variables with the higher order factor partialled out are easily calculated (Gorsuch, 1983), correlations of the original variables with the higher order factors are also desired, and extension analysis is one method to provide such correlation. In this case, the core variables are the primary factors, the factors are the higher order factors, and the extension variables are the original variables being factored at the primary level.
7. One may wish to test the statistical significance of exploratory factors with certain other variables to confirm or disconfirm hypotheses regarding the nature of the factors. For example, the question of whether age is related to a factor is often of interest. If a variable such as age is in the factor analysis, no statistical significance test can be run because the variable helps define the factors. But if such a variable is employed in an extension analysis rather than as a core variable in the factor analysis, a statistical significance test can be run because that variable does not help define the factors.

Other procedures could be used to provide information relative to these purposes, but they are limited to only some of the named applications or are

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

cumbersome. For example, for the third of the seven applications—experimental dependence—one might factor several subsets of the core and extension items. Each subset would contain nondependent variables. However, that procedure requires conceptual integration of at least two factor analyses and would be inappropriate for the other purposes of extension analysis.

The coefficients in extension analysis should be influenced only by the covariance that a core variable and an extension variable both share with the common factors. Additional covariance between two such variables may occur from factors not in the analysis or from correlated error. Such covariation—referred to hereafter as *nonfactor covariance*—should not affect the correlation of an extension variable with a factor.

The rationale is the same as that applied to item-to-total correlations in item analysis. That correlation is increased because the item variance of that item is in both the item scores and in the total scores. Item-remainder methodology was developed to eliminate that problem but has two problems of its own. First, the valid variance of the item is not in the criterion; although this is a minor problem with long scales, it could be more of a problem with short scales. Second, the item-remainder correlation is affected by the nonvalid variance that the item shares with other items in the remainder. A good extension analysis would reduce both of these problems.

Correlating the extension variables with common factor estimated scores is a procedure that has been used for the purpose of extension analysis. But this is only a minor variant of item-total correlation, as illustrated below, and factor score correlations have similar problems. First, the core variables used to estimate factor scores correlate more highly with those factor scores than with the original factor, because the factor and nonfactor portions of a variable score cannot be separated when scoring, just as in item-total correlations. Second, both the estimate and a linear combination of the factored core variables may include the same nonfactor variance (Gorsuch, 1983, chap. 12) and so correlate more highly than is appropriate. Third, there are multiple factor-scoring procedures and factor scores, so these are a source of indeterminacy (Schonemann & Haagen, 1987; Schonemann & Steiger, 1978; Schonemann & Wang, 1972; Steiger, 1979). An extension analysis procedure is needed that does not directly use factor scores.

Accuracy checks of extension procedures are possible for both general (McDonald, 1978) and special cases of extension variables, and such cases provide practical criteria for further developments of extension analysis robust to nonfactor covariance. Two accuracy checks are used in the present article. The first accuracy check uses an extreme case of nonfactor covariance, a variable that is also in the factor analysis. A variable's correlation with itself is a function of both factored variance and nonfactored variance. If the extension solution is accurate, this variable should relate to the factors in an extension analysis just as it does in the factor analysis itself. This condition is hereafter referred to as *Criterion 1*.

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

A second accuracy check is that the correlations of linear combinations of the core variables with the factors should be the same in the extension analysis as that computed by the equations for linear combinations (e.g., Gorsuch, 1983, chap. 12; Guilford & Fruchter, 1973, p. 516f). The correlations from formulae for linear combinations of core variables do not use factor score estimates and so are not influenced by such vagaries of a factor analysis. And because the factor has no nonfactor variance, the extension correlation is not affected by the problems of item-remainder correlations. Hence, *Criterion 2* is that linear combinations of factored variables placed in extension analysis should relate to the factors the same as they relate to the factors by the equation for linear combinations.

Note that the criteria specified here happen to use as extension variables some core variables in the analysis and some linear combinations thereof, because these variables clearly show the impact of nonfactor covariances and because formulas provide exact answers. Variables that (a) correlate with the core variables because they correlate with the same common factors but that have no nonfactor covariation and that (b) have no covariation with the factors outside of their covariation with the core variables are special cases of the variables used for the two criteria. Hence, the results presented below are applicable to all types of variables, including any combination of items that might be used for a scale.

The purpose of the present article is to show the limits of traditional extension analysis as well as correlations with factor scores. We then consider a new approach to extension analysis that meets the two criteria, with a concluding discussion.

Extension Analysis

Traditional Extension Analysis

Traditional extension analysis (Dwyer, 1937; Horn, 1973; McDonald, 1978; Mosier, 1938) begins with this equation (using Gorsuch's [1983] notation in which the subscripts are the number of rows and number of columns; in this notation, a transposed matrix is recognized by a reversal of the subscripts):

$$R_{ev} = P_{ef} P_{fv} + E_{ev}, \tag{1}$$

where R_{ev} is the matrix of correlations between the e extension variables (rows) and the v core variables being factored (columns), P_{ef} is the factor pattern of the e extension variables with the f factors (and, thus, that which is to be solved for below), and P_{fv} is the transpose of the factor pattern of the v core variables with the f common factors. All factors of interest are found

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

in the common factors of the v core variables; factors common to only the extension variables are not of concern. E_{ev} is the matrix of "error" covariances among the e extension and the v core variables after partialling out the f common factors. Note that this matrix contains all covariation not related to the extracted factors regardless of its source. For example, if the common factors were only of abilities, one of the core variables might correlate with an extension variable because both were influenced by the same motivation. That covariation would be in E_{ev}.

In seeking solutions for P_{ef}, E_{ev} has been assumed by some writers (perhaps unrealistically) to be a null matrix, but the more general solution is a least squares solution that minimizes the sum of squares of E_{ef} in Equation 1. Thus, traditional extension analysis is appropriate if the covariation between the core and extension variables is only a function of the extracted factors so that E_{ef} is trivial after minimization. Variables such as those considered in Criteria I and II violate this assumption, which is why these two criteria form tough standards for extension analysis.

Assuming E_{ev} is a null matrix and solving for P_{ef} gives,

$$P_{ef} = R_{ev} P_{vf} (P_{fv} P_{vf})^{-1}, \tag{2}$$

which is the basis of the traditional method and produces a least squares estimate of P_{ef} minimizing the sum of squares of the elements of E_{ev} (McDonald, 1978).

A less traditional but also informative derivation for Equation 2 has used factor score derivations (Gorsuch, 1974, 1983; Horn, 1973). The conclusion of these writers is that traditional extension analysis is a variant of factor scores, namely the idealized factor score method. This suggests that factor score procedures have the same problems as traditional extension analysis.

As the primary references note, the traditional extension method does produce inflated pattern elements in certain conditions. These arise from the fact that moving from Equation 1 to Equation 2 requires assuming that the observed correlation in R_{ev} is solely from the common factors and not from any other covariance that an extension variable might have in common with a core variable. Hence, extension analysis will give an incorrect estimate whenever an extension variable has nonfactor covariation with a core variable. The simplest examples of such cases are linear combinations of variables in the factor analysis, but any other dependence that produces nonfactor covariance (e.g., correlated error or sharing in a common factor not extracted in this analysis) would also produce incorrect results. Hence, this procedure cannot be used for one of the conditions under which it has been recommended, namely, for evaluating variables that are experimentally dependent on core variables.

To illustrate the problems that can arise from the traditional extension analysis, consider the example in Table 1. Table 1 reproduces the correlations

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

among six psychological variables (Gorsuch, 1983) for a factor analysis and gives the correlations with extension variables.

To represent Criterion 1 in the Table 1 example, two variables—the Information Inventory and Tension—are included both as core variables in the factor analysis and as extension variables. Criterion 2 is represented in the extension analysis by two linear combinations of the Z scores of core variables—Variables 1 + 2 + 3 and Variables 5 + 6 – 4. The correlations of these extension variables with the six core variables and with each other are also in Table 1.

In Table 1 are the correlations of the factored variables with varimax-rotated common principal factors (other common factor procedures—including oblique rotation—are also appropriate and would produce the same conclusions). The correct correlations for both Criterion 1 and Criterion 2 variables are in the last two columns of the table; those for the extension variables also in the factor analysis are the original results, and those for the linear combinations were computed by the formula noted above.

Table 2 contains the correct correlations in columns 1 and 2 and the traditional extension analysis correlations as computed by equation (2) in columns 3 and 4. A comparison of columns 3 and 4 with 1 and 2 of Table 2 shows that variables included in both the factor analysis and in the extension analysis had inflated correlations in the extension analysis (Criterion I). Linear combinations of core variables also were inflated (Criterion II), which is obvious in the current example because some correlations are impossible, namely, 1.14 and 1.21. Hence it can be seen that the traditional extension analysis can produce inflated values violating both criteria. This is worse than item-total correlations because the latter are constrained to be no greater than 1.0.

Columns 5 and 6 of Table 2 contain the correlations of the extension variables with regression factor score estimates (computed by equation 12.1.8, Gorsuch, 1983, p. 282). A comparison of these correlations with the correct values of Table 2 illustrates that the regression factor score estimates also produce erroneous values for Criterion 1 and 2 variables. For example, the extension analysis implies that the last linear combination measures the second common factor perfectly when column 2 shows that it is not so—a fact in keeping with the indeterminacy of common factor analysis and the fact that the reliabilities of most scales, including these, cannot support a correlation of 1.0 unless the correlation is based on error variance.

A New Extension Procedure

An extension analysis that would not produce the inflated values noted for traditional extension analyses and for estimated factor score procedures is needed. Such a procedure should meet both Criteria 1 and 2.

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

Table 1

Correlations Among Variables and With Common Factors

	Core Variables						Extension Variables				Common Factors	
Variables	1	2	3	4	5	6	1	6	7	8	Ability	Anxiety
Core variables												
Information inventory	—										.76	–.10
Verbal ability	.67	—									.81	–.07
Verbal analogies	.43	.49	—								.58	–.07
Ego strength	.11	.12	.03	—							.06	–.67
Guilt proneness	–.07	–.05	–.14	–.41	—						–.05	.59
Tension	–.17	–.14	–.10	–.48	.40	—					–.12	.66
Extension variables												
Information inventory	1.00	.67	.43	.11	–.07	–.17	—				.76	–.10
Tension	–.17	–.14	–.10	–.48	.40	1.00	–.17	—			–.12	.66
Variables 1 + 2 + 3	.84	.87	.77	.10	–.10	–.16	.84	–.17	—		.86	–.10
Variables 5 + 6 – 4	–.15	–.13	–.12	–.80	.77	.80	–.15	.80	–.16	—	.09	.81

Note. Taken from Gorsuch, 1983. The last two rows are calculated by the standard equation for the correlation of a weighted linear composite with other variables.

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

Table 2
Correlations of Extension Variables With Factors by Three Methods

	Correct Values		Traditional Extension Analysis		Multiple Regression Scores	
Variable	Ability	Anxiety	Ability	Anxiety	Ability	Anxiety
Information inventory	.76	–.10	.99	–.03	.87	–.12
Tension	–.12	.66	–.10	.73	–.14	.81
1 + 2 + 3	.86	–.10	1.14	–.06	.98	–.12
5 + 6 – 4	–.09	.81	–.05	1.21	–.11	1.00

Conceptually, the procedure suggested here is to consider the hyperspace defined by the core variables being factored and only those variables. This hyperspace includes both common factor and unique variance, and, because the number of dimensions of the hyperspace is no greater than the number of variables, parameters in it can be computed. Common factors can be projected into this variable space because the correlations between factors and variables are the cosines for such projections. Extension variables also can be projected into this variable space because the correlations of the extension and core variables are the cosines for such projections. Then the cosines of the angles between the common factors and the extension variables can be computed. For ease of computation, the variable space can be defined by an orthogonalization of the variables; the method of orthogonalization is a matter of convenience.

Algebraically, we begin with the supermatrix (which is a matrix containing several matrixes) equation for correlations among all variables and factors as computed from the orthogonalized core variables:

$$R_{v+e+f,\,v+e+f} = P_{v+e+f,\,o}P_{o,\,v+e+f} + E_{v+e+f,\,v+e+f}, \tag{3}$$

in which o identifies the orthogonalized core variables and is equal to v. Note that the orthogonalization is for an intermediate step; the exploratory factors can be correlated. The use of the plus sign (+) in the subscripts means that the matrix is a supermatrix formed of several sections. Thus, $v + e + f$ means that the matrix is composed of the v core variables in the factor analysis plus e extension variables plus the f common factors. E contains the residual covariances.

The matrices and Equation 3 may be best understood by examining Figure 1. Rs are correlations among variables and Ss are factor structures, that is, correlations of variables with the factors extracted from R_{vv}; thus, S_{vf} replaces R_{vf} and S_{ef} replaces R_{ef}. Note that R_{vv} and R_{ve} are available from correlating the raw data, and S_{vf} is estimated by the common factor analysis of R_{vv}. Es are residual matrices. Also note that the complete supermatrix is not orthogonalized; only the core variable section is orthogonalized.

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

$$\begin{bmatrix} R_{vv} & R_{ve} & S_{vf} \\ R_{ev} & R_{ee} & S_{ef} \\ S_{fv} & S_{fe} & R_{ff} \end{bmatrix} = \begin{bmatrix} P_{vo} \\ P_{eo} \\ P_{fo} \end{bmatrix} \begin{bmatrix} P_{ov} & P_{oe} & P_{of} \end{bmatrix} + \begin{bmatrix} E_{vv} & E_{ve} & E_{vf} \\ E_{ev} & E_{ee} & E_{ef} \\ E_{fv} & E_{fe} & E_{ff} \end{bmatrix}$$

Figure 1. Matrix layout of supermatrix for Equation 11.

From the supermatrix of Equation 3, the following is true:

$$S_{ef} = P_{eo}P_{of} + E_{ef}, \tag{4}$$

where E_{ef} is a residual matrix of covariances of the extension variables and each factor after all the core variables have been partialled out. If E_{ef} is assumed to be trivial (see below for a discussion of this assumption), Equation 4 simplifies to

$$S_{ef} = P_{eo}P_{of}. \tag{5}$$

P_{eo} and P_{fo} are computed here by diagonal complete component (neither common factor nor truncated component) analysis (Gorsuch, 1983, chap. 5; also known by a host of other names including the staircase method, Gram-Schmidt triangular decomposition, and Choleski method). As many diagonal components are extracted as there are core variables, thus giving P_{vo}. (Other component orthogonalizing methods may also be appropriate; this one was chosen for its ease of illustration because the procedure is well-known and the computations can be followed by hand calculations.)

Diagonal component analysis defines the first core variable as the first component. Given the defining equation of orthogonal component analysis

$$R_{vv} = P_{vo}P_{ov} \tag{6}$$

and the fact that the first variable is the first component, then the pattern/structure of the first variable and the first component is

$$p_{1A} = 1.0, \tag{7}$$

where p_{1A} is the pattern/structure of variable *1* on the first component, A. Then, by definition of the first component being the first variable, the other patterns/structures of the other core variables are,

$$p_{jA} = r_{j1}, \tag{8}$$

where p_{jA} is the pattern/structure of variable j and factor A, and r_{j1} is the correlation of variable j with variable *1*.

After the first component, the residual matrix is

$$R_{vv.A} = R_{vv} - P_{vA}P_{Av}, \tag{9}$$

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

where $R_{vv.A}$ is the residual matrix after the first component, A, has been removed. The next component is computed from this residual matrix and so is orthogonal to the first component. The diagonal analysis is repeated for each core variable (Gorsuch, 1983, chap. 5).

The fact that the diagonal analysis is a complete component analysis means that E_{vv}, E_{ev}, and E_{fv} are all null matrices. Hence, R_{vv} and R_{ve} are reproduced by these components exactly.

The usefulness for extension analysis of diagonal component analysis is that the procedure just given can be seen to be generalizable to anything correlated with the core variables. Because we already have the correlations of the core variables with both the extension variables and the common factors, P_{eo} and P_{fo} can be computed. This is possible because the equations for the diagonal components can be applied to any variables for which the correlation with the core variables is known. The correlations of variable *1* with (a) the core variables are component A's pattern/structure for the core variables, its correlations with (b) the common factors are component A's pattern/structure for the common factors, and its correlations with (c) the extension variables are component A's pattern/structure for the extension variables. In like manner, the procedures for computing the second diagonal component can give the second component pattern/structure for the core variables, extension variables, and common factors.

Carrying the calculations of pattern/structure and residuals into the extension variable section of the matrix, R_{ev}, and the common factor section, P_{vf}, gives P_{eo} and P_{fo}. The resulting P_{eo} and P_{fo} are substituted into Equation 5 to give the desired correlations, S_{ef}.

The new extension procedure was applied to data in Table 1. Table 3 contains the results of the diagonal component analysis of the core variables (P_{vo}) and the relationships of the components with the extension variables and with the common factors (P_{eo} and P_{fo}).

Equation 5 was used with the data of Table 3 to give the correlations of the extension variables with the factors by this new extension procedure. These are in Table 4. Comparing the values of Table 4 with the correct values in columns 1 and 2 of Table 2 shows the values of Table 4 to be correct.

Comparison of Tables 2 and 4 shows that the new extension procedure meets both criteria. The correlations of the extension variables that were also in the factor analysis are the same in the extension analysis as they were in the original factor analysis (Criterion 1). The correlations of the linear combinations are the same as those given by the equations for computing the correlations of linear combinations (Criterion 2). Hence, the new extension procedure gives the correct correlations, thus meeting our two criteria. Comparing Table 4 with columns 3 through 6 of Table 2 shows the new procedures to be superior to both traditional extension analysis and regression estimated factor score methods. The comparison also shows that, for these variables, assuming E_{ef} of Equation 4 is trivial does not adversely affect the

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

Table 3
Diagonal Factor Loadings

Variables	Diagonal Factors for Variables 1 to 6					
	1′	2′	3′	4′	5′	6′
Factored variables (P_{vo})						
Information inventory	1.00	.00	.00	.00	.00	.00
Verbal ability	.67	.75	.00	.00	.00	.00
Verbal analysis	.43	.28	.84	.00	.00	.00
Ego strength	.11	.07	–.04	.99	.00	.00
Guilt proneness	–.07	.00	–.13	–.41	.90	.00
Tension	–.17	–.04	–.02	–.47	.22	.84
Extension variables (P_{eo})						
Information inventory	1.00	.00	.00	.00	.00	.00
Tension	–.17	–.04	–.02	–.47	.22	.84
1 + 2 + 3	.84	.41	.35	.00	.00	.00
5 + 6 – 4	–.15	–.05	–.05	–.79	.47	.35
Factors (P_{fo})						
Ability	.76	.41	.16	–.05	.00	.00
Anxiety	–.10	–.01	–.03	.67	.33	.30

Table 4
Loadings of Extension Variables (P_{ef}) *With Factors by the New Extension Procedure*

Extension Variables	Ability Factors	Anxiety
Information inventory	.76	–.10
Tension	–.12	.66
1 + 2 + 3	.86	–.10
5 + 6 – 4	–.09	.81

Note. Compare against correct values in Table 2.

results, whereas assuming U_{ev} of Equation 1 is trivial produces inflated pattern/structure elements.

Item Analysis Example

For an item analysis example, Beck's Depression Inventory (BDI; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) is used with data from Lee (1995). To make the tables more readable, only the first six items of the BDI were analyzed and one factor was assumed to exist. The items themselves and several linear combinations of the items that might be used to measure the factor were evaluated. This also allows us to consider the total scale as the total domain of items. The best factor extension procedure would be the one that best estimated the correlations with the item domain.

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

For the extension variables, three possible scales were constructed: one with two items, one with three items, and two with four items. These possible scales were then correlated with the factor by several methods.

Table 5 contains the traditional item statistics in the first two columns, the item-total and item-remainder correlations. Technically, the item-total correlations are correlations with a centroid component from the items with the items weighted as a function of their variances; the item-remainder correlations are also correlations with a weighted centroid component formed by the items in each remainder. A comparison of these two columns gives the standard conclusion: Item-total correlations are too high because the item's nonvalid variance correlates with itself, as is characteristic of component factor methods (Snook & Gorsuch, 1989).

The third column of Table 5 has the factor correlations for the items being factored. The factor is from a principal axis extraction, with communalities estimated as squared multiple correlations followed by two iterations. (A principal component analysis gave inflated correlations like those of column 1.)

The fourth column is of particular theoretical interest. It is the correlation of each item or proposed scale with the total 21-item BDI. This is a quasi-independent criterion, because it is a better measure than the six items used and each item has less impact on the criterion. Hence, the best correlations are those that most resemble the correlations of the last column. The common factor analysis produces item correlations that are closest to the item-domain correlations of the last column.

Five possible scales from the six items were correlated with the total six items (first column) as if to evaluate which were best. The scales consisted of the best two items, the best three items, the best four items from the common factor analysis, the best four items from the item-total analysis, and the six items. The scale-to-total six-item scale correlations are too high, being inflated by the shared nonvalid variance between the items in the proposed scale and the items in the total. Item-remainder correlations were not defined for subscales because the definition of remainder could only be the items not on the scale; to correlate a proposed scale with the items not in the proposed scale would be to correlate them with the less valid items.

The five possible scales were correlated with the factor by the new extension analysis in column 3 of Table 5. The correlations with regression factor scores are given in parentheses. As expected, the new extension procedure gives estimates close to the domain correlations. (And the factor score procedure gives inflated values, as expected.) The inflated values improperly suggest that the six items are perfect; the correlations with the factor using the new extension procedure are more realistic.

When the correlations were corrected for attenuation due to unreliability, the proposed scales' correlations were about 1.0 with the common factor but

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

Table 5

Beck's Depression Inventory (BDI) Item Analysis Correlations Using Tradition Item Analysis and Using Factor Analysis With Gorsuch Extension Analysis

Items	Total	Remainder	Factor	Domain
1.	.63	.45	.55	.55
2.	.51	.36	.44	.49
3.	.69	.46	.58	.56
4.	.71	.49	.61	.64
5.	.59	.41	.49	.45
6.	.61	.35	.43	.44
Correlations to Evaluate Possible Scales				
Best two items	.86	—[a]	.73[b] (.88)	.73
Best three items	.90	—	.78 (.93)	.78
Best cf four items[c]	.94	—	.81 (.97)	.79
Best item-total four items[d]	.96	—	.79 (.95)	.80
All six items	1.00	—	.83 (1.00)	.83

a. No value given because *remainder* is difficult to define for proposed scales.

b. Correlations with the common factor using the new extension analysis (followed by correlations using multiple regression factor scores).

c. Based on item to common factor correlations.

d. Based on item-total correlations.

were all greater than 1.0 for the item-domain results. This suggests that a common factor approach is better than a limited item domain approach. The common factor corrected values suggest that the several proposed factor-based scales are measuring the same thing and differ only in reliability. (An analysis with multiple common factors could give considerably different results for another set of items.)

Note that even with these reasonably good items, the item-total analysis would have lead to a slightly poorer quality four-item scale than the common factor analysis four-item scale. The latter gives a higher correlation with the factor than the former. With more items and with poorer quality items, such mis-selection of items could be more dramatic.

Implications

A comparison of Equation 1 for the traditional extension analysis and Equation 4 for the new extension analysis shows that both contain a matrix that is assumed to be trivial for a useful solution. The matrices both include variation not related to the common factors. Assuming the best case scenario, all elements of both matrices would be zero in the population. But in a sample, the likelihood of a high value by chance is a function of the number of cases being analyzed and the number of elements assumed to be zero. Traditional extension analysis assumes $e * v$ elements are zero, whereas the new proce-

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

dure assumes $e * f$ elements are zero. Because the rank f is generally less than v by at least a factor of 3, there is less chance that E_{ef} will contain an unusually high random value than E_{ev}.

To use the traditional extension method, one assumes that the nonfactor covariations between the v core variables being factored and the e extension variables are zero. For the new extension procedure, it is only necessary to assume that covariations between the unique parts of the e extension variables uncorrelated with the v core variables and the f common factors are zero. The former assumption is difficult to defend in the data analyses we have seen, whereas the latter is quite similar to assuming that the factored variables themselves correlate zero with the unique factors.

What happens if the assumption is incorrect? For the traditional extension analysis, assuming E_{ev} is a null matrix can produce misleading pattern/structure, as shown in Table 2. For the examples computed here, the actual values could be computed, but this is possible only for these special cases. In practice, direct computation is seldom possible and nonfactor covariations may well go unnoticed.

Assuming E_{ef} is a null matrix appears to be less serious when the factors are well-defined in the core variable analysis. "Well-defined" is a function of influences such as sufficient number and variety of variables per factors so that the factor would be little changed with the addition or deletion of a core variable and the N. In particular, the factors must be sufficiently well-defined so that if any extension variable without correlated error with the core variables were included in the core variable factor analysis, the original core variable pattern/structure with the factor would be little changed. The reasonableness of the assumption may, for a particular study, be supported by past research or supplementary factor analyses. If it is felt both that E_{ef} is not zero for an extension variable and that the factor would be poorly defined without it, that variable should be included with the factored variables rather than with the extension variables. In the special cases shown in the tables, assuming E_{ef} is null worked well.

What if a core variable correlates with an extension variable because they both correlate with a common factor not among the core variable factors? Elements of E_{ev} would be nontrivial just as they were for the extension variables in the example, and so the traditional extension pattern/structure would be inaccurate. However, E_{ef} would be less affected, and so the extension pattern/structure would be less affected (again assuming that the core factors are well-defined).

Estimating factor scores is not needed with the new extension procedure. This is desirable because, as Table 2 illustrates for regression factor score estimates, correlating extension variables with factor score estimates produces the same problem as traditional extension analysis. Factor score estimates may give inflated correlations.

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

But there are a few situations where factor scores are still desirable. These include (a) forming scatter plots among factor scores, because it has been suggested (Gorsuch, 1974, 1983) that difficulty factors might be recognized by the nature of the scatter plots; (b) using nonlinear transformations of factor scores (e.g., for polynomial regression); and (c) checking the shape of the distribution of factor scores. Such analyses would still need to use estimated factor scores.

Cattell's (1986) procedure for unpacking items from parcels that have been factored to determine the relationship of the items to the factors is more complex than the current approach. It requires two factor analyses to form the relationships. It appears that the current procedure is a more straightforward approach to achieving the same goal.

With ready access to computers and appropriate programs (e.g., Gorsuch, 1990), there is seldom any need to estimate factor scores to estimate correlations of the factor with other variables. Instead, the procedure given here can be used, thus eliminating common factor variations that arise from the multiple ways of estimating common factor scores. The basic source of indeterminacy—having f plus v dimensions—remains, and so do differences that are a function of the several ways of computing a common factor analysis.

In conclusion, the proposed extension analysis:

1. Enables variables that could theoretically distort the factors—including experimentally dependent and ipsative scales—to be related to exploratory orthogonal or oblique factors.
2. Enables significance tests between nominal as well as continuous extension variables and the factors.
3. Requires assuming only that the unique parts of the extension variables are uncorrelated with the factors, rather than assuming the correlations between the unique parts of the core variables and the extension variables are uncorrelated.
4. Gives the correct results in test cases where the actual correlations can be computed but where both traditional extension analysis and correlating regression factor score estimates give incorrect values with the extension variables.
5. Provides an alternative method, reducing the problems of both item-total and item-remainder correlations for selecting scales and evaluating subsets of items as proposed scales.

References

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561-571.

Bernstein, I. H., & Teng, G. (1989). Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization. *Psychological Bulletin*, *105*(3), 467-477.

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010

Cattell, R. B. (1950). *Personality.* New York: McGraw-Hill.

Cattell, R. B. (1986). Dodging the third error source: Psychological interpretation and use of given scores. In R. B. Cattell & R. C. Johnson (Eds.), *Functional psychological testing: Principles and instruments* (pp. 496-543). New York: Brunner/Mazel.

Comrey, A. L. (1973). *A first course in factor analysis.* New York: Academic Press.

Dwyer, P. F. (1937). The determination of the factor loadings of a given test from the known factor loadings of other tests. *Psychometrika, 2,* 173-179.

Gorsuch, R. L. (1974). *Factor analysis.* Philadelphia: W. B. Saunders.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Gorsuch, R. L. (1990). *UniMult guide.* Pasadena, CA: UniMult.

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, 68*(3), 537-560.

Guilford, P. G., & Fruchter, B. (1973). *Fundamental statistics in psychology and education* (3rd ed.). New York: McGraw-Hill.

Horn, J. L. (1973). On extension analysis and its relation to correlations between variables and factor scores. *Multivariate Behavioral Research, 8,* 477-489.

Lee, C-K. E. (1995). *Evaluation of a Korean translation of the State-Trait Anxiety Inventory and the Beck Depression Inventory.* Unpublished doctoral dissertation, Graduate School of Psychology, Fuller Theological Seminary, Pasadena, CA.

McDonald, R. P. (1978). Some checking procedures for extension analysis. *Multivariate Behavioral Research, 13,* 319-325.

Mosier, C. I. (1938). A note on Dwyer: The determination of the factor loadings of a given test. *Psychometrika, 3*(4), 297-299.

Schonemann, P. H., & Haagen, K. (1987). On the use of factor scores for prediction. *Biometrics Journal, 7,* 835-847.

Schonemann, P. H., & Steiger, J. H. (1978). On the validity of indeterminate factor scores. *Bulletin of the Psychonomics Society, 12,* 287-290.

Schonemann, P. H., & Wang, M. M. (1972). Some new results on factor indeterminacy. *Psychometrika, 37,* 61-91.

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis vs. common factor analysis: A Monte Carlo study. *Psychological Bulletin, 106,* 148-154.

Steiger, J. H. (1979). The relationship between external variables and common factors. *Psychometrika, 44*(1), 93-97.

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.

Downloaded from epm.sagepub.com at OKLAHOMA STATE UNIV on November 17, 2010