STRUCTURAL EQUATION MODELING, *12*(1), 173–181

BOOK REVIEW

Making Sense of Factor Analysis. Marjorie A. Pett, Nancy R. Lackey, and John J. Sullivan. Thousand Oaks, CA: Sage. 2003, 348 pages, $34.95 paperback.

Reviewed by Dale N. Glaser
Glaser Consulting, San Diego, California

Factor-analytic texts of varying complexity have been published through the years, with texts authored by Mulaik (1972) and Gorsuch (1983) particularly receiving prominent mention. Generally, the mathematical rudiments of factor analysis (FA) need to be explicated to reinforce a comprehensive grasp of FA. Given that the statistical machinations of FA can indeed be formidable for the introductory researcher or student, for initial purposes, a primer-type chapter on FA may suffice (e.g., Grimm & Yarnold, 1995; Tabachnick & Fidell, 2001). However, once the user is interested in developing an instrument designed for a particular purpose (e.g., screening, assessment, etc.), it is imperative that a more than cursory grasp of psychometric principles be obtained. In many of my consultations, I have come upon assessment devices with designated rubrics (e.g., for a customer satisfaction survey, categories labeled as responsiveness or courtesy), with those rubrics serving as a placeholder for logically related items. Even though logical clustering of items may be a valuable start to survey construction, the dimensionality of the data structure may come into question if statistical or psychometric evidence (e.g., construct validity) does not support the formation of the item clusters. Thus, in fields such as health care, in which assessment is actively pursued, a familiarity with testing and measurement principles, although often neglected, should be a prerequisite prior to test construction. Hence, it is timely that a text specifically geared to instrument development in health care research has been published, that being *Making Sense of Factor Analysis.* As is made clear in this review, the three authors (Marjorie Pett, Nancy Lacky, and John Sullivan) have provided a text that is accessible and sufficiently technical that the introductory user will not feel intimidated or overwhelmed with the complexities of FA. As the authors make clear in the introduction "factor analysis is not a unidimensional approach ... factor analysis involves a series of complex statistical techniques that involve higher order mathematics. There is also much subjectivity" (p. xiv).

Requests for reprints should be sent to Dale N. Glaser, Glaser Consulting, 4003 Goldfinch Street, San Diego, CA 92103. E-mail: glaserconsult@sbcglobal.net

In the first chapter, titled "An Overview of Factor Analysis," the authors delineate the characteristics of FA and what would be the impetus (i.e., data reduction, construct validity) to use such a technique, as opposed to other methods such as analysis of variance, *t* test, and so on. A brief contrast is made between exploratory and confirmatory factor analysis and the attendant assumptions for exploratory factory analysis (EFA). The next section will be familiar to many readers knowledgeable in the domain of intelligence, as the development of FA corresponds closely to advances made in the measure of intelligence (i.e., Spearman and Thurstone's contributions). The acceleration in the use of FA since the 1950s and the reasons for the growth (e.g., development of computers) further confirms the contribution of this technique to test construction. Moreover, although the authors note that psychometric training in the health care discipline has not been as rigorously applied as in the psychological and educational areas, there has been marked growth in the use of FA in health care in the past 15 years.

Chapter 2, "Designing and Testing the Instrument," commences with a description of types of measurement frameworks, such as criterion-referenced and norm-referenced instruments. Those versed in more advanced psychometric theory a la Nunnally and Bernstein (1994) will find these sections a bit on the terse side, although for the introductory reader the brevity will suffice. In the appendix, the authors reference an instrument titled the Concerns About Genetic Testing Scale (CGTS), and much of the subsequent discussion in this text (e.g., norm-referenced instruments) incorporates this instrument. The authors do make a reference to the interval or ratio-level measurement produced by norm-referenced instruments, and suggest that this "can be evaluated by factor analysis" (p. 15). However, as many versed in structural equation modeling (SEM) or item response theory are aware, FA can indeed be conducted with binary data. The next section reviews the notion of latent variables, a domain of much familiarity to readers of this journal. The differentiation between direct and indirect measures is detailed, delineating the requisite steps for defining empirical indicators (e.g., determining the construct of interest, conceptualizing the construct, etc.). This "how-to" section will be valuable to fledgling developers of test instruments, although some SEM researchers may not necessarily agree with the sequence of steps (i.e., operationalizing the construct prior to the literature review). The step of concept analysis may yield some intrigue for seasoned SEM users given its emphasis on determining "the attributes, including the antecedents and consequences of the selected construct" (p. 22). One major shortcoming of this section is the lack of references to those who have made major strides or contributions to the study of latent variables (e.g., Bollen, 2002). In fact, the references in this section are to texts that at best provide a cursory review of latent variables; thus, it will be incumbent that the reader investigate the primary contributors to this area (with a recent article by Borsboom, Mellenbergh, & Van Heerden, 2003, continuing this discussion). Qualitative methods such as phenomenology, naturalistic inquiry, and focus groups are reviewed as alternatives to the identification of empirical indicators. Although most SEM users may be primarily oriented toward quantitative

methods, given the context of nursing research, qualitative research is a well-developed (and often-used) methodology, frequently used in concert with more quantitative techniques.

The next section concerns the global subject of development of the instrument, encompassing instrument format, scaling issues, wording, layout, instructions, and so on. Again, much of this will be familiar to the seasoned developer, but will be a helpful tutorial to the neophyte. Even though it is acknowledged that there are different scale formats, the Likert scale is the primary focus, given it is "one of the most commonly used scaling techniques in psychosocial and health care research" (p. 32). Much of the citation up to and including this section is based on DeVillis's (1991) text, so the reader is urged to also explore other, more updated texts that involve survey research and test construction. Even though one can argue with some of the cursory recommendations such as the use of positively and negatively worded items (the problem being that the wording may compromise construct validity), for the most part the guidelines for survey construction as provided are sufficiently comprehensive. The authors also recommend the use of marker variables as they "are very useful in a factor analysis because they help to clearly define an extracted factor. It also means that fewer subjects per item will be required for the analysis" (p. 41). It would be helpful for the authors to elaborate on the sample size comment and specifically how a marker item reduces the sample size requirement, as the required n for FA is a complex function of the interrelation of unique factors, magnitude of communalities, and overdetermination of factors (MacCallum, Widaman, Zhang, & Hong, 1999). There are also some useful hints about determining the proper number of items, although ideally (and comprehensively) tapping the construct of interest should be the researcher's priority.

The third chapter, titled "Assessing the Characteristics of Matrices," presents an overview of the matrices involved in FA, primarily via SPSS and SAS software. Even though this text is not intended to delve into the mathematics of FA, a rudimentary introduction to the nature of matrices (e.g., dimensions, properties, functions) is provided as a cursory background for the reader. As the book indicates, the reader is referred elsewhere (e.g., Gorsuch, 1983) for a more extended treatment of matrix algebra. Using the CGTS as an example, the fundamentals of matrices, and specifically the correlation matrix, are reviewed. The constituent elements in the correlation coefficient (i.e., covariance, standard deviation, etc.) are described and pointed out in a comprehensive appendix with SAS output. Even though the parallels drawn between the covariance and correlation are necessary, no mention is made of z scores, and how the correlation (although it is mentioned as being "dimensionless") is essentially a standardized covariance. Also, the "suggested rules of thumb" for assessing magnitude (i.e., r^2), although admittedly "not written in stone" (p. 60), are substantively different than Cohen's (1988) offered taxonomy (e.g., $r = .5$ for large correlation per Cohen, whereas in the reviewed text a "strong" correlation is .70–89). Again, the phenomenon under investigation and the historical effect sizes associated with the domain should influence the determination of

what is a small, medium, or large effect. Calculations of the correlation, based on the software output, are provided, with a subsequent detailing of the covariance and identity matrix. Next the determinant and its function in calculating the inverses of a matrix are detailed. The authors also provide an example of calculating the determinant for a matrix larger than 2×2 via the augmentation approach. This somewhat deviates from other multivariate texts (e.g., Stevens, 2002) that arrive at the same calculation via the method of cofactors. Problems associated with matrix inversion as the determinant approaches zero are discussed, enveloping discussion of linear dependencies and singularity. The authors caution that a solution may be obtained in the various software programs (e.g., SPSS) even though the determinant may be very close to zero; thus, interpretation of the output is problematic. Suggestions to deal with ill-conditioned matrices are some that experienced readers will be aware of (e.g., check for duplication, check for high variable intercorrelations, etc.). However, one suggestion is that "if there are similar response sets among subjects, you may need to drop one or more of the subjects from the analysis" (p. 72). I am not sure that is entirely advisable in that similar response sets may be an accurate reflection of the phenomenon or construct being tapped; also, deletion of participants based on similar response sets may appear to be arbitrary and difficult to defend. A generous review of tests of matrices follows, with examples of two tests generated in SPSS: Bartlett's Test of Sphericity and the Kaiser–Meyer–Olkin (KMO) Test, and a Z approximation, $(\chi^2 - (df)) / \sqrt{2(df)}$, that is generally not included in software output. The steps needed to obtain Bartlett's Test and the KMO in SPSS are clearly explicated, with examples of each provided.

The next chapter, titled "Extracting the Initial Factors," is critical insofar as the prevailing practice of resorting to the software's default extraction, that generally being principal component analysis (PCA). However, as made clear by many researchers the selection of extraction method should not be based on the software's default, but rather the objective of the researcher (Marcoulides & Hershberger, 1997; Tabachnick & Fidell, 2001). The chapter starts off recommending the use of the correlation matrix (as opposed to the covariance matrix) for EFA, although the same recommendation is generally not the case for confirmatory factor analysis. The terminology associated with the common factor model is presented (i.e., common variance, specific variance, etc.), even though some clarification is needed as the authors state "common variance, symbolized as h^2, represents the amount of variance that is shared among a set of items" (p. 87). Many users of EFA would characterize communality (i.e., h^2) as the proportion of variance shared among the factors for a specific item (which the authors do correctly describe on the following page, and describe as *item communality* on p. 100). The authors then make the distinction of PCA and common factor analysis, the latter not to be confused with confirmatory factor analysis, and the decomposition of variation. They then briefly highlight the various approaches that fall under common factor analysis, including principal axis factoring, maximum likelihood, and so on. A more detailed,

step-by-step example of extraction for two components (e.g., eigenanalysis, convergence, normalization, etc.) via PCA is provided in the Appendix. The derivations made to arrive at the eigenvalues depart somewhat from the more terse matrix algebra of other texts, but it is made very intelligible to the introductory reader.

The description of PCA in the text covers communalities, eigenvalues, and eigenvectors, as well as certain conditions associated with eigenanalysis (i.e., positive definite matrices). The notion of partialing out variation (and orthogonality) is also discussed in the context of calculating components beyond the first obtained principal component. The method of factor extraction via SPSS and SAS is provided, with the attendant output. The process of reproducing the correlations and deriving the eigenvalues via squared factor loadings is explicated.

After a brief review of the critiques of PCA (e.g., appropriateness of rotation), a description of common factor analysis follows. The strategies by which to arrive at the initial communalities are reviewed, with principal axis factoring (PAF) receiving prominent attention (note that this is the default for many software settings). The authors discuss the iterative process of arriving at the final estimates, and the criteria for convergence. They also recommend that if convergence is not obtained (based on the defaults of the software) that "you may want to use PCA during the initial analyses, at least until some of the items have been discarded" (p. 104). Although this is not an unreasonable suggestion, there are some researchers who would encourage the factor analyst to closely examine the items (and attendant distributions) if convergence is not obtained instead of quickly pursuing a statistical alternative. With an alternative extraction, although it may culminate in convergence, it may also disguise problematic methodology and item construction. The distinction between common variance and unique variance (consisting of both specific and error variance) is briefly defined with detailed examples of reproducing the correlation matrix via PAF and obtaining item communalities following. The output in the chapter is based on SPSS, although the SAS output and accompanying syntax is included in the Appendix. Differences and similarities between the software and the output are also reviewed. Advantages (e.g., unique values, includes error of measurement) and disadvantages (e.g., sample specificity of R^2 values, Heywood cases) of PAF are also delineated. A brief description of alternative extraction methods follows (e.g., unweighted least squares, maximum likelihood, etc.). Other, more comprehensive texts should be sought if the reader is interested in a more elaborate treatment of these methods.

The authors recommend (with the caveat that there are critics of this approach) that

> For an exploratory factor analysis, you start with a PCA solution, solve the problems associated with it ... and come up with a preliminary solution. Then compare these results with a PAF solution on the same matrix and pick the one that is the best fit and that makes the most intuitive sense. (pp. 114–115)

As the authors intimate, there will be critics and adherents to such an approach, especially given the different motives and mathematics associated with PCA and PAF. For a more detailed examination of PCA versus PAF, see Velicer and Jackson (1990). Following the discussion of extraction methods, the factor selection process is detailed. Well-known suggestions include eigenvalue > 1, examination of the scree plot, and percentage of variance extracted. However, they do not delve into alternative methodologies such as minimum average partial rule or parallel analysis method (Zwick & Velicer, 1986), which have shown promise in component retention. Recommendations are then made on the number of factors to retain with the advice rendered "the researcher should err on the side of selecting too many rather than too few factors" (p. 125). Pros and cons of such a strategy could be argued, and it might be worthwhile to investigate the recent research on under- versus overextraction given the type of extraction method.

The next chapter addresses factor rotation. Initially the notion of simple structure (and the attendant conditions) is reviewed and the role it plays in enhancing factor interpretability is discussed. The often-discussed two types of rotation are covered: orthogonal and oblique. To clarify the notion of rotation, a geometric illustration is provided. For a more detailed geometric description, the chapter on PCA and PAF in Wickens (1995) may be of interest. The types of orthogonal rotation explored are the ones generally found in the popular software (i.e., varimax, quartimax, and equamax). The authors state that "varimax ... is the default option in both SPSS for Windows and SAS." (p. 141) However, that is not entirely true because one has to select varimax among the rotation options in SPSS; the default is "no rotation." One problem many versed in EFA have noted is indeed the reliance on the default, as opposed to a thoughtful strategy on the type of rotation that fulfills the objective of the research pursuit. Moreover, many theorists discount the notion (or reality) of a statistical procedure that imposes an uncorrelated component structure. As the authors state, "this assumption is rarely met in health care research" (p. 149); thus, many opine that an oblique rotation should be the prevailing rotation method. However, this method does yield some added complexities given the various matrices (pattern and structure matrix). Even though some have suggested that the pattern matrix be the focus of interpretation (Tabachnick & Fidell, 2001), there have been arguments made that both the structure and pattern matrix be reported and interpreted (Bruce Thompson has done some work in this area). The different types of oblique rotation (direct oblimin, promax, etc.) are then detailed. As with each of the preceding chapters, the computer output from SPSS is reviewed and interpreted. (I recently spoke to a colleague in health care who uses SPSS and recently purchased this text and she found this step-by-step interpretation of the output to be very helpful.)

The next chapter is titled "Evaluating and Refining the Factors." After a brief reiteration of the "art" of EFA interpretation, and the examination of pattern versus structure matrices, a strategy of setting the output to generate loadings only greater than a certain value (e.g., |.4|) is discussed. The notion of item-total correlations

and creating a cutoff for item deletion (e.g., |.3|) is suggested. However, it is also noted that at times an item that is deemed conceptually noteworthy does not yield a substantive loading on any of the factors. Thus, the researcher needs to make a decision about the implications of maintaining the problematic item. It is also necessary to keep in mind that unless factorial invariance across multiple strata has been supported, sampling error may be the root of an item with a low loading. Thus item deletion (or inclusion) should not be a hasty decision. This same line of thinking dovetails with the following discussion in the text regarding items that cross-load across multiple factors. The suggestion is made (which I agree with) that the item should be retained on the factor that bears the most conceptual sense. The rest of the chapter primarily delves into evaluating the internal consistency of an instrument, reviewing split-half reliability and Cronbach's alpha. As well, the implication of increasing or decreasing the number of items on the reliability estimate is detailed using the Spearman–Brown formula. Strategies for item deletion, using such statistical output as the corrected item-total correlation or coefficient alpha if item deleted in SPSS for Windows are detailed, keeping in mind some of the same caveats as EFA when item deletion is considered. Even though the extended discussion of internal consistency is germane to psychometric testing, it is unusual to couch it, in such a detailed fashion, within the context of EFA. Generally, the factor analyst has made the decision of item inclusion or deletion after an in-depth examination, and then the testing of reliability estimate is made subsequent to a decision about the formation of the item parcels. Even though the strategy of assessing internal consistency may be helpful in the instance of cross-loading items, there should still be a conceptual basis for item deletion and inclusion.

The penultimate chapter, titled "Interpreting Factors and Generating Factor Scores," initially reviews the interpretation of factors, emphasizing the magnitude of the loadings as well as their conceptual interconnectedness. Some guidelines are offered for item inclusion (e.g., the shared variance of the item, such as loading = .71 equaling 50% of shared variance); however, as with any guidelines, they should only be loosely adhered to. Strict conformance to the guidelines may culminate in under- or overextraction given the pattern and strength of loadings. As the authors state, "the researcher needs to consider the breadth and complexity of the factor as well as its relationship to the initial conceptualization" (p. 209). "Naming the factor," a process that is part art and part science, is detailed next. Readers may want to especially take note of the following passage: "If the items for the factor analysis were derived from theory or from a conceptualization, the researcher should return to the original theory or conceptualization to name the factors" (p. 210). If indeed the items parallel the initial conceptualization (or a priori theory), one could argue the researcher could have proceeded directly to conducting a confirmatory factor analysis. Although criteria for using EFA versus confirmatory factor analysis have been bandied about (e.g., Hurley et al., 1997), it is up to the researcher's judgment to assess when the formulation of the constructs are theoretically sound enough to

bypass EFA and proceed to confirmatory factor analysis. One alternative that has been recommended is to randomly split the database and conduct EFA on one half and confirmatory factor analysis on the other. The next section of this chapter reviews factor scores and their potential use "in subsequent analysis" (p. 213). The controversy and implications of factor score indeterminacy when using such scores from common factor analysis is discussed, although for a more recent broaching of this topic, the reader is recommended to avail themselves of a special section in *Multivariate Behavioral Research* (Vol. 31, No. 4, 1996). For the final section, the pros and cons of using factor scores (which use all the items in their derivation) as opposed to factor-based scales that are "obtained by summing the scores for only those individual items that have been selected for inclusion on a given factor" (p. 221) is highlighted.

The final chapter tackles the subject of "Reporting and Replicating the Results." The authors assert that:

> An instrument is ready for reporting when the developer can clearly state the conceptual bases for the instrument, how it was constructed, the minimal reliability and validity that have been established, and the type of subjects and research situations in which the instrument can be used. (p. 227)

It may also behoove the researcher to consult the most recent version of *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) for further elaboration on these related topics. One point that may need clarification is when the authors recommend that "at least three studies be completed before a report of the instrument development" (p. 227). The first being content validity, the second being "field testing of the items ... [including] initial reliabilities of both the instrument as a whole and the subscales if appropriate" (p. 227), and the third being the testing of construct validity. However, it can be argued that the second and third steps are inextricably intertwined, for the reliabilities of an instrument (as per the text) are estimated after the factorial structure of the instrument has been established. Indeed, some would opine that EFA is a method by which to test construct validity. Various elements for what should be included in a factor-analytic report are offered, including the design and administration of the instrument, types of tables, rotation and extraction methods, and so on. The authors conclude the chapter (and text) with the comment that "construct validity is a never-ending, ongoing, complex process that is determined over a series of studies in a number of different ways" (p. 239). To support that claim, one can peruse the issues of recent assessment-oriented journals (including *Structural Equation Modeling*) to view many articles that have examined the invariance of factorial structures for certain tools across multiple strata, as well as articles that have assessed the stability of the factorial structure across time.

Overall, *Making Sense of Factor Analysis* is a welcome text for those eager to tutor themselves in the complex technique of EFA. Thus, this text goes beyond the single chapters dedicated to EFA found in many multivariate texts, but falls short of the more mathematically demanding texts by Gorsuch, Mulaik, and others. Even though the emphasis is on health care research, there is nothing that precludes the incorporation of this text in other disciplines. One recommendation I would make is, although this is not technically a research text, the citations are fairly predictable and, to some degree, antiquated. A cursory glance at journals such as *Psychological Methods, Multivariate Behavioral Research, Applied Psychological Measurement, Educational and Psychological Measurement, British Journal of Mathematical and Statistical Psychology,* and others, would serve in addressing more recent topics such as factor underextraction and overextraction, alternative methodologies for factor extraction, factorial invariance, and so on.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology, 53,* 605–634.

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110,* 203–219.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

DeVillis, R. F. (1991). *Scale development.* Newbury Park, CA: Sage.

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

Grimm, L. G., & Yarnold, P. R. (1995). *Reading and understanding multivariate statistics.* Washington, DC: American Psychological Association.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., et al. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior, 18,* 667–683.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84–99.

Marcoulides, G. A., & Hershberger, S. L. (1997). *Multivariate statistical methods: A first course.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mulaik, S. A. (1972). *The foundation of factor analysis.* New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25,* 1–28.

Wickens, T. D. (1995). *The geometry of multivariate statistics.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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