Paper 330-2012

Exploring the Dimensionality of Large-Scale Standardized Educational Assessments using PROC FACTOR

Jonathan Steinberg, Educational Testing Service, Princeton, NJ, USA

ABSTRACT

Standardized educational assessments test students in specific subject areas or measure certain core competencies. Educational researchers regularly use exploratory factor analysis (EFA) to understand a test's internal structure related to its design. PROC PRINCOMP may be used; yet, it has limitations in dealing with the potentially complex structure of standardized test data. This paper will demonstrate how PROC FACTOR is more useful in two ways. First, chi-square hypothesis tests can determine whether a specified number of factors fit the data, particularly when no a priori hypotheses exist about the test's internal structure. Secondly, rotation of multiple factors can be employed to account for inherent inter-factor correlations. This paper is intended for those with good knowledge of multivariate statistics and moderate levels of SAS® programming experience.

INTRODUCTION

There has been a significant increase in large-scale state K-12 standardized testing due to federal legislation (e.g. No Child Left Behind). This has generated increasing interest among educational researchers to explore the relationships among the items on these tests for various sub-groups (e.g. students with disabilities and English language learners) and content areas (e.g. mathematics, science, and English). In analyzing large-scale assessment data as is generally true with any large data set, univariate procedures in SAS® such as PROC FREQ and PROC MEANS, are useful diagnostic tools to run initial checks on the data set to ensure that basic results make sense. PROC CORR would also be useful, but as is often the case with test data where test questions are marked as right (1) or wrong (0), the correlations need to be treated differently to account for guessing or perhaps difficulty factors. A discussion about what types of correlations are appropriate can be found in Steinberg, Cline, and Sawaki (2011).

Large-scale K-12 assessments often contain a large range in the number of items administered according to grade level and/or subject area. Examples of this variability have been reported in Young, Holtzman, and Steinberg (2011). Students often receive sub-scores that are aligned with state standards. There are complex issues regarding the advisability of reporting sub-scores (Haberman, Sinharay, & Puhan, 2006), but this will not be discussed here. Educational researchers often employ factor analysis as a useful data reduction technique to understand the test's internal structure in relation to its design. This is important to investigate because test developers will create items or item sets to measure specific content. Factor analysis can help answer questions such as how many underlying dimensions are present in the data, whether those dimensions are comparable across different populations, and whether there are as many underlying dimensions in the data as reported sub-scores. The common input data for this kind of analysis is either a correlation matrix or a variance-covariance matrix, depending on the nature of the data.

There are two steps to answer these types of research questions: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA methods try to suggest what the hypotheses should be about the internal structure of the test, and CFA methods later test those hypotheses. However, if a priori hypotheses exist about the underlying structure of the data, CFA can be attempted first to confirm those hypotheses, and EFA will only then be performed if those hypotheses cannot be confirmed. Bollen (1989) also discussed the recommended practice of splitting the sample for the two different types of analyses for cross-validation purposes.

There are several methods for extracting factors to perform EFA. This paper focuses on two common extraction techniques available in SAS to perform EFA. PROC PRINCOMP performs principal component analysis (PCA). This technique allows the researcher to understand the dimensionality of a complex multivariate data set at a rudimentary level without accounting for the inter-correlations of the items (Bentler & Kano, 1990). This method allows for easy exploration into how many dimensions may exist in the data based on how much variance is accounted for by each factor and the degree to which test items or sub-scores relate to those dimensions, as expressed by the magnitudes of factor loadings. A factor loading represents the correlation between an observed variable and an unobserved factor (Tabachnick & Fidell, 1989). As many components as variables are extracted, but with PROC PRINCOMP these components are all treated as orthogonal, or uncorrelated, to each other.

Sub-scores on a standardized test usually exhibit at least moderate inter-correlations because on subject-area tests such as mathematics, items pertaining to numerical operations and algebra for example will be grouped in different sub-scales, but will often have some similar characteristics. Findley, Turnbull and Conrad (1947) raised concerns about this phenomenon in test construction. Nonetheless, these inter-correlations often need to be considered in doing factor analysis of administered assessments, so the functionality of PROC PRINCOMP for specific kinds of educational research is limited.

Another commonly-used method of factor extraction is maximum likelihood (ML) estimation. This method is preferred in analyzing survey or assessment data because the additional option of oblique factor rotations is possible, which allows the factors to be correlated with each other. As more dimensions are extracted, it is desired to perform a statistical test to see if each additional dimension in the data adds or detracts from its explanatory value. These options are not available using PROC PRINCOMP, but rather with PROC FACTOR.

The remainder of this paper will utilize a hypothetical example to walk the reader through each of the required steps to conduct an EFA using PROC FACTOR. Maximum likelihood estimation will be used for factor extraction and an oblique form of factor rotation, known as promax, will be employed so that multiple factors can be correlated within the underlying measurement model (Hendrickson & White, 1964). PROC FACTOR is limited though since items cannot be allocated to designated factors. However, the insights gained from the EFA with PROC FACTOR can be insightful in shaping a CFA approach later in the potential analysis.

DESCRIPTION OF A HYPOTHETICAL EXAMPLE

The reader is asked to consider a hypothetical large-scale standardized test that contains 65 multiple-choice items that are marked as right or wrong. Each of these 65 items corresponds to one of five sub-scales based on the test specifications. There are various populations who took the test, ranging from 500 to 30,000 students. The large disparity in sample sizes discourages the analysis of the data using the individual items because of (a) concerns with positive definite correlation matrices (Mislevy, 1986) and (b) non-linear relationships between binary scored items that can create more factors than are really present (Rock, Bennett, & Kaplan, 1985).

As an alternative approach, groups of items within each of the sub-scales can be put together into mini-tests, called item parcels. These parcels generally consist of items with similar content and on average, have a similar level of difficulty. This type of approach was undertaken by Cook, Dorans, and Eignor (1988) using data from the SAT®. A generic description of the pros and cons of item parceling can be found in Little, Cunningham, Shahar, and Widman (2002). Consider in this example that 12 of these smaller units have been created each containing 4 to 7 items, which will minimize potential non-normality concerns with the binary item data. This approach also allows for greater comparability between the test-taking populations by selecting random samples of student test records equal to the size of the smallest population and comparing the results across the samples. More information can be found in Steinberg, Cline, and Sawaki (2011). The factor analysis presented here will utilize a variance-covariance matrix of the sums of scores obtained by students on each of these 12 item parcels for one example test-taking population.

CREATING A DATA STEP

Since a variance-covariance matrix of the parcel scores will be used, in creating the DATA Step, **_TYPE_ = COV** is coded to describe the data set and the matrix of scores used as the input. Next, the number of observations is coded using the **_N_** = syntax. The variable names are then entered following the introductory text **input _name_**. After a **cards** statement, the lower triangle of the variance-covariance matrix of the parcel scores is sufficient. The variable names are included along with the matrix. Figure 1 summarizes these steps.

```
data test (type=cov);
    _type_ = 'COV';
    _N_ = 500;
    input _name_$ par01a par02a par03a par04a par05a par06a
                  par07a par08a par09a par10a par11a par12a;
    cards;
par01a  2.305 . . . . . . . . . . .
par02a  1.742 2.738 . . . . . . . . . .
par03a  1.367 1.485 2.232 . . . . . . . . .
par04a  0.858 1.009 0.944 1.296 . . . . . . . .
par05a  1.419 1.511 1.645 0.940 2.554 . . . . . . .
par06a  1.193 1.303 1.100 0.766 1.233 1.910 . . . . . .
par07a  0.923 0.979 0.941 0.532 0.866 0.931 1.314 . . . . .
par08a  1.039 1.141 0.954 0.616 0.991 1.087 0.786 1.485 . . . .
par09a  1.084 1.140 1.032 0.631 1.036 0.984 0.590 0.796 1.638 . . .
par10a  1.236 1.354 1.178 0.748 1.219 1.139 0.780 0.938 1.141 1.843 . .
par11a  1.140 1.308 1.159 0.755 1.114 1.054 0.788 0.864 1.081 1.179 1.863 .
par12a  0.563 0.531 0.526 0.323 0.592 0.495 0.306 0.431 0.453 0.494 0.519 0.664
;
run;
```

Figure 1. Example DATA Step for Exploratory Factor Analysis Based on Item Parcel Scores

The matrix displayed in Figure 1 could also be generated directly from PROC CORR for a correlation matrix and also for a variance-covariance matrix with the COV option added. PROC CORR would be appropriate in this instance because the underlying data now contain variables with values that are not just 0 or 1, but range between 0 and the number of items in the respective parcels.

THE PROC FACTOR STATEMENT

The PROC FACTOR statement requires several initial arguments including: a valid data set **(DATA=)**, the variables to be analyzed **(VAR)**, the number of observations in the data set **(NOBS=)**, and the number of factors to extract **(N=** or **NFACT=)** when testing whether that number of factors is sufficient to fit the data provided in the variance-covariance matrix. The **N=** or **NFACT=** specifications are only one way that SAS allows the analyst to specify the number of factors. The number of factors to extract can also be designated by a minimum eigenvalue **(MINEIGEN=)** or a threshold on the proportion of explained variance **(PROPORTION=)**. An additional option can be submitted, to create a scree plot, which is a useful picture of the magnitude of the eigenvalues compared to the eigenvalue number. A set of eigenvalues are the roots of a polynomial equation involving the input matrix.

Next, the method of factor extraction is defined along with the type of rotation to be applied if more than one factor is being extracted. There are two kinds of factor rotations: orthogonal and oblique. Orthogonal rotations such as varimax and quartimax, do not allow for multiple factors to be correlated. Oblique rotations such as promax (used in this paper), oblimin, and quartimin, do allow for correlations among multiple factors. Please refer to the SAS documentation for more information on these and other factor rotations. An additional argument referring to the initial value of the communalities, or the portion of the variance of a variable explained by the common factors, may also be required. However, depending on the method of factor extraction, certain specifications need to be made for computing the communalities.

Principal component extraction is only one method of factor extraction. Please see the SAS documentation for more information on other extraction methods, such as principal axis, alpha, and unweighted least squares. If no extraction method is specified, principal component extraction will be employed by default.

Factor extraction methods and prior communality estimates can be paired in different ways because of how the eigenvalues are ultimately scaled. The illustration in this paper will pair maximum likelihood extraction with prior communalities equal to one based on recommendations by Fabriger, Wegener, MacCallum, and Strahan (1999). This avoids the generation of negative eigenvalues which can result when maximum likelihood extraction is paired with prior communalities equal to the squared multiple correlation, which could detract from the interpretability of the factor solution. Figure 2 displays an example PROC FACTOR statement. Figure 3 displays the SAS log from this example statement. The resulting output will be explained later.

```
proc factor data=test scree
    method=ml rotate=promax priors=one nobs=500 n=1;
    var par01a -- par12a;
    title 'ML Method with 1 Factor';
run;
```

Figure 2. Complete Example of a PROC FACTOR Statement

```
NOTE: The means of one or more variables in the input data set WORK.TEST are missing and are
      assumed to be 0.
NOTE: 1 factor will be retained by the NFACTOR criterion.
NOTE: Convergence criterion satisfied.
NOTE: Rotation not possible with 1 factor.
NOTE: PROCEDURE FACTOR used (Total process time):
      real time           0.71 seconds
      cpu time            0.07 seconds
```

Figure 3. Example of SAS log from a PROC FACTOR Statement

The first note in the log represents the fact that the lower triangle of the variance-covariance matrix was entered to be analyzed. The second note indicates that one factor will be retained as specified in the code (n=1). The third note is important because it lets the analyst know that the factor solution properly converged. The fourth note reminds the analyst that factor rotation is not possible with only one factor. Finally, it is evident that the procedure ran very quickly. However, the processing time is dependent upon the size of the variance-covariance matrix or the correlation matrix being used and the internal processing power of the analyst's computer.

ISSUES IN EXTRACTING MULTIPLE FACTORS

In educational assessment, factor analytic studies have tried to test whether a number of dimensions equal to the number of sub-scores can be extracted (Rock, Bennett, Kaplan, & Jirele, 1988). This is not always an attainable goal, yet this does not undermine the utility of the reported sub-scores for providing diagnostic information to students and other stakeholders. In exploratory factor analysis, when the prior communality estimates are set to either one or the squared multiple correlations, the resulting estimates may exceed a value of one. This causes a lack of convergence represented by the statement in the SAS output shown in Figure 4.

```
ERROR: Communality greater than 1.0.
```

Figure 4. PROC FACTOR Output Message for Invalid Communalities

These values are Heywood cases, meaning that the specific variances, or the portion of the total variance due to the specific factor, are negative, resulting in an improper factor solution.

The causes of Heywood cases are often outliers (Bollen, 1987), variability due to sampling (Anderson & Gerbing, 1984), or misspecification of the measurement model (Bollen, 1989). This error can be corrected by adding the HEYWOOD argument to the PROC FACTOR statement to facilitate model convergence or the ULTRAHEYWOOD argument may be added to permit communalities greater than 1.0. However, these modifications can lead to unreliable and misleading results. In such instances, the original hypotheses about the structure of the measurement model may need to be revised or additional sampling may be needed to address the variability issues.

SUMMARY OF OUTPUT FROM THE PROC FACTOR STATEMENT

The output from a PROC FACTOR statement is quite extensive, so only aspects relevant to the discussion will be covered in this section. This section will in many ways be comparable to Suhr (2005), but applied to the context of educational assessment. The hypothetical test under study has five sub-scales, and the task is to determine whether five factors, or perhaps fewer, are sufficient to fit the data. As described in the PROC FACTOR statement in Figure 2, maximum likelihood was the method of factor extraction. Since the sub-scales were inherently correlated, a promax rotation method was applied except when only one factor was extracted. Prior communalities were set to one. Since this test had five sub-scales, several factor solutions were attempted, from one to five. However, Heywood cases were detected when a four-factor solution was requested, so analyses stopped at this point.

EIGENVALUES

The eigenvalues are presented first as shown in Table 1. A scaling factor is applied by SAS when the prior communality estimates are set to one, so the eigenvalues need to be adjusted accordingly to be more in line with the number of variables. The adjustment is to divide the eigenvalues by 999. However, the proportion of explained variance is still valid as reported in the output. If the prior communality estimates are set to the squared multiple correlations, this is equivalent to looking at the weighted reduced correlation matrix obtained when the prior communality estimates are set to one. Generally, an eigenvalue greater than 1.0 is considered to be significant (Kaiser, 1960).

#	Eigenvalue	Difference	Variance Proportion	Cumulative Variance Proportion
1	6.8362		57.0%	57.0%
2	0.7428	6.0933	6.2%	63.2%
3	0.6997	0.0432	5.8%	69.1%
4	0.6332	0.0665	5.3%	74.3%
5	0.5214	0.1118	4.4%	78.7%
6	0.4971	0.0243	4.2%	82.8%
7	0.4618	0.0353	3.9%	86.7%
8	0.3614	0.1004	3.0%	89.7%
9	0.3477	0.0137	2.9%	92.6%
10	0.3299	0.0177	2.8%	95.4%
11	0.2868	0.0431	2.4%	97.8%
12	0.2701	0.0167	2.3%	100.0%

Table 1. Eigenvalue Summary from PROC FACTOR

As is evident in the output and is true of many standardized tests, the first eigenvalue represents a large share of the explained variance compared to subsequent eigenvalues (Steinberg, Cline, & Sawaki, 2011), indicating that only one factor may be adequate to fit the data. This can also be confirmed by looking at the scree plot shown in Figure 5. When the plot bends, known as the elbow, this helps determine the number of factors to retain (Johnson & Wichern, 2002). The elbow occurs at a value of 2 on the X-axis representing the eigenvalue number, indicating that factor should ideally be retained. However, further exploration of the data is necessary in order to fully investigate the number of factors required to account for the data.

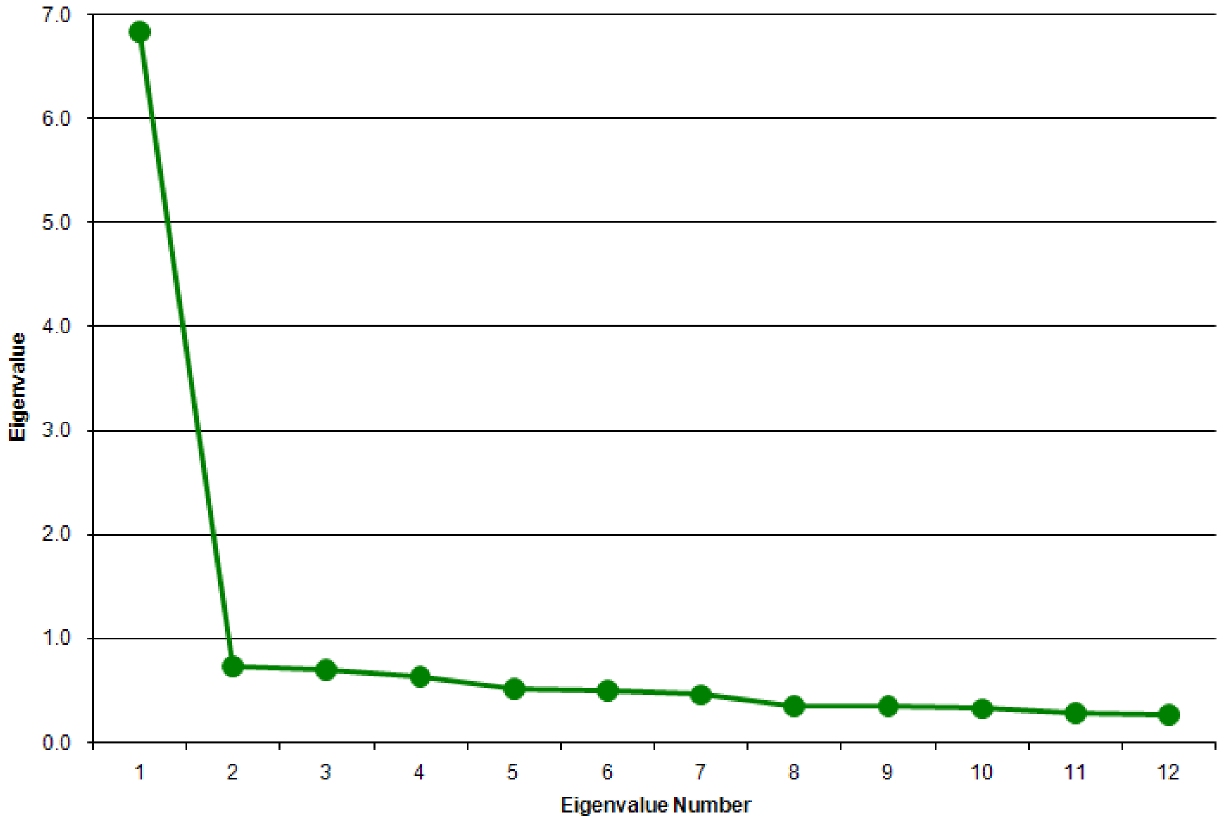

Figure 5. Scree Plot from Working Example

It should be noted that when ODS GRAPHICS are enabled and the **scree** option is changed to **plots=scree**, some of the information provided in Table 1 along with the scree plot from Figure 5 can be simultaneously produced as shown in Figure 6. Please note though that the Y-axis of the graph in the left panel uses the original scaling of the eigenvalues, which is why Figure 5 is also included in this paper.

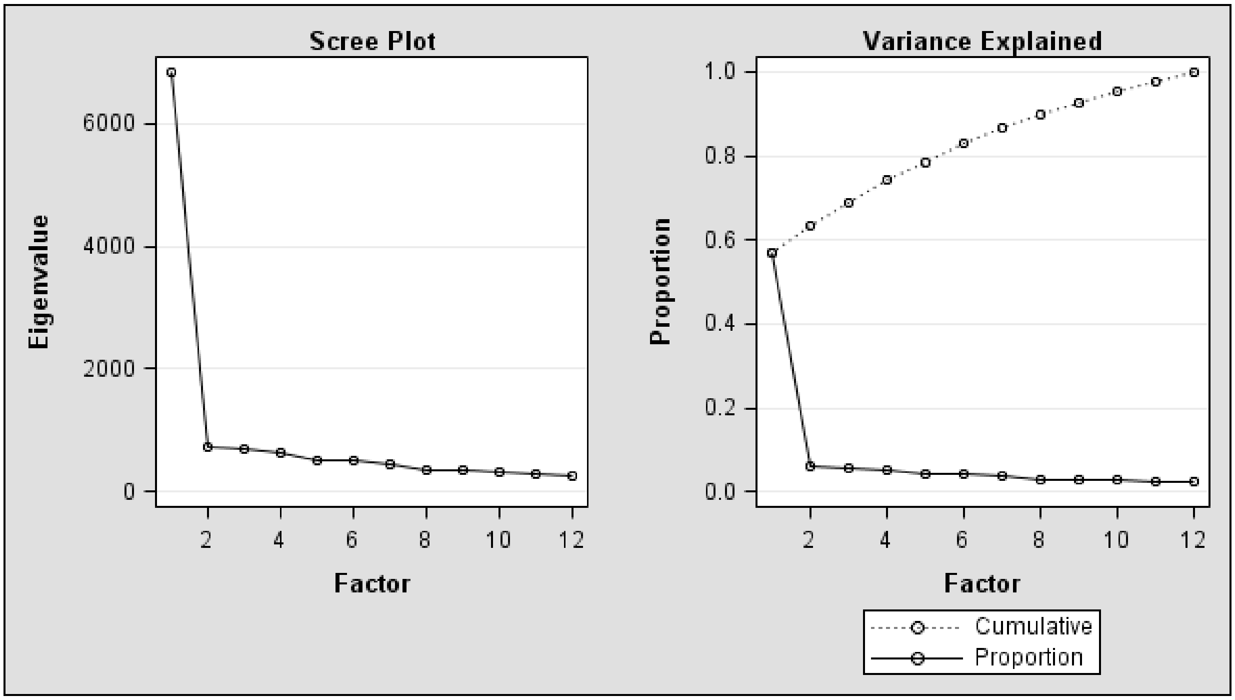

Figure 6. Scree Plot and Variance Explained Plot from Working Example using ODS GRAPHICS

CHI-SQUARE HYPOTHESIS TESTS

SAS implements a statistical test to determine whether the number of extracted factors is sufficient to fit the data. The chi-square test examines the degree of difference between the observed and expected correlation or variance-covariance matrix. The decision rule for evaluating whether the number of factors is sufficient is based on the magnitude of the p-value, as in most statistical tests. According to this test, if the p-value is greater than 0.05, the number of extracted factors is sufficient to fit the input data matrix. If the p-value is less than 0.05, more factors may need to be extracted, which can sometimes be a dubious conclusion if the p-value is significant when the number of specified factors is equal to the number of sub-scores on the test. The Tucker-Lewis Fit Index (TLI) shows the degree to which the model fits the data based on the average correlation between items in the data matrix (Bollen & Long, 1993). The closer the TLI value is to 1.0, the better. Table 2 displays the results from the hypothesis testing across different models.

Factors	Degrees of Freedom	Chi-Square	p-Value	Tucker-Lewis Index
1	54	229.134	<.0001	0.937
2	43	154.197	<.0001	0.950
3	33	81.022	<.0001	0.972

The model fit seems to improve as more dimensions are extracted as the values of the Tucker-Lewis Index increase from 0.937 to 0.972. However, it is apparent that little additional explanatory variance can be found in these dimensions. Additionally, a Heywood case was detected when attempting to extract four factors. While chi-square hypothesis tests are useful diagnostic tools in exploratory factor analysis, these tests are sensitive to sample size. The fact that the significance tests indicate that more than three factors are needed to adequately fit the data when only one eigenvalue is larger than 1.0 confirms the caution that is needed in interpreting results from the chi-square test in isolation and why fit indices should be considered, along with further testing in the CFA stage of analysis.

FACTOR LOADINGS

One of the primary pieces of information derived from a factor analysis is the degree to which each input variable is associated with each extracted dimension, known as a factor loading. As described earlier, this association is generally measured in terms of a correlation. The factor loadings are standardized in the output and range from -1 to +1. A general rule of thumb is that a variable meaningfully contributes to an underlying dimension if its factor loading is at or above 0.32 (Tabachnick & Fidell, 2007). The location of this information in the output is dependent upon the number of factors specified to be extracted in the PROC FACTOR statement for testing. For one factor, the “Factor

Pattern" section shows the loadings of the variables. Accounting for factor rotations with multiple factors, these values are found in the "Standardized Regression Coefficients" section. The factor loadings for this example are displayed in Table 3. Values rounded to or above 0.32 are bolded and all other values are italicized.

	1-Factor Solution	2-Factor Solution		3-Factor Solution		
Parcel	**Factor 1**	**Factor 1**	**Factor 2**	**Factor 1**	**Factor 2**	**Factor 3**
1	**0.782**	**0.464**	**0.372**	**0.361**	*0.238*	*0.271*
2	**0.781**	**0.467**	**0.367**	**0.360**	*0.228*	*0.278*
3	**0.773**	*0.073*	**0.799**	**0.781**	*0.075*	*0.026*
4	**0.647**	*0.233*	**0.472**	**0.463**	*0.125*	*0.135*
5	**0.743**	*0.128*	**0.700**	**0.671**	*0.087*	*0.082*
6	**0.762**	**0.680**	*0.128*	*0.020*	*0.211*	**0.638**
7	**0.672**	**0.438**	*0.281*	*0.211*	*-0.077*	**0.630**
8	**0.729**	**0.636**	*0.135*	*0.029*	*0.135*	**0.666**
9	**0.724**	**0.670**	*0.100*	*0.044*	**0.811**	*0.027*
10	**0.787**	**0.697**	*0.140*	*0.120*	**0.597**	*0.170*
11	**0.749**	**0.635**	*0.162*	*0.156*	**0.549**	*0.136*
12	**0.570**	**0.407**	*0.202*	*0.187*	**0.317**	*0.133*

The factor loadings indicate that the item parcels load strongly on a single factor, confirming the information from the scree plot. Yet, the parcels may load on two or perhaps even three distinct dimensions. This does diminish from the parsimony of a model with multiple dimensions because it is not entirely evident to which factor the parcel is more strongly correlated. This is evident in the two-factor solution as Parcels 6 through 12 appear to separate from Parcels 1 through 5. In the three-factor solution, Parcels 6 through 12 split into two distinct factors. Multiple factors found in an educational assessment may apply to certain specific content areas covered by the test, such as geometry or algebra for a mathematics test. While not shown, the results from the four-factor solution where a Heywood case had been detected would have shown a potential source for producing the Heywood case as evidenced by a factor loading close to 1.0.

In this case, the one-factor solution is quite strong and might be preferred because it is simple to understand. According to Johnson and Wichern (2002), the final decision about the measurement model should be based on the proportion of explained variance, knowledge of the subject of interest, and the degree to which the results look reasonable. The methodology behind the final solution in terms of extraction method and rotation is less important as long as the results from multiple rotation methods essentially confirm the same underlying structure of the data. Therefore, in relation to the previous section on chi-square tests, while models with increasing numbers of dimensions will always show better fit to the data, parsimony needs to be considered along with fit statistics.

INTER-FACTOR CORRELATIONS

The last piece of information in carefully evaluating results from an exploratory factor analysis is the set of inter-factor correlations, which are not the same as the observed correlations between the sub-scales. Inter-factor correlations are computed when at least two factors are extracted and account for factor rotations. If such correlations are extremely high, approaching 0.90 (Stricker, Rock, & Lee, 2005) it is thought that multiple dimensions can be possibly consolidated (Bagozzi & Yi, 1988). However, in exploratory factor analyses, these correlations are not often so high.

In our example, a value of 0.724 was obtained in the two-factor solution and the inter-factor correlations in the three-factor solution were between 0.684 and 0.694, so a solution with multiple dimensions might also be suitable to fit the data. However, the strength of the single-factor solution cannot be dismissed because it represents the most parsimonious model available, making it difficult to obtain meaningful guidance as to how best to proceed with the analysis at this stage, which is why more formal decisions about the measurement model of interest are made during the CFA stage.

SUMMARY AND CONCLUSIONS

This paper has attempted to show the power of PROC FACTOR in analyzing large-scale standardized educational assessment data. The topic is particularly relevant for the SAS Global Forum audience because of the increased use of SAS within the educational field. The analytical capabilities presented here show some advantages for this procedure over PROC PRINCOMP within the context of factor extractions for exploratory factor analysis (EFA).

Whereas PROC PRINCOMP identifies how many dimensions may exist within a data set, PROC FACTOR does a better job at determining how observed scores on test items or item parcels may correspond to unobserved factors.

The results of the exploratory factor analyses obtained using PROC FACTOR in this example demonstrated why often, the simpler model is the preferred model to retain. The chi-square hypothesis tests suggested multiple factors existed even though the scree plot suggested a single factor would best fit the data. This again demonstrates the sensitivity of this test with large samples relative to the number of variables. While the inter-factor correlations were not high enough to suggest factor consolidation and the factor loadings indicated a possible two-dimensional or three-dimensional structure, little additional explanatory variance (5 to 6 percent) was found with these dimensions, causing the uncertainty in interpretation.

Many supplementary options exist in exploratory factor analysis, such as the computation of factor scores which can sometimes be useful in regression models. The reader is directed to the SAS documentation for more information on these features. In most instances, the next step in the analysis would be to perform confirmatory factor analysis (CFA), which can be done using PROC CALIS in SAS. This allows for the assignment of variables to specific factors, which cannot be done in exploratory factor analysis. The results of the confirmatory analysis would attempt to show that while multiple sub-scores are reported to students taking this large-scale assessment, the internal structure of the test might best be represented by a single dimension.

REFERENCES

- Anderson, J. C., & Gerbing, D. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49*, 155–173.
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*, 74-94.
- Bentler, P. M., & Kano, Y. (1990). On the equivalence of factors and components. *Multivariate Behavioral Research, 25,* 67-74.
- Bollen, K. A. (1987). Outliers and improper solutions: A confirmatory factor analysis example. *Sociological Methods and Research, 15*, 375–384.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models.* Newbury Park, CA: Sage.
- Cody, R. P., & Smith, J. K. (1991). *Applied statistics and the SAS® programming language* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Cook, L. L., Dorans, N. J., & Eignor, D. R. (1988). An assessment of the dimensionality of three SAT – Verbal test editions. *Journal of Educational Statistics, 13,* 19–43.
- Fabriger, 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*(3), 272–299.
- Findley, W. G., Turnbull, W. W., & Conrad, H. S. (1947). Construction, evaluation, and applications of intelligence tests. Review of Educational Research*, 17*(1), 10-32.
- Goldberg, R. J. (1997). PROC FACTOR: How to interpret the output of a real-world example. In *Proceedings of the SUGI 22 Conference*. Retrieved from http://www2.sas.com/proceedings/sugi22/STATS/PAPER268.pdf
- Haberman, S., Sinharay, S., & Puhan, G. (2006). *Subscores for institutions* (ETS Research Report No. RR-06-13). Princeton, NJ: Educational Testing Service.
- Hendrickson, A. E., & White, P. O. (1964). PROMAX: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology, 17,* 65–70.
- Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,*141-151.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*(2), 151–173.
- Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics, 11*, 3–31.

- Rock, D. A., Bennett, R. E., & Kaplan, B. (1985). *The internal construct validity of the SAT across handicapped and nonhandicapped populations* (ETS Research Report No. RR-85-50), Princeton, NJ: Educational Testing Service.
- Rock, D. A., Bennett, R. E., Kaplan, B., & Jirele, T. (1988). Factor structure of the Graduate Record Examinations General Test in handicapped and nonhandicapped groups. Journal *of Applied Psychology, 73*, 383-392.
- Santos, J. R. A., Lippke, L., & Pope, P. (1998). PROC FACTOR: A tool for extracting hidden gems from a mountain of variables. In *Proceedings of the SUGI 23 Conference*. Nashville, TN. Retrieved from http://www2.sas.com/proceedings/sugi23/Stats/p240.pdf
- SAS® 9.2 online help and documentation. (2009). Cary, NC: SAS Institute Inc.
- Steinberg, J., Cline, F., & Sawaki, Y. (2011). *Examining the factor structure of a state standards-based science assessment for students with learning disabilities* (ETS Research Report No. RR-11-38). Princeton, NJ: Educational Testing Service.
- Stricker, L. J., Rock, D. A., & Lee, Y.-W. (2005). *Factor structure of the LanguEdge test across language groups* (TOEFL Monograph Series No. MS-32). Princeton, NJ: ETS.
- Suhr, D. D. (2005). Principal component analysis vs. exploratory factor analysis. In *Proceedings of the SUGI 30 Conference*. Retrieved from http://www2.sas.com/proceedings/sugi30/203-30.pdf
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.
- Young, J. W., Holtzman, S., & Steinberg, J. (2011). *Score comparability for language minority students on the content assessments used by two states* (ETS Research Report No. RR-11-27). Princeton, NJ: Educational Testing Service.

ACKNOWLEDGMENTS

The author would like to thank Rachael Biel, Ted Blew, Linda Cook, Kim Fryer, Marna Golub-Smith, Bruce Kaplan, Ed Kulick, Guangming Ling, Anastasiya Osborne, Tyler Smith, and Elizabeth Stone for their support and assistance in proofreading this paper.

Your comments and questions are valued and encouraged. Contact the author at:

Name: Jonathan Steinberg
Enterprise: Educational Testing Service
Address: Rosedale & Carter Roads – Mail Stop 20-T
City, State ZIP: Princeton, NJ 08541
Work Phone: (609) 734-5324
Fax: (609) 734-5010
E-mail: jsteinberg@ets.org
Web: