Informed Decision-Making in Exploratory Factor Analysis

Melissa Cater
Louisiana State University AgCenter
Harrisonburg, LA
mcater@agcenter.lsu.edu

Krisanna Machtmes
Louisiana State University
Baton Rouge, LA
machtme@lsu.edu

Informed Decision-Making in Exploratory Factor Analysis

Melissa Cater and Krisanna Machtmes
Louisiana State University

Abstract: As the field of youth development grows, evaluation of program outcomes has become more embedded in the program process. It is incumbent on youth development professionals to understand more clearly the process of instrument development. One phase of the process involves taking a closer look at the data to determine if an instrument measures one overall construct or if items cluster together to reveal underlying constructs. Exploratory factor analysis is a technique used to determine the underlying constructs, or subgroups, of an instrument. While professionals may use a variety of decision-making steps, researchers have identified some commonly agreed upon considerations when conducting an exploratory factor analysis. This article seeks to make the decision-making steps of this process more transparent while also offering guidance to youth development professionals who may be seeking to use this procedure for the first time.

Introduction

In the field of youth development, evaluating programs based on participant perceptions is a common practice. When developing instruments to measure youth perceptions, it is not uncommon to ask questions about several different aspects of learning that are targeted in a program. Factor analysis is a statistical technique "whose common objective is to represent a set of variables in terms of a smaller number of hypothetical variables" (Kim, & Mueller, 1978, p. 9). Youth development professionals can use this technique to determine the underlying constructs, or subgroups, that an instrument measures. This is important because it allows the professional to purposefully consider a specific construct (i.e., verbal communication skills), and

how this skill is impacted by a treatment or specific demographic variables, instead of simply looking at the impact of the program on overall communication skills.

Choosing the Appropriate Analysis

There are two types of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Exploratory factor analysis is appropriate for use with newly developed instruments where there is no prior knowledge about the existence of underlying patterns in the data, while confirmatory factor analysis is appropriate for testing hypotheses about the structuring of variables into significant factors (Kim, 1970).

Exploratory Factor Analysis. Exploratory factor analysis "seeks to uncover the underlying structure of a relatively large set of variables" (Garson, 2006, paragraph 12). In other words, a survey instrument with a given number of questions is analyzed to see if the questions separate into smaller subgroups. An example of this is a 40-question survey developed to measure youths' perceptions of learning in a leadership program. Exploratory factor analysis would be used to analyze the instrument and determine what, if any, underlying commonalities the questions share. These themes might include a subset of questions which measure perceptions of communication skills, perceptions of problem-solving skills, or numerous other concepts.

Confirmatory Factor Analysis. Confirmatory factor analysis "seeks to determine if the number of factors and the loadings of measured (indicator) variables on them conform to what is expected on the basis of pre-established theory" (Garson, 2006, paragraph 13). This analysis is routinely used with existing instruments, where there is interest in verifying that previously identified latent constructs hold under different circumstances (e.g., a larger population or different ethnic groups). This type of analysis would be appropriate for use with an instrument such as the Free Time Motivation Scale (Baldwin, & Caldwell, 2003), which was tested with a sample of 6th through 8th grade youth whom were 95% Euro-American. If the scale was used in a different population, perhaps one that was largely African-American or Hispanic, then confirmatory factor analysis would be used to ensure that the questions from the survey factored into the same constructs as with the predominantly Euro-American youth.

The choice between exploratory factor analysis and confirmatory factor analysis is driven by the type of instrument that is being analyzed; whether it is a newly developed instrument or a previously developed instrument. Discussing the decision-making process for each of these processes is beyond the scope of this article, therefore the remainder of this article will focus on the specific choices facing users of a newly developed instrument where exploratory factor analysis would be the appropriate procedure to use.

Sample Size Considerations

One of the first considerations impacting the advisability of conducting an exploratory factor analysis is sample size. Is the sample large enough to assure good recovery of population factors? Recommendations are commonly expressed as a proportion of questionnaire respondents to instrument questions or items. Suggestions for this proportion range from a minimum of five respondents for each item, to as many as twenty respondents per item (Gorsuch, 1983; Habing, 2003; Nunnally, 1978). More recently, researchers have suggested that the issues of which variables to include, the communality of variables, and

overdetermination of factors are of more importance than sample size (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Hogarty, Hines, Kromrey, Ferron, & Mumford, 2005).

MacCallum et al. (1999) suggest that when communalities are high (greater than .6), sample size has less impact on recovery of population factors, although this "rule" does not occur in a vacuum. Determination of factors, and the factor solution, also plays a part in recovering population factors. Mundfrom, Shaw, Ke, Knofczinski, and McFann (2003) extend the discussion of factor determination by recommending a greater number of variables per factor in order to further elucidate the factor. This is supported by Hogarty, et al. (2005), who noted that not only did overdetermination reduce the influence of sample size, but it also improved the factor analysis solution. Henson and Roberts (2006) sum it up best with their contention that "the best rule of thumb to follow is to get the largest possible sample for a factor analysis" (p. 402).

Extracting the Factors

At this step of the process, the main objective of EFA is to determine the minimum number of factors needed to produce the correlations among the observed variables (Kim, & Mueller, 1978). Researchers often disagree on the type of extraction to use in EFA. Extraction models fall into two categories: components models [e.g., principal components analysis (PCA)] and common factor models (e.g., principal axis factoring and maximum likelihood factoring).

Some researchers suggest that components models and common factor models may be used interchangeably in EFA; however, when communalities are .40 or lower and there are three or fewer measured variables per factor, the results of PCA and those of common factor models may be different (Fabrigar, et al., 1999). Additionally, the existence of random error in social science research is assumed. This assumption indicates the need for an extraction model that allows the user to take this error into account.

To better understand this, it is perhaps easier to think of a variable in terms of the following equations: variable = common variance + specific variance + error and unique variance = specific variance + error. The common variance portion of a variable indicates the amount of influence that a variable has among all of the factors. The unique variance part of a variable shows the amount of specific variance and error that is distinctive to that particular variable.

Use of a common factors method of extraction allows the researcher to estimate the pattern of relations between the common factor(s), with the unique variance partitioned from the common variance (Fabrigar, et al., 1999). This method shows the amount of variance shared among factors as well as indicating the amount of variance that is unique to a factor. Components models (like PCA) combine the unique and common variance, resulting in a matrix which is not representative of latent variables, and simply seek a reduction in the number of variables in the solution (Conway, & Huffcutt, 2003; Fabrigar, et al., 1999). This method does not distinguish between common and unique variance of factors but instead combines it all together. Therefore, when analyzing a newly developed instrument, the best decision is one where a common factor model is utilized.

Determining Number of Factors

Once the factors are extracted, the number of factors to retain must be determined. Popular strategies include employing Kaiser's (1960) criteria that stipulates that eigenvalues associated with the extracted factors should equal or exceed unity (EV>1), utilizing Cattell's (1966) scree plot, and using the solution that accounts for the highest proportion of variance while still producing an interpretable solution (Conway, & Huffcutt, 2003; Fabrigar, et al., 1999). Henson and Roberts (2006) report that researchers commonly only use one strategy to determine the number of factors retained, with the EV>1 rule being the most commonly used approach. Used independently, each method has its own share of pros and cons. The optimal approach is to employ multiple strategies to determine the number of factors to retain, with special attention given to reporting the decision-making strategies used.

Rotating to a Terminal Solution

Once the number of factors is determined, the final step in exploratory factor analysis is to simplify the interpretation of the factors. This is accomplished via rotation. While a graphical rotation illustrates the process more clearly, an analytical rotation is more interpretable. Choices for analytical rotation in statistical packages are classified as either orthogonal or oblique. Orthogonal rotations produce a solution in which factors are uncorrelated. The most commonly used orthogonal rotation is varimax. Oblique rotations produce correlated factor solutions.

The most commonly used procedures for oblique rotation include Promax and direct oblimin. The consensus is that oblique rotation is the preferred method of analytical rotation (Conway, & Huffcutt, 2003; Fabrigar, et al., 1999). Oblique rotation is warranted particularly in a situation where theoretical and empirical evidence exists to suggest an underlying correlation among the constructs. Using orthogonal rotation will force the factors into an uncorrelated solution. An oblique rotation is more likely to produce a simple factor structure, and will also show if the factors are uncorrelated. If this occurs, an orthogonal rotation can be utilized. Factors are then named based on the underlying theme that the variables within a factor share.

Reporting EFA Results

When reporting the results of an exploratory factor analysis, it is important to fully report the analytical decisions made during the EFA process. Include the target population and sample size. Report the method of extraction, as well as the strategies used to determine the number of factors to retain. When reporting the results of an oblique rotation, include both the pattern and structure matrices. This allows readers to evaluate decision-making, and rotate results to alternative criteria (Henson, & Roberts, 2006). When giving the results of either oblique or orthogonal rotations, include communalities as a part of the pattern/structure matrix table. This allows the reader to determine the tightness of the solution and what, if any, effect sample size may have had on the stability of the loadings (MacCallum, et al., 1999).

Conclusion

Exploratory factor analysis is a very powerful procedure that can be utilized when developing a new instrument. The following is a summary of the steps to follow:

1. Plan for good population recovery by collecting as large a sample as possible;

2. Choose an appropriate extraction method [i.e., common factor model (like principle axis factoring or maximum likelihood factoring)];
3. Use multiple methods for determining the number of factors to retain;
4. Choose a rotation method appropriate to the theoretical and empirical evidence of correlation among factors; and
5. Accurately report all steps of the decision-making process.

References

Baldwin, C., & Caldwell, L. (2003). Development of the free time motivation scale for adolescents. Journal of Leisure Research, 35(2), 129-151.

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

Conway, J., & Huffcutt, A. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. Organizational Research Methods, 6(2), 147-168.

Fabrigar, L., Wegener, D., MacCallum, R., & Strahan, E. (1999). Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods, 4(3), 272-299.

Garson, G.D. (2006). Factor analysis. Retrieved April 10, 2006, from:
www2.chass.ncsu.edu/garson/pa765/factor.htm

Gorsuch, R. (1983). Factor analysis. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Habing, B. (2003). Exploratory factor analysis. Retrieved June 23, 2007, from:
http://www.stat.sc.edu/~habing/courses/530EFA.pdf

Henson, R., & Roberts, J. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. Educational and Psychological Measurement, 66(3), 393-416.

Hogarty, K., Hines, C., Kromrey, J., Ferron, J., & Mumford, K. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. Educational and Psychological Measurement, 65(2), 202-226.

Kaiser, H. (1960). The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Kim, J. (1970). Factor analysis. In N. Nie, C. Hull, J. Jenkins, K. Steinbrenner, D. Bent (Eds.), SPSS: Statistical package for the social sciences (468-514). New York: McGraw-Hill.

Kim, J., & Mueller, C. (1978). Introduction to factor analysis: What it is and how to do it. Newbury Park, CA: Sage.

MacCallum, R., Widaman, K., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. Psychological Methods, 4(1), 84-99.

Mundfrom, D., Shaw, D., Ke, T. Knofczinski, G., & McFann, K. (2003, April). Minimum sample size for exploratory factor analysis. Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

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