

Paper 203-30

# Principal Component Analysis vs. Exploratory Factor Analysis

Diana D. Suhr, Ph.D.

University of Northern Colorado

## Abstract

Principal Component Analysis (PCA) and Exploratory Factor Analysis (EFA) are both variable reduction techniques and sometimes mistaken as the same statistical method. However, there are distinct differences between PCA and EFA. Similarities and differences between PCA and EFA will be examined. Examples of PCA and EFA with PRINCOMP and FACTOR will be illustrated and discussed.

## Introduction

You want to run a regression analysis with the data you've collected. However, the measured (observed) variables are highly correlated. There are several choices

- use some of the measured variables in the regression analysis (explain less variance)
- create composite scores by summing measured variables (explain less variance)
- create principal component scores (explain more variance).

The choice seems simple. Create principal component scores, uncorrelated linear combinations of weighted observed variables, and explain a maximal amount of variance in the data.

What if you think there are underlying latent constructs in the data? Latent constructs

- cannot be directly measured
- influence responses on measured variables
- include unreliability due to measurement error.

Observed (measured) variables could be linear combinations of the underlying factors (estimated underlying latent constructs and unique factors). EFA describes the factor structure of your data.

## Definitions

An **observed variable** can be measured directly, is sometimes called a **measured variable** or an **indicator** or a **manifest variable**.

A **principal component** is a linear combination of weighted observed variables. Principal components are uncorrelated and orthogonal.

A **latent construct** can be measured indirectly by determining its influence to responses on measured variables. A latent construct could also be referred to as a **factor**, **underlying construct**, or **unobserved variable**.

**Unique factors** refer to unreliability due to measurement error and variation in the data.

**Principal component analysis** minimizes the sum of the squared perpendicular distances to the axis of the principal component while least squares regression minimizes the sum of the squared distances perpendicular to the x axis (not perpendicular to the fitted line) (Truxillo, 2003).

**Principal component scores** are actual scores.

**Factor scores** are estimates of underlying latent constructs.

**Eigenvectors** are the weights in a linear transformation when computing principal component scores.

**Eigenvalues** indicate the amount of variance explained by each principal component or each factor.

**Orthogonal** means at a 90 degree angle, perpendicular.

**Oblique** means other than a 90 degree angle.

An observed variable "**loads**" on a factor if it is highly correlated with the factor, has an eigenvector of greater magnitude on that factor.

**Communality** is the variance in observed variables accounted for by a common factor. Communality is more relevant to EFA than PCA (Hatcher, 1994).

## Principal Component Analysis (PCA)

- Is a variable reduction technique
- Is used when variables are highly correlated
- Reduces the number of observed variables to a smaller number of principal components which account for most of the variance of the observed variables
- Is a large sample procedure

The total amount of variance in PCA is equal to the number of observed variables being analyzed. In PCA, observed variables are standardized, e.g., mean=0, standard deviation=1, diagonals of the matrix are equal to 1. The amount of variance explained is equal to the trace of the matrix (sum of the diagonals of the decomposed correlation matrix).

The number of components extracted is equal to the number of observed variables in the analysis. The first principal component identified accounts for most of the variance in the data. The second component identified accounts for the second largest amount of variance in the data and is uncorrelated with the first principal component and so on. Components accounting for maximal variance are retained while other components accounting for a trivial amount of variance are not retained. Eigenvalues indicate the amount of variance explained by each component. Eigenvectors are the weights used to calculate components scores.

### Exploratory Factor Analysis (EFA)

- Is a variable reduction technique which identifies the number of latent constructs and the underlying factor structure of a set of variables
- Hypothesizes an underlying construct, a variable not measured directly
- Estimates factors which influence responses on observed variables
- Allows you to describe and identify the number of latent constructs (factors)
- Includes unique factors, error due to unreliability in measurement
- Traditionally has been used to explore the possible underlying factor structure of a set of measured variables without imposing any preconceived structure on the outcome (Child, 1990).

Figure 1 below shows 4 factors (ovals) each measured by 3 observed variables (rectangles) with unique factors. Since measurement is not perfect, error or unreliability is estimated and specified explicitly in the diagram. Factor loadings (parameter estimates) help interpret factors. Loadings are the correlation between observed variables and factors, are standardized regression weights if variables are standardized (weights used to predict variables from factor), and are path coefficients in path analysis. Standardized linear weights represent the effect size of the factor on variability of observed variables.

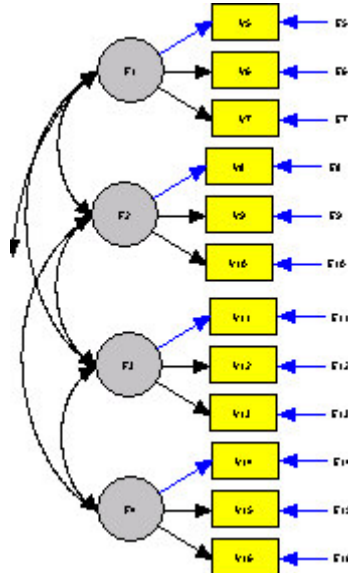


Figure 1. Diagram of four factor model

Variables are standardized in EFA, e.g., mean=0, standard deviation=1, diagonals are adjusted for unique factors, 1-u. The amount of variance explained is the trace (sum of the diagonals) of the decomposed adjusted correlation matrix. Eigenvalues indicate the amount of variance explained by each factor. Eigenvectors are the weights that could be used to calculate factor scores. In common practice, factor scores are calculated with a mean or sum of measured variables that "load" on a factor.

In EFA, observed variables are a linear combination of the underlying factors (estimated factor and a unique factor). Communality is the variance of observed variables accounted for by a common factor. Large communality is strongly influenced by an underlying construct.

Community is computed by summing squares of factor loadings  
 $d_i^2 = 1 - \text{communality} = \% \text{ variance accounted for by the unique factor}$   
 $d_i = \text{square root } (1 - \text{community}) = \text{unique factor weight (parameter estimate)}$

### Similarities

- PCA and EFA have these assumptions in common:
- Measurement scale is interval or ratio level
- Random sample - at least 5 observations per observed variable and at least 100 observations.
- Larger sample sizes recommended for more stable estimates, 10-20 observations per observed variable
- Over sample to compensate for missing values
- Linear relationship between observed variables
- Normal distribution for each observed variable
- Each pair of observed variables has a bivariate normal distribution
- PCA and EFA are both variable reduction techniques. If communalities are large, close to 1.00, results could be similar.

PCA assumes the absence of outliers in the data. EFA assumes a multivariate normal distribution when using Maximum Likelihood extraction method.

### Differences

Principal Component Analysis	Exploratory Factor Analysis
Principal Components retained account for a maximal amount of variance of observed variables	Factors account for common variance in the data
Analysis decomposes correlation matrix	Analysis decomposes adjusted correlation matrix
Ones on the diagonals of the correlation matrix	Diagonals of correlation matrix adjusted with unique factors
Minimizes sum of squared perpendicular distance to the component axis	Estimates factors which influence responses on observed variables
Component scores are a linear combination of the observed variables weighted by eigenvectors	Observed variables are linear combinations of the underlying and unique factors

PCA decomposes a correlation matrix with ones on the diagonals. The amount of variance is equal to the trace of the matrix, the sum of the diagonals, or the number of observed variables in the analysis. PCA minimizes the sum of the squared perpendicular distance to the component axis (Truxillo, 2003). Components are uninterpretable, e.g., no underlying constructs. Principal components retained account for a maximal amount of variance.

The component score is a linear combination of observed variables weighted by eigenvectors. Component scores are a transformation of observed variables ( $C1 = b11x1 + b12x2 + b13x3 + \dots$ ).

The PCA Model is  $Y = XB$

Where  $Y$  is a matrix of observed variables  
 $X$  is a matrix of scores on components  
 $B$  is a matrix of eigenvectors (weights)

SAS code to run PCA is

```
proc factor method=prin priors=one;
```

where priors specify that the prior communality estimate for each variable is set to one, e.g., ones on the diagonals of the correlations matrix. Using `Method=prin` with `priors=one` options runs principal component analysis.

EFA decomposes an adjusted correlation matrix. The diagonals have been adjusted for the unique factors. The amount of variance explained is equal to the trace of the matrix, the sum of the adjusted diagonals or communalities. Factors account for common variance in a data set. Squared multiple correlations (SMC) are used as communality estimates on the diagonals. Observed variables are a linear combination of the underlying and unique factors. Factors are estimated, ( $X1 = b1F1 + b2F2 + \dots + e1$  where  $e1$  is a unique factor).

The EFA Model is  $Y = X\beta + E$

where  $Y$  is a matrix of measured variables  
 $X$  is a matrix of common factors  
 $\beta$  is a matrix of weights (factor loadings)  
 $E$  is a matrix of unique factors, error variation

An example of SAS code to run EFA is

```
proc factor method=ml priors=smc;
```

where priors specify that the prior communality estimate for each variable is set to its squared multiple correlation with all other variables, e.g., diagonals of the correlation matrix are adjusted by unique factors. Using `method=ml` with `priors=smc` options runs a maximum likelihood factor analysis. Using `method=uls` with `priors=smc` runs an unweighted least squares factor analysis. Using `method=prin` with `priors=smc` options runs a principal factor analysis.

### PCA STEPS

- 1) initial PCA – number of components equal to number of variables, only the first few components will be retained
- 2) Determine the number of components to retain
  - a. Eigenvalue > 1 criterion (Kaiser criterion, (Kaiser, 1960))  
Each observed variable contributes one unit of variance to the total variance. If the eigenvalue is greater than 1, then each principal component explains at least as much variance as 1 observed variable.
  - b. Scree test – look for an elbow and leveling, large space between values, sometimes difficult to determine the number of components
  - c. Proportion of variance for each component (5-10%)
  - d. Cumulative proportion of variance explained (70-80%)
  - e. Interpretability – principal components do not exhibit a conceptual meaning
- 3) Rotations is a linear transformation of the solution to make interpretation easier (Hatcher, p.28) With an orthogonal rotation, loadings are equivalent to correlations between observed variables and components.
- 4) Interpret rotated solution
- 5) Summarize results in a table
- 6) Prepare paper
  - a. Describe observed variable
  - b. Method, rotation
  - c. Criterion to determine number of components, eigenvalue greater than 1, proportion of variance explained, cumulative proportion variance explained, number of components retained
  - d. Table, criterion to load on component

### EFA Steps (similar to PCA).

- 1) initial extraction
  - each factor accounts for a maximum amount of variance that has not previously been accounted for by the other factors
  - factors are uncorrelated
  - eigenvalues represent amount of variance accounted for by each factor
- 2) determine number of factors to retain
  - scree test, look for elbow
  - proportion of variance
  - prior communality estimates are not perfectly accurate, cumulative proportion must equal 100% so some eigenvalues will be negative after factors are extracted, e.g., if 2 factors are extracted, cumulative proportion equals 100% with 6 items, then 4 items have negative eigenvalues
  - interpretability
    - at least 3 observed variables per factor with significant factors
    - common conceptual meaning
    - measure different constructs
    - rotated factor pattern has simple structure (no cross loadings)
- 3) rotation – a transformation
- 4) interpret solution
- 5) calculate factor scores
- 6) results in a table
- 7) prepare results, paper

**PROC and options for PCA**

PRINCOMP – a procedure to perform principal component analysis

Data= specifies dataset to be analyzed

Out= output dataset containing original data and principal component scores

Prefix= specifies prefix for naming principal components

N= specifies number of components to retain

FACTOR – a procedure to perform exploratory factor analysis, principal component analysis

Method=principal, performs PCA with priors=one

Rotate=varimax, promax – rotation methods are orthogonal, oblique

Scree requests a scree plot of eigenvalues

Priors=one, sets values of the diagonals of correlation matrix

Flag, \* flags loadings above a specified value

Reorder, sort pattern matrix and rotated pattern matrix from largest to smallest loading for each factor

Mineigen= specifies minimum eigenvalue

**PROC and options for EFA**

PROC FACTOR is a procedure to run exploratory factor analysis and principal component analysis

DATA= specifies dataset to be analyzed

PRIORS=SMC, squared multiple correlations used as adjusted diagonals of the correlation matrix

METHOD=ML,ULS, specifies maximum likelihood and unweighted least squares methods

ROTATE=PROMAX (ORTHOGONAL), VARIMAX(OBLIQUE)

SCREE, requests a scree plot of the eigenvalues

N= specifies number of factors

MINEIGEN=1, specifies select factors with eigenvalues greater than 1

OUT= data and estimated factor scores, use raw data and N=

FLAG=, include a flag (\*) for factor loadings above a specified value

**PCA or EFA?**

Determine the appropriate statistical analysis to answer research questions a priori. It is inappropriate to run PCA and EFA with your data. PCA includes correlated variables with the purpose of reducing the numbers of variables and explaining the same amount of variance with fewer variables (principal components). EFA estimates factors, underlying constructs that cannot be measured directly.

In the examples below, the same data is used to illustrate PCA and EFA. The methods are transferable to your data. Do not run both PCA and EFA, select the appropriate analysis a priori.

**Data and Participants**

Data is from the National Longitudinal Survey of Youth, a longitudinal study of achievement, behavior, and home environment. The original NLSY79 sample design enabled researchers to study longitudinal experiences of different age groups as well as analyze experiences of women, Hispanics, Blacks, and economically disadvantaged. The NLSY79 is a nationally representative sample of 12,686 young men and women who were 14- to 22-years old when first surveyed in 1979 (Baker, Keck, Mott, & Quinlan, 1993).

As part of the NLSY79, mothers and their children have been surveyed biennially since 1986. Although the NLSY79 initially analyzed labor market behavior and experiences, the child assessments were designed to measure academic achievement as well as psychological behaviors. The child sample consisted of all children born to NLSY79 women respondents participating in the study biennially since 1986. The number of children interviewed for the NLSY79 from 1988 to 1994 ranged from 4,971 to 7,089. The total number of cases available for the analysis is 2212.

**Measurement Instrument**

The PIAT (Peabody Individual Achievement Test) was developed following principles of item response theory (IRT). PIAT Reading Recognition, Reading Comprehension, and Mathematics Subtest scores are measured variables in the analysis. Tests were administered in 1988, 1990, 1992, and 1994.

**Examples****PCA with PRINCOMP**

The PRINCOMP procedure runs principal component analysis on the dataset named rawfl. Correlated variables are PIAT math, reading recognition, and reading comprehension scores from 1988, 1990, 1992, and 1994. Eigenvalues are output to a dataset named meigen. The eigenvalues are plotted with the PROC PLOT procedure to show a scree plot.

```
proc princomp data=rawfl;
  var math88 math90 math92 math94 readr88 readr90 readr92 readr94
      readc88 readc90 readc92 readc94;
  ods output eigenvalues=meigen;
proc plot data=meigen;
  plot eigenvalue * Number;
```

Output below from the PRINCOMP procedure shows eigenvalues of the correlation matrix.

A prior criteria has been set to select the number of principal components (PC) that will explain a maximal amount of variance. Criteria are that each PC explain at least 5% of the variance, the cumulative variance is at least 75%, and eigenvalues are greater than one. Remember an eigenvalue greater than one explains at least as much variance as the variance explained by one variable (ones on the diagonals of the correlation matrix).

From the output below, 2 principal components have eigenvalues of 8.28815058 and 1.04770385. A third PC has an eigenvalue of 0.69715466, below the 1.0 eigenvalue criteria. Proportion of variance explained by the first 3 PCs is 69%, 9%, and 6%, meeting the criteria. Cumulative variance explained by 2 PCs is 78% and by 3 PCs is 84%. Two principal components are retained, they meet all criteria.

## The PRINCOMP Procedure

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	8.28814048	7.24043663	0.6907	0.6907
2	1.04770385	0.35054919	0.0873	0.7780
3	0.69715466	0.31370076	0.0581	0.8361
4	0.38345390	0.06882399	0.0320	0.8680
5	0.31462990	0.02470323	0.0262	0.8943
6	0.28992667	0.04134964	0.0242	0.9184
7	0.24857703	0.01229147	0.0207	0.9391
8	0.23628556	0.03297872	0.0197	0.9588
9	0.20330684	0.06144279	0.0169	0.9758
10	0.14186405	0.04657411	0.0118	0.9876
11	0.09528994	0.04162283	0.0079	0.9955
12	0.05366711		0.0045	1.0000

The analysis was run with n=2, to retain 2 principal components. Output below illustrates eigenvalues indicating the amount of variance explained, and eigenvectors, the weights to calculate components scores.

```
proc princomp data=rawfl n=2;
  var math88 math90 math92 math94 readr88 readr90 readr92 readr94
      readc88 readc90 readc92 readc94;
```

## The PRINCOMP Procedure

Eigenvalues of the Correlation Matrix				
	Eigenvalue	Difference	Proportion	Cumulative
1	8.28814048	7.24043663	0.6907	0.6907
2	1.04770385		0.0873	0.7780

Eigenvectors		
	Prin1	Prin2
math88	0.276591	-.397613
math90	0.292381	-.118885
math92	0.277473	0.162226
math94	0.256202	0.303288
readr88	0.295449	-.431803
readr90	0.316860	-.059956
readr92	0.307683	0.179242
readr94	0.289108	0.302284
readc88	0.290135	-.440298
readc90	0.301597	-.046715
readc92	0.292798	0.231146
readc94	0.261855	0.382680

Principal components scores are calculated with eigenvectors as weights. Values are rounded for illustration. SAS calculates the PC scores with eigenvalues shown above.

$$\text{Prin1} = 0.28\text{math88} + 0.29\text{math90} + 0.28\text{math92} + 0.26\text{math94} + 0.30\text{readr88} + 0.32\text{readr90} \\ + 0.31\text{readr92} + 0.29\text{readr94} + 0.29\text{readc88} + 0.30\text{readc90} + 0.29\text{readc92} + 0.26\text{readc94}$$

$$\text{Prin2} = -0.40\text{math88} - 0.12\text{math90} + 0.16\text{math92} + 0.30\text{math94} - 0.43\text{readr88} - 0.05\text{readr90} \\ + 0.18\text{readr92} + 0.30\text{readr94} - 0.44\text{readc88} - 0.5\text{readc92} + 0.23\text{readc92} + 0.38\text{readc94}$$

### PCA with FACTOR and rotation

The FACTOR procedure with compute a principal components analysis including rotation. The code below performs the same analysis as PRINCOMP and includes a varimax rotation (orthogonal). Datasest is rawfl, method is principal, rotation is varimax, diagonals of the correlation matrix are equal to one, 2 PC retained, original data and component scores output to dataset pclsb, component loadings are reordered from largest to smallest values for each factor, variables in the analysis are math, reading recognition, reading comprehension over 4 timepoints.

```
proc factor data=rawfl method=p rotate=v priors=one plot n=2 out=pclsb reorder;
  var math88 math90 math92 math94 readr88 readr90 readr92 readr94
      readc88 readc90 readc92 readc94;
```

Results of the factor procedure illustrate eigenvalues, eigenvectors, and variance explained by each component.. The rotated factor pattern shows weights to calculate component scores. The factor procedure labels items as "factor" even though PCA was run. In this analysis "factor" could be replaced with "principal component".

The FACTOR Procedure

Initial Factor Method: Principal Components

Prior Communality Estimates: ONE

Eigenvalues of the Correlation Matrix: Total = 12 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	8.28814048	7.24043663	0.6907	0.6907
2	1.04770385	0.35054919	0.0873	0.7780
3	0.69715466	0.31370076	0.0581	0.8361
4	0.38345390	0.06882399	0.0320	0.8680
5	0.31462990	0.02470323	0.0262	0.8943
6	0.28992667	0.04134964	0.0242	0.9184
7	0.24857703	0.01229147	0.0207	0.9391
8	0.23628556	0.03297872	0.0197	0.9588
9	0.20330684	0.06144279	0.0169	0.9758
10	0.14186405	0.04657411	0.0118	0.9876
11	0.09528994	0.04162283	0.0079	0.9955
12	0.05366711		0.0045	1.0000

2 factors will be retained by the NFACTOR criterion.

Rotated Factor Pattern		
	Factor1	Factor2
readr94	0.82158	0.33687
readc94	0.81971	0.22319
readc92	0.78003	0.39760
readr92	0.77550	0.46571
math94	0.75265	0.27189
math92	0.69977	0.41954
readr88	0.32556	0.90157
readc88	0.30842	0.89759
math88	0.30938	0.83904
readr90	0.62892	0.66359
math90	0.53623	0.66015
readc90	0.60581	0.62384

Variance Explained by Each Factor			(principal components)
	Factor1	Factor2	
	4.9598291	4.3760152	

Component scores are calculated with the following equations (rounded weights).

$$\text{Prin1} = 0.82\text{readr94} + 0.82\text{readc94} + 0.78\text{readc92} + 0.78\text{math94} + 0.75\text{math94} + 0.70\text{math92} \\ + 0.33\text{readr88} + 0.31\text{readc88} + 0.31\text{math88} + 0.63\text{readr90} + 0.54\text{math90} + 0.61\text{readc90}$$

$$\text{Prin2} = 0.34\text{readr94} + 0.22\text{readc94} + 0.40\text{readc92} + 0.47\text{readr92} + 0.27\text{math94} + 0.42\text{math92} \\ + 0.90\text{readr88} + 0.90\text{readc88} + 0.84\text{math88} + 0.66\text{readr90} + 0.66\text{math90} + 0.62\text{readc90}$$

### EFA with FACTOR

Exploratory factor analysis with dataset rawfl, method is maximum likelihood, scree plot of eigenvalues, diagonals of the correlation matrix are equal to squared multiple correlations, measured variables are math, reading recognition, reading comprehension over 4 timepoints.

```
proc factor data=rawfl method=ml scree priors=smc;
  var math88 math90 math92 math94 readr88 readr90 readr92 readr94
      readc88 readc90 readc92 readc94;
```

Three factors are retained, shown as cumulative variance of 1.0215 (below). Preliminary eigenvalues are 43.7192325, 4.9699785, 1.4603799, larger than the amount of variance explained by one variable. Each factor explains 89%, 10% and 3% of the variance.

```
The FACTOR Procedure
Initial Factor Method: Maximum Likelihood
Preliminary Eigenvalues: Total = 49.0952995 Average = 4.09127496
```

	Eigenvalue	Difference	Proportion	Cumulative
1	43.7192325	38.7492540	0.8905	0.8905
2	4.9699785	3.5095987	0.1012	0.9917
3	1.4603799	0.8226791	0.0297	1.0215
4	0.6377007	0.5299327	0.0130	1.0345
5	0.1077680	0.0819166	0.0022	1.0367
6	0.0258514	0.0874121	0.0005	1.0372
7	-0.0615606	0.1892583	-0.0013	1.0359
8	-0.2508189	0.0309477	-0.0051	1.0308
9	-0.2817666	0.0602684	-0.0057	1.0251
10	-0.3420350	0.0922321	-0.0070	1.0181
11	-0.4342670	0.0208965	-0.0088	1.0093
12	-0.4551635		-0.0093	1.0000

3 factors will be retained by the PROPORTION criterion.



Hypothesis tests are both rejected, no common factors and 3 factors are sufficient. In practice, we want to reject the first hypotheses and accept the second hypothesis. Squared multiple correlations indicate amount of variance explained by each factor. Eigenvalues of the weighted reduced correlation matrix show. eigenvalues of 62.4882752, 7.5300396, and 1.9271076. Proportion of variance explained is 0.8686, 0.1047, and 0.268. Cumulative variance for 3 factors is 100%.

Significance Tests Based on 995 Observations

Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors	66	13065.6754	<.0001
HA: At least one common factor			
H0: 3 Factors are sufficient	33	406.9236	<.0001
HA: More factors are needed			

Chi-Square without Bartlett's Correction	409.74040
Akaike's Information Criterion	343.74040
Schwarz's Bayesian Criterion	181.94989
Tucker and Lewis's Reliability Coefficient	0.94247

Squared Canonical Correlations

Factor1	Factor2	Factor3
0.98424906	0.88276725	0.65836583

Eigenvalues of the Weighted Reduced Correlation Matrix:

	Eigenvalue	Difference	Proportion	Cumulative
Total = 71.9454163				Average = 5.99545136
1	62.4882752	54.9582356	0.8686	0.8686
2	7.5300396	5.6029319	0.1047	0.9732
3	1.9271076	1.2700613	0.0268	1.0000
4	0.6570464	0.4697879	0.0091	1.0091
5	0.1872584	0.0202533	0.0026	1.0117
6	0.1670051	0.0724117	0.0023	1.0141
7	0.0945935	0.1415581	0.0013	1.0154
8	-0.0469646	0.1141100	-0.0007	1.0147
9	-0.1610746	0.0355226	-0.0022	1.0125
10	-0.1965972	0.0469645	-0.0027	1.0097

### EFA with FACTOR and rotation

Options added to the factor procedure are a varimax rotation (orthogonal), reorder to arrange the pattern matrix from largest to smallest loading for each factor, plot to plot factors, n=3 to keep 3 factors, out=facsub3 to save original data and factor scores.

The factor pattern and rotated factor pattern matrices and final communality estimates (diagonals of the adjusted correlation matrix) are shown. Hypothesis tests are both rejected, no common factors and 3 factors are sufficient.

```
proc factor data=rawfl method=ml rotate=v n=3 reorder plot
      out=facsub3 priors=smc;
  var math88 math90 math92 math94 readr88 readr90 readr92 readr94
      readc88 readc90 readc92 readc94;
```

## The FACTOR Procedure

Initial Factor Method: Maximum Likelihood

Preliminary Eigenvalues - same as above

Significance Tests - same as above

Squared Canonical Correlations - same as above

Eigenvalues of the Weighted Reduced Correlation Matrix - same as above

	Factor Pattern		
	Factor1	Factor2	Factor3
readr88	0.94491	-0.26176	0.02539
readc88	0.92409	-0.26406	0.00801
readr90	0.89145	0.20878	0.11313
readc90	0.83206	0.19923	0.06487
readr92	0.83040	0.40006	0.16396
math88	0.81886	-0.13400	-0.21690
math90	0.79779	0.12447	-0.30123
readc92	0.75910	0.39092	0.01863
readr94	0.75071	0.46900	0.15896
math92	0.70865	0.32157	-0.37062
readc94	0.64675	0.41433	-0.05782
math94	0.62910	0.37008	-0.35975

## The FACTOR Procedure

Rotation Method: Varimax

	Orthogonal Transformation Matrix		
	1	2	3
1	0.54103	0.75766	0.36502
2	0.65147	-0.65205	0.38783
3	0.53186	0.02797	-0.84637

	Rotated Factor Pattern		
	Factor1	Factor2	Factor3
readr92	0.79710	0.37288	0.31950
readr94	0.79624	0.26741	0.32138
readr90	0.67849	0.54244	0.31062
readc92	0.67528	0.32076	0.41293
readc90	0.61446	0.50233	0.32608
readc94	0.58909	0.21823	0.44571
readr88	0.35420	0.88731	0.22191
readc88	0.33219	0.87255	0.22813
math88	0.24037	0.70172	0.43051
math92	0.39578	0.31687	0.69707
math94	0.39013	0.22527	0.67765
math90	0.35250	0.51487	0.59444

## Final Community Estimates and Variable Weights

Total Communality: Weighted = 71.945422 Unweighted = 9.364379

Variable	Communality	Weight
math88	0.73553429	3.7804035
math90	0.74270969	3.8855883
math92	0.74295833	3.8909129
math94	0.66215006	2.9610723
readr88	0.96202122	26.3349931
readr90	0.85107493	6.7170126
readr92	0.87649299	8.0949108
readr94	0.80879089	5.2288375
readc88	0.92372805	13.1051008
readc90	0.73621916	3.7926495
readc92	0.72939551	3.6952039
readc94	0.59330358	2.4587311

Factor scores could be calculated by weighting each variable with the values from the rotated factor pattern matrix.. In common practice, factor scores are calculated without weights. The mean or sum of variables is calculated for variables that load, are highly correlated with the factor. Factor scores could be calculated with a mean as illustrated below.

```
Factor1 = mean(readr92, readr94, readr90, readc92, readc90, readc94);
Factor2 = mean(readr88, readc88, math88);
Factor3 = mean(math92, math94, math90);
```

### Conclusion

Principal Component Analysis and Exploratory Factor Analysis are powerful statistical techniques. The techniques have similarities and differences. A priori decisions on the type of analysis to answer research questions allow you to maximize your knowledge.

### WAM (Walk away message)

Principal components are a weighted linear combination of correlated variables, explain a maximal amount of variance of the variables, are uninterpretable. Factors are underlying constructs, not directly measured, influence responses on measured variables, and are influenced by measurement error (unique factors).

### References

Baker, P. C., Keck, C. K., Mott, F. L. & Quinlan, S. V. (1993). NLSY79 child handbook. A guide to the 1986-1990 National Longitudinal Survey of Youth Child Data (rev. ed.) Columbus, OH: Center for Human Resource Research, Ohio State University.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245-276.

Child, D. (1990). The essentials of factor analysis, second edition. London: Cassel Educational Limited.

Hatcher, L. (1994). A step-by-step approach to using the SAS® System for factor analysis and structural equation modeling. Cary, NC: SAS Institute Inc.

Nunnally, J. C. (1978). Psychometric theory, 2<sup>nd</sup> edition. New York: McGraw-Hill.

SAS® Language and Procedures, Version 6, First Edition. Cary, N.C.: SAS Institute, 1989.

SAS® Online Doc 9. Cary, N.C.: SAS Institute.

SAS® Procedures, Version 6, Third Edition. Cary, N.C.: SAS Institute, 1990.

SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 1. Cary, N.C.: SAS Institute, 1990.

SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 2. Cary, N.C.: SAS Institute, 1990.

Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen E. P. (1991). Measurement and evaluation in psychology and education. New York: Macmillan Publishing Company.

Truxillo, Catherine. (2003). *Multivariate Statistical Methods: Practical Research Applications Course Notes*. Cary, N.C.: SAS Institute.

### Contact Information

Diana Suhr is a Statistical Analyst in the Office of Institutional Research at the University of Northern Colorado. In 1999 she earned a Ph.D. in Educational Psychology at UNC. She has been a SAS programmer since 1984.

**email:** [diana.suhr@unco.edu](mailto:diana.suhr@unco.edu)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.