Psychological Methods
2001, Vol. 6, No. 4, 430–450

1082-989X/01/$5.00 DOI: 10.1037//1082-989X.6.4.430

Computing and Evaluating Factor Scores

James W. Grice
Oklahoma State University

A variety of methods for computing factor scores can be found in the psychological literature. These methods grew out of a historic debate regarding the indeterminate nature of the common factor model. Unfortunately, most researchers are unaware of the indeterminacy issue and the problems associated with a number of the factor scoring procedures. This article reviews the history and nature of factor score indeterminacy. Novel computer programs for assessing the degree of indeterminacy in a given analysis, as well as for computing and evaluating different types of factor scores, are then presented and demonstrated using data from the Wechsler Intelligence Scale for Children—Third Edition. It is argued that factor score indeterminacy should be routinely assessed and reported as part of any exploratory factor analysis and that factor scores should be thoroughly evaluated before they are reported or used in subsequent statistical analyses.

Exploratory factor analysis is a widely used multivariate statistical procedure in psychological research. Undoubtedly one of its appeals is the ability to compute scores for the individuals in the analysis on the extracted factors. These novel *factor scores* can be used in a wide variety of subsequent statistical analyses. For instance, they can be correlated with measures of different constructs to help clarify the nature of the factors or they can be entered as predictor variables in multiple regression analyses or as dependent variables in analyses of variance. Factor scores are also routinely computed as simple sum scores in the scale development process and are often referred to as scale, composite, or total scores. Given these varied uses, it is not surprising that factor scores are common in the psychological literature.

Computing factor scores is not a straightforward process, however, as researchers must choose from a number of competing computational methods. For example, consider three researchers who conduct an exploratory common factor analysis on a questionnaire. Each decides on the same number of factors; uses the same extraction, rotation, and interpretation procedures; and then decides to compute scores for the identified factors. The first researcher selects the default option from his or her favorite computer program, resulting in continuous factor scores with means equal to 0 and standard deviations close to 1. The second researcher examines the structure coefficients (the correlations between the items and the factors) for salient items using a conventional criterion such as .30 or .40. The factor scores are then computed by summing the raw or standardized scores for the items that are deemed salient, much like the procedures employed in scale construction. The final researcher also uses a simple summing procedure but selects the salient items from the factor score coefficient matrix (regression weights for predicting the factors from the items) rather than from the structure coefficient matrix. All three researchers will likely arrive at a different set of factor scores, possibly yielding widely discrepant rankings of the individuals along the extracted factors. Which set of factor scores is most appropriate, and should yet another method for computing the factor scores have been chosen? The purpose of the present article is to help researchers and practitioners address these important questions.

It is widely known that decisions made regarding the various extraction and rotation methods used in an

Development of this article was supported by a summer research fellowship awarded to James W. Grice from Southern Illinois University at Edwardsville. I thank Richard J. Harris at the University of New Mexico for his comments on an earlier draft of this article. Naturally, any errors that remain are my own.

Correspondence concerning this article should be addressed to James W. Grice, Psychology Department, 215 North Murray, Oklahoma State University, Stillwater, Oklahoma 74078. Electronic mail may be sent to jgrice@okstate.edu.

exploratory factor analysis can greatly influence the quality of the results (Comrey, 1978; Fabrigar, Wegener, MacCallum, & Strahan, 1999). Similarly, the choice made regarding how factor scores are computed can significantly affect their quality (Grice, 2001; Grice & Harris, 1998) as well as the outcomes of subsequent analyses in which the scores are used (e.g., see Morris, 1980). The present article hence reviews the different methods of computing factor scores as well as the history of the factor score indeterminacy debate that ultimately led to the creation of the various methods. A number of computer programs are then presented that provide the necessary tools for thoroughly evaluating the adequacy of a set of factor scores. These programs are demonstrated using a genuine data set from the Wechsler Intelligence Scale for Children—Third Edition (WISC–III; Wechsler, 1991), and practical implications are finally discussed.

Historical Review

The common factor model was conceived and developed as a vehicle of discovery in the early part of the 20th century. Charles Spearman, Cyril Burt, Louis Thurstone, and other notable psychologists believed that this model could be used to isolate and describe the fundamental categories of human ability. An early and particularly cogent description of the assumptive framework underlying the common factor model was offered by Thurstone (1935):

> If abilities are to be postulated as primary causes of individual differences in overt accomplishment, then the widely different achievements of individuals must be demonstrable functions of a limited number of reference abilities. This implies that individuals will be described in terms of a limited number of faculties. (p. 46)

The mathematical framework that provided the means for identifying these yet undiscovered primary cognitive abilities was initially developed by Spearman (1904) and then expanded by Garnett (1919a, 1919b) and Thurstone (1931, 1934). A number of important texts also introduced new methodologies or summarized the emerging mathematical structure and conceptual nature of the common factor model (e.g., Burt, 1940; Holzinger & Harman, 1941; Spearman, 1927; Thomson, 1939; Thurstone, 1935, 1947). Rotational procedures, the importance of simple structure, identification of the correct number of factors to extract, and the process of labeling factors by examining the loadings were all introduced in this early period and are topics familiar to the modern student of factor analysis. The latter part of the 20th century of course yielded a prodigious number of analytical and empirical developments that greatly refined and extended these early procedures. A number of extraction or parameter estimation procedures are now readily available, such as image, principal axis, alpha, or maximum likelihood factor solutions; the number of factors to extract can be evaluated according to a scree plot, parallel analysis, minimum average partial rule, or one of many other competing methods; rotation can be accomplished using varimax, quartimax, equamax, promax, or oblimin transformations, among others; and the factors can be interpreted on the basis of the structure, pattern, reference vector, or factor score coefficient matrices.

What may be unfamiliar to many modern consumers of factor analytic technology, however, is the controversy surrounding the common factor model. As early as the 1920s researchers recognized that, even if the correlations among a set of ability tests could be reduced to a subset of factors, the scores on these factors would be indeterminate (Wilson, 1928). In other words, an infinite number of ways for scoring the individuals on the factors could be derived that would be consistent with the same factor loadings. Under certain conditions, for instance, an individual with a high ranking on *g* (general intelligence), according to one set of factor scores, could receive a low ranking on the same common factor according to another set of factor scores, and the researcher would have no way of deciding which ranking is "true" based on the results of the factor analysis. As startling as this possibility seems, it is a fact of the mathematics of the common factor model. Not surprisingly, the discovery of factor score indeterminacy threatened the integrity of Spearman's *g* and competing multifactor theories of cognitive ability, and the ensuing debate grew fairly heated at times but was ultimately left unresolved. From the 1930s to the 1970s the issue was seemingly ignored by the popular purveyors of factor analytic technology. For example, Thurstone (1935) discussed procedures for computing factor scores—almost as a footnote—in the shortest and final chapter of *The Vectors of Mind* and made no direct mention of the indeterminacy debate. Spearman (1932) similarly made no reference to the debate in the second printing of *The Abilities of Man* (Steiger, 1996a). The motives behind this apparently nonchalant attitude are not clear. On the one hand, Steiger and Schönemann (1978) and Steiger (1996a) present evidence that suggests the issue was consciously ignored and perhaps suppressed. On the other hand, Lovie and Lovie (1995) argue that Spearman thought the indeterminacy problem had been settled, and Thurstone (1935) himself admitted that the principal problem of factor analysis was that of "isolating and identifying primary factors in a battery of traits" (p. 226)—calculating factor scores was a secondary concern. Regardless of the motives, however, the effects on modern practice have been indisputably negative because modern researchers routinely compute factor scores. These same researchers appear to be completely unaware of factor score indeterminacy. Consequently, factor scores are derived and left unevaluated, and their potentially adverse effects on the results of subsequent analyses are ignored.

Factor Score Indeterminacy

Several discussions of factor score indeterminacy are readily available. Steiger and Schönemann (1978) offer a thorough and readable summary of the problem, including a straightforward outline of the important contributions of Guttman's (1955) classic article on the subject. Maraun (1996) offers a concise presentation of factor score indeterminacy in matrix form and profiles two contemporary views of the issue, and Mulaik (1972, 1976) discusses the issue from a geometrical perspective. Given these excellent resources, indeterminacy will be discussed in a relatively nontechnical fashion here.

A simple look at the common factor model reveals the essential nature of the problem:

$$\mathbf{Z}_{nk} = \mathbf{F}_{n(f+k)}\mathbf{P}'_{(f+k)k}, \tag{1}$$

where n, k, and f indicate the dimensions of the matrices and are equal to the number of observations, items, and common factors, respectively. $\mathbf{Z}_{nk}$ is the matrix of standardized observed scores, $\mathbf{F}_{n(f+k)}$ is an augmented matrix of common and unique factor scores, and $\mathbf{P}'_{(f+k)k}$ is a transposed augmented matrix of common and unique factor pattern weights. If $\mathbf{P}_{k(f+k)}$ is an invertible matrix, a unique solution for the factor scores, $\mathbf{F}_{n(f+k)}$, can be found as follows:

$$\mathbf{F}_{n(f+k)} = \mathbf{Z}_{nk}\mathbf{P}^{-1}_{k(f+k)}. \tag{2}$$

The total number of common and unique factors, $f + k$, however, will exceed the number of items, k, and hence $\mathbf{P}_{k(f+k)}$ is not a square, invertible matrix in the common factor model. The consequence is that a unique solution for the factor scores does not exist because the model is attempting to define $f + k$ variables uniquely by k equations. In essence, one is faced with a situation in which the number of unknowns exceeds the number of equations, making an infinite number of solutions possible. It should be understood that factor scores can be computed for Equation 2 that satisfy the stipulations of the common factor model. For instance, factor scores can be computed that meet the following criterion for an orthogonal factor solution (Kestelman, 1952):

$$\mathbf{F}_{n(f+k)}\mathbf{F}'_{(f+k)n}n^{-1} = \mathbf{I}_{nn}. \tag{3}$$

In other words, the factor scores will be standardized and perfectly orthogonal. The indeterminacy problem is not that the factor scores cannot be directly and appropriately computed; it is that an infinite number of sets of such scores can be created for the same analysis that will all be equally consistent with the factor loadings.

What Equations 1, 2, and 3 fail to make explicit, however, is that factor scores that satisfy Equation 3 can be divided into indeterminate and determinate portions (Green, 1976; Guttman, 1955; Maraun, 1996; McDonald, 1974; Mulaik, 1976). The degree of indeterminacy will not be equivalent across studies and is related to the ratio between the number of items and factors in a particular design (Meyer, 1973; Schönemann, 1971). It may also be related to the magnitude of the communalities (Gorsuch, 1983). Small amounts of indeterminacy are obviously desirable, and the consequences associated with a high degree of indeterminacy are extremely unsettling. Least palatable is the fact that if the maximum possible proportion of indeterminacy in the scores for a particular factor meets or exceeds 50%, it becomes entirely possible to construct two orthogonal or negatively correlated sets of factor scores that will be equally consistent with the same factor loadings (Guttman, 1955). As mentioned previously, how individuals are ranked along the factor can therefore be completely different depending on which set of factor scores is chosen. This effect also carries over to other variables not included in the factor analysis, as Steiger (1979) showed that the relationships between indeterminate factor scores and external criteria are similarly indeterminate (see also Schönemann & Steiger, 1978). When the degree of indeterminacy is small, however, the competing sets of factor scores will all be highly correlated. As the proportion of indeterminacy approaches 0, wildly different rankings of particular individuals will also become impossible, and the relationships between the factor scores and external criteria will grow more determinate.

Factor Score Approximations

The indeterminate nature of factor scores is inextricably tied to the numerous methods that have been developed for their computation. These methods can be divided into two general classes (Horn, 1965), and in practice they yield scores that are only approximations of the factor scores that would satisfy Equation 3. The first class uses all of the (typically) standardized variables, z_is. For instance, the estimated factor score on Factor j for Person i can be represented as follows:

$$\mathbf{F}_{ji} = w_{j1}z_{i1} + w_{j2}z_{i2} + \ldots + w_{jk}z_{ik}. \qquad (4)$$

The regression weights, w_js, are multidecimal values referred to as factor score coefficients. There are a number of different approaches for obtaining these weights. Each places a different set of constraints on the estimated factor scores and seeks to minimize a particular estimate of error. The most popular solution is probably Thurstone's (1935) least squares regression approach in which the factor score coefficients are computed from the original item correlations, $\mathbf{R}_{kk}$ (subscripts represent the dimensions of the matrix), and structure coefficients, $\mathbf{S}_{kf}$ (the correlations between the factors and the items):

$$\mathbf{W}_{kf} = \mathbf{R}_{kk}^{-1}\mathbf{S}_{kf}. \qquad (5)$$

The standardized observed scores can then be multiplied by the matrix of factor score coefficients, $\mathbf{W}_{kf}$, to obtain the estimated factor scores:

$$\mathbf{F}_{nf} = \mathbf{Z}_{nk}\mathbf{W}_{kf}. \qquad (6)$$

The values found in $\mathbf{F}_{nf}$ represent only the determinate portion of the factor scores, which is maximized in this instance (conversely, the indeterminate portion is minimized), and hence will not satisfy the stipulations of the common factor model fully. For example, the estimated factor scores will often be intercorrelated when the factors are orthogonal and they will not have unit variance.

Additional linear prediction methods were developed as alternatives to the method shown in Equation 5. Anderson and Rubin (1956; see Gorsuch, 1983) developed a procedure for estimating factor scores that are constrained to orthogonality:

$$\mathbf{W}_{kf} = \mathbf{U}_{kk}^{-2}\mathbf{P}_{kf}\,(\mathbf{P}'_{fk}\mathbf{U}_{kk}^{-2}\mathbf{R}_{kk}\mathbf{U}_{kk}^{-2}\mathbf{P}_{kf})^{-1/2}, \qquad (7)$$

where $\mathbf{P}_{kf}$ represents the pattern coefficients and $\mathbf{U}_{kk}$ is a diagonal matrix of the reciprocals of the squared unique factor weights. This method was generalized by McDonald (1981) to oblique (i.e., correlated) factor solutions and developed further by ten Berge, Krijnen, Wansbeek, and Shapiro (1999):

$$\begin{aligned}\mathbf{W}_{kf} &= \mathbf{R}_{kk}^{-1/2}\,\mathbf{C}_{kf}\,\mathbf{\Phi}_{ff}^{1/2}, \text{ where } \mathbf{C}_{kf}\\ &= \mathbf{R}_{kk}^{-1/2}\,\mathbf{L}_{kf}\,(\mathbf{L}'_{kf}\,\mathbf{R}_{kk}\,\mathbf{L}_{kf})^{-1/2} \text{ and } \mathbf{L}_{kf}\\ &= \mathbf{P}_{kf}\,\mathbf{\Phi}_{ff}^{1/2}, \end{aligned} \qquad (8)$$

where $\mathbf{\Phi}_{ff}$ is the factor correlation matrix. This equation is appropriate even when the covariance matrix for the unique factors is nonsingular, although in practice this matrix is assumed to be diagonal. The correlations among estimated factor scores computed from the weight matrix in Equation 8 are constrained to match the correlations among the factors themselves. Bartlett (1937) developed a method that minimizes the sum of squares for the unique factors across the range of variables:

$$\mathbf{W}_{kf} = \mathbf{U}_{kk}^{-2}\mathbf{P}_{kf}(\mathbf{P}'_{fk}\mathbf{U}_{kk}^{-2}\mathbf{P}_{kf})^{-1}. \qquad (9)$$

Harman (1976) reports the "idealized variable" strategy based on the reproduced correlation matrix rather than on the original item correlations:

$$\mathbf{W}_{kf} = (\mathbf{P}_{kf}\mathbf{P}'_{fk})^{-1}\mathbf{P}_{kf}. \qquad (10)$$

Correlations between the factors and noncorresponding factor score estimates computed from the weight matrices in Equations 9 and 10 are constrained to 0 when the factors are orthogonal. The correlations among the estimated factor scores, however, will not necessarily match the correlations among the factors, and the proportion of indeterminacy will not be minimized for either set of factor score estimates.

Several other methods have also been developed (Heermann, 1963; Ledermann, 1939) that provide linear approximations of the factor scores using a complex, "refined" $\mathbf{W}_{kf}$ weight matrix. Previous authors (e.g., Gorsuch, 1983; Grice & Harris, 1998; Horn, 1965) have referred to this general class of scores as exact factor scores. In the present article, however, these estimated factor scores will be referred to as "refined factor scores."

The second class of methods involves a simplified ("coarse") weighting process. Specifically, the estimated factor score on Factor j for Person i is still of the form:

$$\mathbf{F}_{ji} = \sum_{i=1}^{k} w_{kj}z_{ik}. \qquad (11)$$

The values for w_{kj}, however, are restricted to simple unit weights. In other words, the multidecimal regression weights are replaced by values of +1, −1, and 0; hence factor scores are estimated by simply adding, subtracting, or ignoring the (typically) standardized scores on the original items. The simplified weights are determined by identifying salient items in the structure, pattern, or factor score coefficient matrices. It is common practice, for instance, to conduct an exploratory factor analysis and examine the structure matrix for salient values using some conventional criterion (e.g., .30 or .40). Factor scores are then estimated by summing either the observed or standardized scores of those items deemed salient (items with negative structure coefficients are subtracted rather than added, and items with nonsalient structure coefficients are ignored). This simple cumulative scoring scheme is quite popular and can be seen in the index scores of the WISC–III (Wechsler, 1991), the facet and domain scores of the Revised NEO Personality Inventory (Costa & McCrae, 1992), and a host of other scale and subscale scores for instruments from a wide variety of domains. Variations on this scoring strategy include (a) allowing a particular item to contribute only to the score of a single factor, even if it is salient on several factors, (b) omitting an item that is salient on more than one factor, and (c) examining the pattern or factor score coefficient matrices in lieu of the structure matrix for salient items. These variations will often lead to different factor score estimates for the same set of extracted factors. Because each procedure incorporates a simple cumulative scheme, however, these types of estimated factor scores will be referred to as "coarse factor scores" herein.

The refined and coarse factor scoring strategies grew directly out of the factor score indeterminacy debate. The former methods were developed to minimize the "damage" incurred from indeterminacy using several approaches; for example, by maximizing the correlation between the refined factor scores and their respective factors (i.e., maximizing the determinate proportion in each set of scores; Thurstone, 1935) or by eliminating the intercorrelations among the scores for orthogonal factors (Anderson & Rubin, 1956; Heermann, 1963). As alluded to previously, however, each of these methods suffers from one or more particular defects (McDonald & Burr, 1967) and should therefore not be considered to offer a solution to the indeterminacy problem. The inadequacy of a particular set of refined factor scores as representations of the common factors will in fact go completely unnoticed unless the researcher takes the time to evaluate their properties.

Coarse factor scoring methods were introduced as simple and efficient alternatives to the more complex weighting schemes used to compute refined factor scores (Cattell, 1952; Thurstone, 1947). Because the structure coefficients are commonly used to determine which items are summed, the unit-weighted coarse factor scores are also believed to be more consistent with the process of factor interpretation, which is typically based on the same coefficients. Another attractive feature of coarse factor scores is their stability across independent samples of observations relative to refined factor scores (Grice & Harris, 1998; Wackwitz & Horn, 1971). Despite these advantages, however, coarse factor scores suffer from a number of defects. For instance, they may be highly correlated even when the factors are orthogonal and they will be less valid representations of the factors in comparison with the refined factor scores. Another concern involves basing the coarse factor scores on the structure coefficients. Two recent Monte Carlo studies (Grice, 2001; Grice & Harris, 1998) suggest that this practice yields scores that are poor representations of the common factors. These studies showed that the correlations between the coarse and known factor scores were generally low when the former were computed on the basis of the structure coefficients. The coarse factor scores were greatly improved, however, when the factor score coefficient matrix was instead used to determine which items (as well as their signs) to include in the computations. Use of the structure coefficients appeared to be justified only when the factors were orthogonal and demonstrated unrealistic simple structure (Grice & Harris, 1998; Wackwitz & Horn, 1971).

Regardless of how one finally chooses to compute coarse factor scores, the scores should routinely be evaluated to detect consequential inadequacies. Moreover, such scores should be evaluated even when they are derived from the results of a principal-components analysis or image analysis where the component or image scores are completely determinate in nature. Simplified scores computed from either of these analyses may be highly correlated even when the components or factors are orthogonal and they may be relatively inaccurate representations of the very dimensions they are meant to quantify. As mentioned previously, such inadequacies are most likely to emerge if the structure coefficients are used to determine which items are salient and how they are to be weighted and summed. Simplified scores computed from principal-components and image analysis as well as those computed from a traditional common factor analysis should therefore be routinely evaluated.

Evaluating Factor Score Approximations

Procedures for evaluating factor scores also emerged from the indeterminacy debate and are readily available in equation form (Gorsuch, 1983, pp. 272–273; Guttman, 1955; Mulaik, 1976). Unfortunately, these procedures have scarcely made their way into modern statistical software and have consequently been ignored in practice. Fortunately, leading analysis packages are extremely flexible, and the necessary procedures can be programmed into their respective languages. Such programs for SAS are presented and demonstrated below using genuine data from a recently published study.

Scale scores from the WISC–III (Wechsler, 1991) for 215 children diagnosed with learning disabilities served as the example data set (for details, see Grice, Krohn, & Logerquist, 1999). Consistent with the initial, exploratory common factor analyses reported in the WISC–III manual (see pp. 188–194), four factors were extracted and transformed using an orthogonal varimax rotation. The resulting structure coefficients for the 12 subtests (Picture Completion, Information, Coding, Similarities, Picture Arrangement, Arithmetic, Block Design, Vocabulary, Object Assembly, Comprehension, Symbol Search, and Digit Span) are reported in Table 1. These values were compared with the loadings on pages 192 and 193 of the WISC–III manual yielding congruence coefficients of .95, .95, .86, and .79 for the verbal, performance, processing speed, and freedom from distractibility factors, respectively (note that the last two factors switched order in the present sample such that freedom from distractibility is the fourth factor rather than processing speed). Refined and coarse factor scores can be computed for each of these factors, although the WISC–III manual only incorporates the latter and refers to them as index scores.

Refined Factor Scores

A SAS interactive matrix language (IML) program was written to conduct the necessary analyses for evaluating the refined factor scores computed from Equations 5 and 6.[1] The unique output for the WISC–III data is presented in Appendix A. As shown, the factor score coefficients ($\mathbf{W}_{kf}$) are listed first and are followed by indeterminacy indices: (a) the multiple correlation between each factor and the original variables, ρ, as well as its square, ρ^2 (Green, 1976; Mulaik, 1976), and (b) the minimum possible correlation between two sets of competing factor scores, $2\rho^2 - 1$ (Guttman, 1955; Mulaik, 1976; Schönemann, 1971). The former index ranges from 0 to 1, with high values being desirable, and indicates the maximum possible degree of determinacy for factor scores that satisfy Equation 3. Its square, ρ^2, represents the maximum proportion of determinacy. The second indeterminacy index ranges from −1 to +1, and high positive values are desirable. As discussed previously, when $\rho \leq .707$ (at least 50% indeterminacy), $2\rho^2 - 1$ will be less than or equal to 0, meaning that two sets of competing factor scores can be constructed for the same common factor that are orthogonal or even negatively correlated. Values for ρ that do not appreciably exceed .71 are therefore particularly problematic. The results for the first three extracted factors of the WISC–III (see Appendix A) are all above .80 and

Table 1
Structure Coefficients for Varimax-Rotated Factors From the WISC–III

Subtest	Structure coefficients 1	2	3	4
Picture Completion	.248	**.606**[a]	.164	.068
Information	**.701**[a]	.148	.091	.214
Coding	−.021	.137	**.679**[a]	.074
Similarities	**.695**[a]	.211	.013	.070
Picture Arrangement	.256	**.427**[a]	**.414**	.071
Arithmetic	.274	.031	.291	**.510**[a]
Block Design	.189	**.730**[a]	.237	.265
Vocabulary	**.793**[a]	.217	.092	.139
Object Assembly	.127	**.735**[a]	.097	−.057
Comprehension	**.579**[a]	.140	.136	**.306**
Symbol Search	.143	.198	**.728**[a]	.048
Digit Span	.132	.058	−.019	**.492**[a]

Note. Salient structure coefficients (≥.30) are in bold.
[a] Corresponding structure coefficients that were salient in the original Wechsler Intelligence Scale for Children—Third Edition (WISC–III) standardization sample.

[1] This program as well as other programs described in this article can be downloaded from James W. Grice's Web site at http://psychology.okstate.edu/faculty/jgrice/factorscores/.

seem sufficient, but the $2\rho^2 - 1$ values are fairly low and suggest that the best possible factor score estimates may still be too indeterminate. The results for the fourth factor, freedom from distractibility, however, are clear because ρ is only .683 and $2\rho^2 - 1$ is −.066. Such low values indicate that two orthogonal sets of factor scores could be created that are both equally consistent with the factor loadings. Therefore, refined or coarse factor scores should not be estimated for this factor. High degrees of indeterminacy may also be an impetus to reexamine the scree plot or factor-selection criterion and consider extracting fewer factors (see Schönemann & Wang, 1972).

The ρ values represent upper bounds on the determinacy of factor score estimates that can be computed for each of the factors. The refined factor scores that are actually computed, however, may have lower proportions of determinacy. The *validity coefficients* reported next in the output (see Appendix A) provide the means for assessing this possibility. These values represent the correlations between the factor score estimates and their respective factors, and may range from −1 to +1. They should be interpreted in the same manner as ρ described previously. Gorsuch (1983, p. 260) recommends values of at least .80, but much larger values (>.90) may be necessary if the factor score estimates are to serve as adequate substitutes for the factors themselves. In the present example, the validity coefficients are equal to the ρs because the refined approach in Equation 5 minimizes the proportion of indeterminacy (i.e., it maximizes validity) in the estimated factor scores. The refined methods in Equations 8, 9, and 10, however, minimize different functions and may therefore produce lower validities that must be scrutinized in relation to ρ.

Another useful criterion for evaluating factor scores is *univocality,* which represents the extent to which the estimated factor scores are excessively or insufficiently correlated with other factors in the same analysis. As shown in Appendix A, two matrices, labeled UNIV and FACTCOR, are to be compared to assess univocality. The values in the latter matrix are the interfactor correlations, which are all 0 for the present set of orthogonal factors. The values in the former matrix are the correlations between the refined factor scores (the rows) and the other, noncorresponding factors in the analysis (the columns). For example, .082 in the UNIV matrix represents the correlation between the second factor and the refined factor scores for the first factor, whereas .080 represents the correlation between the first factor and the refined factor scores for the second factor. The values in the UNIV matrix should match those in the FACTCOR matrix if the estimated factor scores are univocal. The results for the WISC–III factors are fairly good because most values are similar across the two matrices and the maximum absolute difference is only .170. These results therefore indicate that the estimated factor scores are not heavily contaminated by variance from other orthogonal factors in the same analysis.

The final criterion for evaluating the factor scores is *correlational accuracy,* which indicates the extent to which the correlations among the estimated factor scores match the correlations among the factors themselves.[2] This criterion can be assessed from the final two matrices shown in Appendix A. The SCORECOR matrix represents the correlations among the refined factor scores, and the FACTCOR matrix again represents the correlations among the factors. The estimated factor scores reveal superior levels of correlational accuracy when the values in the two matrices match. In the present example, the correlations among the refined factor scores compare favorably with the correlations among the four orthogonal factors extracted from the current WISC–III data (see Appendix A). The coefficients for the refined factor scores in the SCORECOR matrix are generally small, and the largest difference between the two matrices is for the joint first and fourth factors (r = .191).

In summary, the refined factor scores for the WISC–III four-factor orthogonal solution were found to possess several desirable and undesirable characteristics. The multiple correlations, ρ, appeared adequate for three of the factors, but the $2\rho^2 - 1$ values seemed low. These two indeterminacy indices for the

[2] The validity coefficients are taken from the diagonal of $\mathbf{R}_{fs}$, a matrix of correlations between the f factors and s factor scores, which is computed as follows:

$$\mathbf{R}_{fs} = \mathbf{S}'_{fk}\mathbf{W}_{kf}\mathbf{L}_{ss}^{-1}, \tag{12}$$

where $\mathbf{L}_{ss}$ is a diagonal matrix of factor score standard deviations. These values are the square roots of the diagonal elements of $\mathbf{C}_{ss}$, which is computed as follows:

$$\mathbf{C}_{ss} = \mathbf{W}'_{fk}\mathbf{R}_{kk}\mathbf{W}_{kf}. \tag{13}$$

The off-diagonal elements of $\mathbf{R}_{fs}$ constitute the values for univocality (see Gorsuch, 1983, p. 273). The correlational accuracy values are computed by calculating the Pearson product-moment correlations among the estimated factor scores.

freedom from distractibility factor, however, were clearly too low. The validity coefficients matched the ρ values, and the refined factor scores demonstrated admirable levels of univocality and correlational accuracy for all four factors; that is, the factor scores were not overly contaminated with variance from other orthogonal factors in the same analysis, nor were they highly correlated with one another.

The factor analysis options of the program described previously can be adjusted to derive oblique factors. Also, the program itself can be slightly modified to compute refined factor scores on the basis of Equation 10 rather than on Equation 5. These modifications were made to provide another example of the process of evaluating indeterminacy, validity, univocality, and correlational accuracy (see Footnote 1). Four factors were again extracted, but an oblique, promax transformation was applied to the factors. The pattern coefficients and factor correlations are presented in Table 2. The loadings could not be compared to the WISC–III manual because results for oblique rotations were not reported. It is interesting to note, however, that the four orthogonal factors from the exploratory analyses were allowed to become correlated in the confirmatory factor analyses (see Wechsler, 1991, pp. 191–195 and p. 281).

The unique output generated by the evaluation program for the oblique factors and the refined factor scores is presented in Appendix B. As shown, the multiple correlations for the first three factors are greater than .87 and deemed adequate. The multiple correlation for the fourth factor, although better than the corresponding result for the orthogonal factors reported previously, is marginal (ρ = .792). The $2\rho^2 - 1$ indeterminacy indices are also relatively high compared with the same values for the orthogonal factors, but the fourth factor still appears suspect. The validity coefficients for the refined factor scores compare favorably with the multiple correlations, with the greatest loss in validity (.734 compared with .792) occurring for the troubled fourth factor. As stated previously, only refined factor scores computed using Equation 5 ensure maximum validity. With respect to univocality, the scores for the oblique factors are again sufficient. As can be seen in Appendix B, the values in the UNIV and FACTCOR matrices are highly similar. The largest absolute difference is found for the fourth factor's correlation with the refined factor scores for the first factor (.408 compared with .556; absolute difference = .148). It is also worth mentioning that the correlations between the factors and the refined factor scores are generally lower than the correlations among the factors, as revealed by the lower values in the UNIV matrix compared with the FACTCOR matrix. A similar effect can be seen in terms of correlational accuracy, because the values in the SCORECOR matrix are less extreme than the values in the FACTCOR matrix. The correlational accuracy of the refined factor scores, however, is sufficient as the SCORECOR and FACTCOR matrices are highly similar. The largest difference is found for the correlation between the first and fourth factors (.374 compared with .556; absolute difference = .182). On the whole, the refined factor scores computed using Equation 10 for the oblique solution are superior to the scores from the orthogonal solution reported previously, but the scores for the fourth factor (freedom from distractibility) are still marginal at best.

Table 2
Pattern Coefficients and Factor Correlations for Promax-Rotated Factors From the WISC–III

	Pattern coefficients			
Subtest	1	2	3	4
Picture Completion	.108	**.616**[a]	.013	−.014
Information	**.739**[a]	−.038	−.012	.062
Coding	−.146	−.019	**.750**[a]	−.010
Similarities	**.777**[a]	.059	−.084	−.101
Picture Arrangement	.143	**.328**[a]	**.353**	−.049
Arithmetic	.115	−.112	.195	**.510**[a]
Block Design	−.078	**.764**[a]	.020	.231
Vocabulary	**.864**[a]	.018	−.013	−.055
Object Assembly	−.024	**.822**[a]	−.062	−.127
Comprehension	**.553**[a]	−.025	.023	.200
Symbol Search	.047	−.005	**.791**[a]	−.093
Digit Span	−.035	.046	−.168	**.572**[a]
Factor correlations				
1. Verbal	—			
2. Performance	.504	—		
3. Processing speed	.373	.508	—	
4. FD	.556	.321	.443	—

Note. Salient structure coefficients (≥.30) are in bold. FD = freedom-from-distractibility factor.
[a] Corresponding structure coefficients that were salient in the original Wechsler Intelligence Scale for Children—Third Edition (WISC–III) standardization sample.

Coarse Factor Scores

The method for computing the four index scores (coarse factor scores) for the WISC–III is based on the structure coefficients derived from exploratory factor analyses of the original standardization sample. For example, the Information, Similarities, Vocabulary, and Comprehension subtests were found to possess salient, positive structure coefficients (≥.30) on the first factor, labeled as *verbal comprehension,* for an orthogonally rotated four-factor solution. Scores on these subtests are therefore summed to create the coarse factor score for any given child who completes the WISC–III. Because the subtests are all on the same scale, it is not necessary to standardize the data before performing this computation. As stated previously, the structure coefficients from orthogonal factors for the present group of children diagnosed with learning disabilities (see Table 1) match the WISC–III standardization sample quite well. Consequently, the specific procedures for summing the subtest scores into coarse factor scores found in the WISC–III manual were initially followed here, and the resulting scores were assessed using a SAS IML evaluation program (see Footnote 1).

The novel output generated from this program is reported in Appendix C. As shown, the matrix of whole weights used to create the coarse factor scores is listed first. This matrix is essentially a simplified factor score coefficient matrix, and examination of its columns reveals how the items are summed to compute the scores for each factor. For example, scores on the third factor are computed by summing the Coding and Symbol Search subtests. Because all of the nonzero weights are positive, no subtests are subtracted in the present scoring scheme. The indeterminacy indices—ρ, ρ^2, and $2\rho^2 - 1$—are reported next to provide a basis of comparison. Validity coefficients for the coarse factor scores are then listed and compared with the multiple correlations for the factors. Because the coarse factor scores entail a loss of information, their validity coefficients will most assuredly be less than the ρs, and the goal is to achieve as little discrepancy between the values as possible. The results in Appendix C reveal that the coarse factor scores for the fourth factor are clearly inadequate (which is necessarily the case given the low multiple correlation), and the scores for the second factor compare least favorably with their respective multiple correlation (difference = .064). Univocality follows the validity information and reveals that the coarse factor scores are only modestly contaminated by other orthogonal factors in the analysis (maximum = .293). The correlational accuracy matrix, however, reveals moderate overlap among several coarse factor scores. The scores on the second factor, for instance, correlate .457 and .414 with scores on the first and third factors, respectively. Hence, even though the factors themselves are uncorrelated, several of the coarse factor scores are moderately oblique. In summary, the validity indices reveal that the coarse factor scores for the fourth factor (freedom from distractibility) are clearly inadequate. Validity coefficients for the first three factors appear adequate, but attempts to improve the coarse factor scores for the second factor (i.e., trying different sets of weights) might be fruitful. This latter suggestion is supported by several poor correlational accuracy values for the second factor.

The method for deriving the unit-weighted, coarse scoring scheme used in the WISC–III (i.e., determining how to weight the items on the basis of the structure coefficients) is common in practice and widely endorsed in the literature (Alwin, 1973; Gorsuch, 1997; ten Berge & Knol, 1985). Two recent Monte Carlo studies conducted by Grice and Harris (1998) and Grice (2001), however, suggest that the simplified scoring scheme should be based on the factor score coefficients instead. These coefficients are designed explicitly for determining the relative spacings of the individuals on the factors and are computed using least squares methods that can maximize the determinate nature of the estimated factor scores (viz., Equation 5), whereas the structure coefficients simply represent the correlations between the items and the factors. The relative values in these two matrices will typically correspond when ideal simple structure is obtained, which is certainly the exception rather than the rule in practice. It is therefore entirely possible that using the structure coefficients to determine item saliency for scoring one's factors may result in coarse factor scores that have poor validity, univocality, and correlational accuracy. This potentiality may explain the poor performance of the coarse scores for the second factor described previously relative to the first and third factors.

A SAS IML program was written that provides the means for selecting salient items from the factor score coefficients (see Footnote 1). The output generated from this program for the current WISC–III data is reported in Appendix D. As shown, the factor score coefficients for the first factor are initially ranked in descending order. These ranked values are then plotted in a two-dimensional graph with the abscissa comprising the ranks and the ordinate comprising the factor score coefficients. As can be seen in the top panel of Figure 1, the resulting graph is similar to a scree plot. A clear break in the graph is present that sepa-

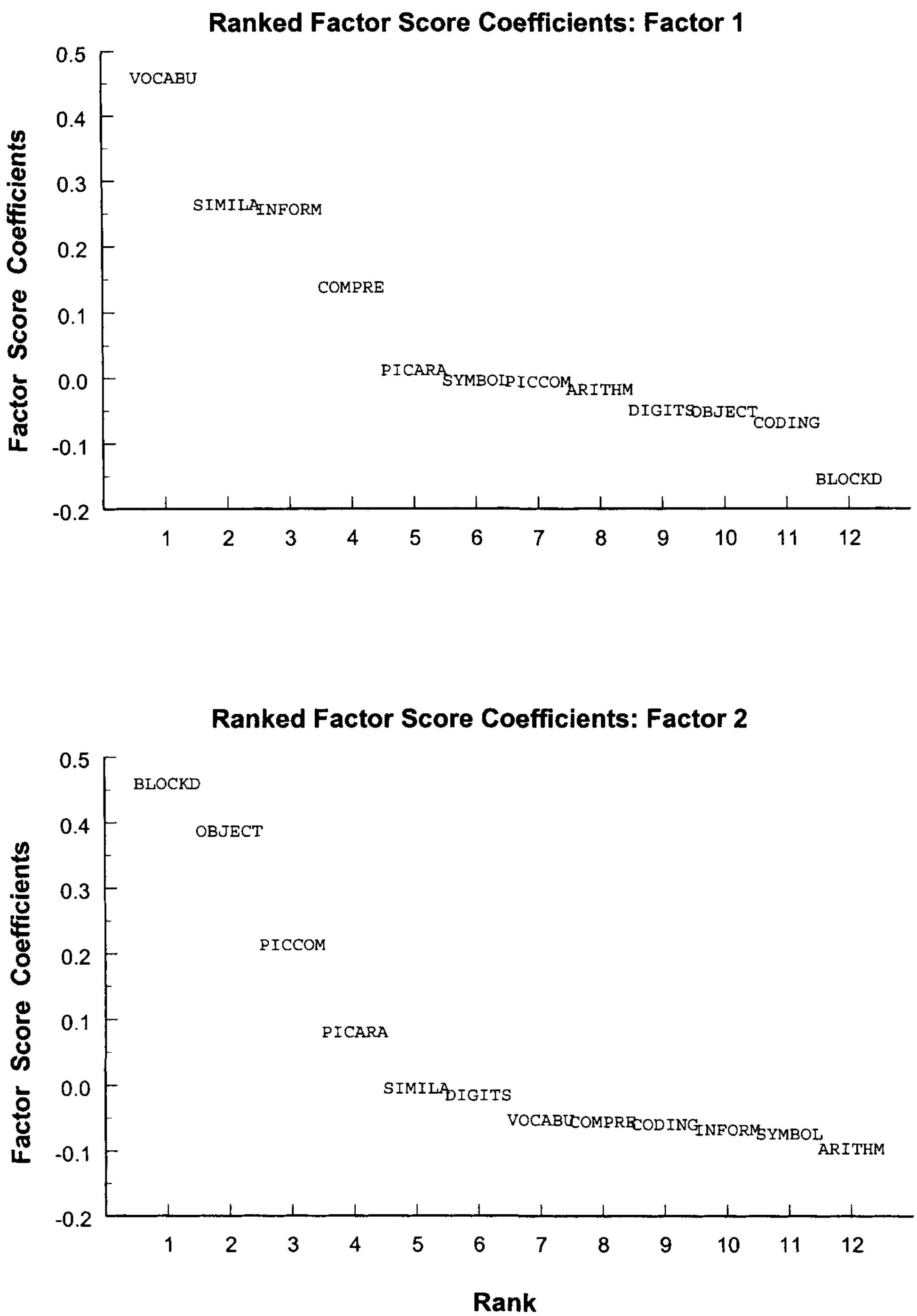

Figure 1. Plots of ranked factor score coefficients (Equation 5) for first and second orthogonal factors. Abbreviations for subtests are as follows: VOCABU = Vocabulary; SIMILA = Similarities; INFORM = Information; COMPRE = Comprehension; PICARA = Picture Arrangement; SYMBOL = Symbol Search; PICCOM = Picture Completion; ARITHM = Arithmetic; DIGITS = Digit Span; OBJECT = Object Assembly; CODING = Coding; BLOCKD = Block Design.

rates Vocabulary, Similarities, Information, and Comprehension from the remaining subtests on the first factor. Moreover, a second break in the graph reveals a particularly extreme negative factor score coefficient for Block Design. The coarse factor scores are hence computed by summing Vocabulary, Similarities, Information, and Comprehension, and subtracting Block Design. This method differs from the original scoring procedure reported in the WISC–III manual by including the Block Design subtest. The graph of ranked factor score coefficients for the second factor is presented in the bottom panel of Figure 1 and reveals that the coarse factor scores should be computed by summing Block Design, Object Assembly, and Picture Completion. Unlike the original scoring procedure, the Picture Arrangement subtest is not included for the second factor. It should also be noted that although Picture Arrangement shows slight separation from Similarities in the graph, its absolute value is similar to Arithmetic, which shows virtually no separation from Symbol Search and a number of other subtests. Including Picture Arrangement would therefore justify including these other negatively weighted subtests as well. As with interpreting a traditional scree plot, a degree of subjectivity is obviously involved. Unlike a scree test, however, the researcher can use the evaluation program described previously to assess the adequacy of the coarse factor scores and make adjustments if necessary. According to the scree-type plot for the third factor (not shown), coarse factor scores are computed from the Coding and Symbol Search subtests, which matches the original scoring scheme. Coarse factor scores for the fourth factor are computed by summing the Arithmetic, Block Design, Comprehension, and Digit Span subtests, and subtracting the Object Assembly subtest. In the original scoring scheme, only the Arithmetic and Digit Span subtests are summed. It is interesting to note that the troubled fourth factor produced a plot (not shown) that did not reveal clear separation among the factor score coefficients, which more than doubled the number of subtests used to estimate the factor scores.

Standard errors and t values for the factor score coefficients are also listed in the output (see Appendix D).[3] A common criticism against least squares regression weights (e.g., see Gorsuch, 1997) is the potential for wildly different standard errors that would obscure the interpretation of their relative magnitudes. Basing the coarse factor scoring scheme on such coefficients would therefore be risky business. The results for this particular data set, however, reveal that the standard errors within each factor are fairly homogeneous, and the most extreme t values correspond to the most extreme factor score coefficients. For example, the Picture Completion (5.57), Block Design (10.53), and Object Assembly (9.84) subtests clearly have the most extreme t values (listed in parentheses) for the second factor. The t value for Picture Arrangement (2.23) on the second factor was not unusual, justifying its exclusion from the coarse factor scores. Hence, the standard errors and t values support the weighting schemes for computing the factor scores derived from the graphs in Figure 1. The output concludes with the total contribution of each item (WISC-III subtest) to the squared multiple correlation, ρ^2, for each factor.[4] The factor score coefficients represent the direct contribution of each item to ρ^2, whereas the numbers in this final matrix include both direct and indirect effects. These values can consequently be examined to determine whether additional items need to be included in the computation of the coarse factor scores. In this data set, the largest values correspond to the largest factor score coefficients for each factor, and hence no additional items are deemed necessary.

These new coarse factor scores based on the factor score coefficients for the orthogonal four-factor solution were evaluated using the program described previously. The results are reported in Appendix E and can be compared with the original output based on the structure coefficients in Appendix C. This comparison reveals that the validity coefficients for the first and second factors were improved by the adjustments. The first factor increased from .852 to .864 (ρ = .886), and the second factor improved from .801 to .831 (ρ = .865). The new coarse factor scores for the beleaguered fourth factor, however, showed little improvement (.627 to .629), although the changes were

---

[3] The standard error for the factor score coefficient of Item i and Factor j was computed as follows:

$$se_{ij} = (r_{ii}(1 - R_j^2)/N - k - 1)^{1/2}, \qquad (14)$$

where r_{ii} is the ith main diagonal entry of the inverse of the item correlation matrix, R_j^2 is the squared validity coefficient for Factor j (not ρ^2), N is the total number of observations, and k is the number of items (see Harris, 1985b, p. 65).

[4] The total direct and indirect contributions to the squared multiple correlations for the factors were computed using element-wise multiplication of the structure and factor score coefficient matrices.

in vain given the already inadequate multiple correlation (ρ = .683). The validity coefficient for the third factor did not change because the scoring schemes based on the structure and factor score coefficients were equivalent in this case. Comparison of the univocality matrices in Appendices E and C also reveal improvement, as values in the former Appendix are generally closer to the comparison matrix of zeros. In other words, the new coarse factor scores are less contaminated by other factors in the same analysis. The exception to this general improvement involves the fourth factor. The new coarse factor scores show more contamination from the fourth factor than from the original scores. This result is not completely surprising, however, given that two of the three items added to this factor are shared by other factors in the analysis. Finally, a comparison of the correlational accuracy matrices across Appendices E and C again shows improvement for the correlations involving only the first three factors and a decrement in performance for the fourth factor. The correlations among the new coarse factor scores for the first three factors resembles the comparison matrix (in this case, an identity matrix) more closely than the original coarse factor scores.

In summary, the coarse factor scores based on the factor score coefficients revealed superior levels of validity, univocality, and correlational accuracy for the first three factors compared with the original scores based on the structure coefficients. The coarse factor scores for the fourth factor showed a slight improvement in validity but decrements in univocality and orthogonality. It should be kept in mind, however, that the fourth factor was retained in all of the analyses to provide a thread of consistency throughout the examples. It was therefore retained solely for pedagogical reasons. Given all of the results stated previously for both the refined and coarse factor scoring methods, a more appropriate strategy may have been to reconduct the factor analyses and extract only three factors, or retain four factors and apply an oblique transformation. Even with oblique factors, the fourth factor was marginal, as revealed by its indeterminacy indices, and the best alternative may consequently be its exclusion.

Discussion

Factor scores computed from the common factor model are indeterminate in nature. For any single common factor, an infinite number of sets of scores can be derived that are equally consistent with the factor loadings. Under particular circumstances, competing sets of scores for the same factor can actually be orthogonal or negatively correlated, thus yielding completely different rankings of the individuals. This inherent indeterminacy creates both conceptual (Steiger & Schönemann, 1978) and empirical (Schönemann & Steiger, 1978; Steiger, 1979) difficulties for the factor analyst that should not be ignored. Indeterminacy has also led to a large number of methods for estimating factor scores, some of which appear to be defective. In other words, not all means of calculating factor scores are adequate; hence, even if highly determinate factor scores can be created for a given set of results, the researcher may still choose a method that is severely flawed (e.g., summing standardized scores on the basis of salient structure coefficients). The foregoing procedures and computer programs were specifically designed to provide researchers with the knowledge and tools necessary for effectively addressing, rather than neglecting, the issues surrounding the computation of factor score estimates. The structure and style by which this information was presented was intentionally pedagogical, introducing and demonstrating procedures and criteria for evaluating factor scores that are unknown to a vast majority of researchers and consumers of factor analytic technology. A number of different questionnaires, inventories, or ability tests other than the WISC–III (Wechsler, 1991) could certainly have been chosen for this purpose, as the evaluative computer programs were designed for maximum flexibility. Large numbers of items and factors, different factor extraction algorithms, and orthogonal or oblique factor transformations can all be managed by the computer programs reported in this article.

Choosing the WISC–III for the previously described demonstrations, however, proved fortuitous for a number of reasons. First, it is a widely used assessment tool in psychology and education and is therefore familiar to a large and diverse audience of readers. Second, it possesses a number of manageable items (subtests) that could be displayed and discussed efficiently. Finally, and most importantly, a good deal of controversy surrounds the validity of the four-factor solution reported in the test manual. Some researchers have argued that a two- or three-factor solution provides superior fit to the extant data, directing most of their criticism toward the freedom from distractibility factor (Kamphaus, Benson, Hutchinson, & Platt, 1994; Kush & Watkins, 1994; Sattler, 1992). The results reported previously may therefore be interpreted in light of a genuine measurement controversy, further exhibiting the importance of factor score indeterminacy. Viewed in this light, the freedom from distractibility factor was clearly inadequate. Its multiple correlation, ρ, was less than .707, and the minimum correlation among competing sets of factor scores, $2\rho^2 - 1$, was $-.066$. Because the multiple correlation was low, the validity coefficient for the coarse factor scores (index scores) was necessarily low as well. Applying an oblique transformation to the extracted factors improved the multiple correlation ($\rho = .792$), but it was still inadequate.

In a recent article, Fabrigar et al. (1999) offered a number of recommendations to aid researchers with the various decisions that must be made when conducting an exploratory factor analysis (EFA). Sadly, they did not recommend evaluating factor score indeterminacy or the adequacy of one's computed factor score estimates. The example reported previously for the WISC–III, however, demonstrates clearly the necessity for incorporating such evaluations into any EFA. Even if the researcher will not finally compute factor scores for his or her data, the maximum proportion of determinacy for each factor, ρ^2, should at least be reported along with the eigenvalues, rotation method, and structure or pattern coefficients that are routinely published. SAS reports ρ if the factor score coefficients are requested, and SPSS reports ρ for orthogonal factors if factor scores from Equation 5 (labeled as *regression* factor scores in the program) are requested. None of the commercial programs, however, provide the means for computing factor scores from Equations 5, 8, 9, and 10. For instance, SAS allows one to compute refined factor scores using Equation 5, and SPSS provides options for computing refined factor scores from Equations 5, 7, and 9. The necessary tools for evaluating refined as well as coarse factor scores in terms of validity, univocality, and correlational accuracy are not available. Moreover, although most major programs provide the structure, pattern, and factor score coefficients from Equation 5, they do not provide the graphical displays, t values, and standard errors for the factor score coefficients, $\mathbf{W}_{kf}$, nor do they provide the combined direct and indirect contributions of each item to ρ^2. The programs reported in this article should therefore prove vital to EFA researchers for constructing and evaluating factor score estimates as well as the degree of indeterminacy in their analyses.

If factor scores are computed for a particular set of results, or a method of computing factor scores is provided for future researchers using the same instrument (e.g., as in scale construction), then information regarding the validity, univocality, and correlational accuracy of the factor score estimates should be reported. Validity represents the extent to which the factor score estimates correlate with their respective factors in the sample, and values approaching 1.00 are desirable. Low validity will likely reduce the statistical power of subsequent decisions based on the factor scores because a large proportion of their variance may be random in nature. Validity must also be judged, however, within the context of univocality and correlational accuracy because a small decrement in the former index may correspond with substantial losses in the latter indices. As a consequence, the factor score estimates may become overly saturated with variation from other factors and factor score estimates in the same analysis. Such an outcome could confound the process of interpreting the relationships between factor score estimates for a particular factor and external criteria. For example, consider a researcher who extracts two orthogonal factors, labels them as *depression* and *hostility,* uses a scoring method that produces highly correlated factor score estimates, and then fails to evaluate the scores for validity, univocality, and correlational accuracy. Subsequent analyses will all be interpreted in light of orthogonal factors, even though the factor score estimates themselves are not independent. All three evaluative criteria should therefore be carefully examined and reported as standard output of any EFA.

When factor scores are computed and evaluated, a choice between refined or coarse factor score estimates must be made. Refined factor scores will typically have superior levels of validity compared with their coarse counterparts for a given sample, and particular constraints, such as orthogonality for uncorrelated factors, can be placed on these scores. Consequently, if one wishes to employ a complex weighting scheme that uses all of the items, the refined factor scores would be suitable. The researcher will have to decide, however, what constraints are to be employed. Table 3 serves as a quick summary of the different properties of the factor score estimates generated from Equations 5, 8, 9, and 10. As shown, each method fails to meet all three criteria, and the methods in Equations 9 and 10 only ensure univocality when the factors are orthogonal. Some authors have additionally argued that the purposes for which factor scores are to be used should be included in one's choice. For example, Tucker (1971) showed analytically that the

Table 3
Properties of Different Methods for Estimating Refined Factor Scores

Method (Equation no.)	Maximizes validity	Univocal for orthogonal factors	Correlation preserving
5	Yes	No	No
8	No	No	Yes
9	No	Yes	No
10	No	Yes	No

method in Equation 5 should be preferred when the estimated factor scores are to be entered into subsequent regression analyses. This suggestion, however, was not supported when tested in an empirical study (Lastovicka & Thamodaran, 1991). Skrondal and Laake (in press) further argued that if regression analyses are used in lieu of structural equation models with latent variables, a modified type of blockwise factor score regression should be employed. Specifically, factor scores for latent explanatory (predictor) variables should be computed from Equation 5, whereas factor scores for latent response (criterion) variables should be computed from Equations 9 or 10. Future research should add to this store of knowledge in helping to direct researchers in their choice of the refined factor scoring methods.

Coarse factor scores are extremely popular in the literature and go by different names, such as scale scores, index scores, cluster scores, and sum scores. Although some authors may object to naming such quantities *factor scores* (e.g., see Thorndike, 1978, p. 321), the values are intended to provide the rankings of the individuals on the identified factors in the analysis. As such, they are justifiably called factor scores (estimates) and are generally believed to be simple, effective, and stable alternatives to the refined methods. They are also considered to be more consistent with the process of factor interpretation, which is based on an examination of subsets of items in the analysis. As discussed previously, these beliefs may be well founded as long as the researcher is willing (a) to avoid using the structure coefficients as the basis for selecting salient items to include in the coarse factor scores and (b) to examine these scores using the procedures and programs provided in this article. As shown previously, the coarse factor scores for the second factor from the WISC–III were greatly improved when salient items were selected on the basis of factor score coefficients, $\mathbf{W}_{kf}$. These coarse factor scores were more valid and showed superior levels of univocality and correlational accuracy compared with similar scores based on the structure coefficients, $\mathbf{S}_{kf}$. A number of criteria can be used when selecting items from the $\mathbf{W}_{kf}$ matrix. Scree-type plots of the factor score coefficients for each factor can be examined to determine which items yield the most extreme values, as well as their signs. The *t* values and standard errors for the factor score coefficients can also be examined, and the former values can be evaluated for statistical significance. The total contribution of each item to ρ^2 is also provided and can be incorporated in the decision process. Certainly, all of this information can be used to select the items that will finally be summed into the coarse factor scores, and several versions of estimated factor scores can be attempted and evaluated in the early stages of an exploratory factor analysis. It should also be mentioned that differential weighting strategies can be used. For instance, items with extreme factor score coefficients may be weighted by a factor of two. In essence, a slightly more complex weighting scheme (−2, −1, 0, 1, 2) is used rather than the more common, simple scheme (−1, 0, 1). Rozeboom (1979) argued that a complex strategy is most likely to be fruitful when the items are heterogeneous in nature (i.e., they tend to load on more than one factor) and few in number. Regardless, the final coarse factor scores should be assessed with respect to their validity, univocality, and correlational accuracy.

Coarse factor scores are often computed from the results of principal-components and image analyses. Even though refined factor scores from such analyses are determinate in nature, the coarse scores may still suffer from poor validity, univocality, and correlational accuracy. As mentioned previously, inadequate component or image scores are likely to be derived when one uses the structure coefficients rather than the factor score coefficients to select salient items to include in the coarse scores (see Halperin, 1976). One exception to this prescription, however, is the case of an unrotated principal-components analysis in which the factor score coefficients are a simple function of the structure coefficients and eigenvalues. Namely, $\mathbf{W}_{kf}$ is computed by dividing the elements in $\mathbf{S}_{kf}$ by their respective eigenvalues (Kaiser, 1962). The relative magnitudes of the two matrices will therefore be equivalent, and the same items to include in the coarse factor scores will be selected. Once the principal components are transformed orthogonally or obliquely, discrepancies in the relative magnitudes of the $\mathbf{W}_{kf}$ and $\mathbf{S}_{kf}$ matrices can emerge. Consequently, radically different coarse factor scores can be derived (Harris, 1985a, 1985b). Because the factor score coefficients are designed specifically for scoring the components, and the structure coefficients represent the correlations between the components and the items, the former values will yield coarse factor scores that have superior levels of validity, univocality, and correlational accuracy compared with the latter.

In conclusion, many early psychologists believed that the common factor model and exploratory factor analytic techniques would help shape our understanding of human abilities and individual differences. Support for these beliefs can indeed be seen in modern theories of intelligence, personality, and self-esteem, to name only a few domains or constructs that have been modeled with factor analysis. Critics have argued, however, that factor score indeterminacy seriously hinders the effectiveness of the common factor model and may in fact render it misleading (e.g., see Schönemann, 1997; Schönemann & Wang, 1972; Steiger, 1996b) or even meaningless (see Schönemann & Steiger, 1978). The question posed by most critics is: Of what scientific value is a common factor if the researcher cannot score the individuals in an unambiguous fashion along the identified dimension? For instance, imagine a measure of temperature that has a high degree of factor score indeterminacy. Two researchers could use such a measure to derive completely different rankings, both equally valid, of the temperatures in the rooms of their buildings. It is difficult to imagine that such a measure would be deemed as adequate or would propel the science of temperature forward. Yet, this is exactly the issue facing psychologists who use exploratory factor analysis techniques and fail to evaluate factor score indeterminacy or their estimated factor scores. Thurstone's (1935) attempt to separate the factor identification and scoring processes may be at the root of this dilemma, but as Steiger (1996a) recently wrote, "If we wish to continue using models with more latent than observed variables, we need to discuss and develop methods for the measurement and evaluation of factor indeterminacy, so that the problem is properly controlled" (p. 619). The present article provides researchers and practitioners with these much needed methods.

References

Alwin, D. F. (1973). The use of factor analysis in the construction of linear composites in social research. *Sociological Methods & Research, 2,* 191–213.

Anderson, R. D., & Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the Third Berkeley Symposium of Mathematical Statistics and Probability, 5,* 111–150.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology, 28,* 97–104.

Burt, C. (1940). *The factors of mind.* London: London University Press.

Cattell, R. B. (1952). *Factor analysis: An introduction and manual for the psychologist and social scientist.* New York: Harper.

Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology, 46,* 648–659.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO–PI–R) and NEO Five-Factor Inventory (NEO–FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

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

Garnett, J. C. M. (1919a). General ability, cleverness and purpose. *British Journal of Psychology, 9,* 345–366.

Garnett, J. C. M. (1919b). On certain independent factors in mental measurements. *Proceedings of the Royal Society, Series A, 96,* 91–111.

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

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment, 68,* 532–560.

Green, B. F. (1976). On the factor score controversy. *Psychometrika, 41,* 263–266.

Grice, J. W. (2001). A comparison of factor scores under conditions of factor obliquity. *Psychological Methods, 6,* 67–83.

Grice, J. W., & Harris, R. J. (1998). A comparison of regression and loading weights for the computation of factor scores. *Multivariate Behavioral Research, 33,* 221–247.

Grice, J. W., Krohn, E. J., & Logerquist, S. (1999). Cross-validation of the WISC–III factor structure in two samples of children with learning disabilities. *Journal of Psychoeducational Assessment, 17,* 236–248.

Guttman, L. (1955). The determinacy of factor score matrices with applications for five other problems of common factor theory. *British Journal of Statistical Psychology, 8,* 65–82.

Halperin, S. (1976). The incorrect measurement of components. *Educational and Psychological Measurement, 36,* 347–353.

Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: University of Chicago Press.

Harris, R. J. (1985a, November). *Further features of a factor fable.* Paper presented at the annual meeting of the Society for Multivariate Psychology, Berkeley, CA.

Harris, R. J. (1985b). *A primer of multivariate statistics* (2nd ed.). Orlando, FL: Academic Press.

Heermann, E. F. (1963). Univocal or orthogonal estimators of orthogonal factors. *Psychometrika, 28,* 161–172.

Holzinger, K. J., & Harman, H. H. (1941). *Factor analysis.* Chicago: University of Chicago Press.

Horn, J. L. (1965). An empirical comparison of methods for estimating factor scores. *Educational and Psychological Measurement, 25,* 313–322.

Kaiser, H. F. (1962). Formulas for component scores. *Psychometrika, 27,* 83–87.

Kamphaus, R. W., Benson, J., Hutchinson, S., & Platt, L. O. (1994). Identification of factor models for the WISC–III. *Educational and Psychological Measurement, 54,* 174–186.

Kestelman, H. (1952). The fundamental equations of factor analysis. *British Journal of Psychology, 5,* 1–6.

Kush, J. C., & Watkins, M. W. (1994). *Factor structure of the WISC–III for Mexican-American, learning disabled students* (Clearing House No. TM 022 083). East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED 374 154)

Lastovicka, J., & Thamodaran, K. (1991). Common factor score estimates in multiple regression problems. *Journal of Marketing Research, 28,* 105–112.

Ledermann, W. (1939). On a shortened method of estimation of mental factors by regression. *Psychometrika, 4,* 109–116.

Lovie, P., & Lovie, A. D. (1995). The cold equations: Spearman and Wilson on factor indeterminacy. *British Journal of Mathematical and Statistical Psychology, 48,* 237–253.

Maraun, M. D. (1996). Metaphor taken as math: Indeterminacy in the factor analysis model. *Multivariate Behavioral Research, 31,* 517–538.

McDonald, R. P. (1974). The measurement of factor indeterminacy. *Psychometrika, 39,* 203–222.

McDonald, R. P. (1981). Constrained least squares estimators of oblique common factors. *Psychometrika, 46,* 337–341.

McDonald, R. P., & Burr, E. J. (1967). A comparison of four methods of constructing factor scores. *Psychometrika, 32,* 381–401.

Meyer, E. P. (1973). On the relationship between ratio of number of variables to number of factors and factorial determinacy. *Psychometrika, 38,* 375-378.

Morris, J. D. (1980). The predictive accuracy of full-rank variables vs. various types of factor scores: Implications for test validation. *Educational and Psychological Measurement, 40,* 389–396.

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

Mulaik, S. A. (1976). Comments on "The measurement of factorial indeterminacy." *Psychometrika, 41,* 249–262.

Rozeboom, W. W. (1979). Sensitivity of a linear composite of predictor items to differential item weighting. *Psychometrika, 44,* 289–296.

Sattler, J. (1992). *Assessment of children: WISC–III and WPPSI–R supplement.* San Diego, CA: Author.

Schönemann, P. H. (1971). The minimum average correlation between equivalent sets of uncorrelated factors. *Psychometrika, 36,* 21–30.

Schönemann, P. H. (1997). On models and muddles of heritability. *Genetica, 99,* 97–108.

Schönemann, P. H., & Steiger, I. H. (1978). On the validity of indeterminate factor scores. *Bulletin of the Psychonomic Society, 12,* 287–290.

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

Skrondal, A., & Laake, P. (in press). Regression among factor scores. *Psychometrika.*

Spearman, C. (1904). 'General intelligence' objectively determined and measured. *American Journal of Psychology, 5,* 201–293.

Spearman, C. (1927). *The abilities of man, their nature and measurement.* New York: Macmillan.

Spearman, C. (1932). *The abilities of man (2nd Impression).* London: Macmillan.

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

Steiger, J. H. (1996a). Coming full circle in the history of factor indeterminacy. *Multivariate Behavioral Research, 31,* 617–630.

Steiger, J. H. (1996b). Dispelling some myths about factor indeterminacy. *Multivariate Behavioral Research, 31,* 539–550.

Steiger, J. H., & Schönemann, P. H. (1978). A history of factor indeterminacy. In S. Shye (Ed.), *Theory construction and data analysis* (pp. 136–178). Chicago: University of Chicago Press.

ten Berge, J., & Knol, D. (1985). Scale construction on the

basis of components analysis: A comparison of three strategies. *Multivariate Behavioral Research, 20,* 45–55.
ten Berge, J., Krijnen, W., Wansbeek, T., & Shapiro, A. (1999). Some new results on correlation-preserving factor scores prediction methods. *Linear Algebra and its Applications, 289,* 311–318.
Thomson, G. (1939). *The factorial analysis of human ability.* London: University of London Press.
Thorndike, R. M. (1978). *Correlational procedures for research.* New York: Gardner Press.
Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review, 38,* 406–427.
Thurstone, L. L. (1934). The vectors of mind. *Psychological Review, 41,* 1–32.
Thurstone, L. L. (1935). *The vectors of mind.* Chicago: University of Chicago Press.
Thurstone, L. L. (1947). *Multiple-factor analysis.* Chicago: University of Chicago Press.
Tucker, L. R. (1971). Relations of factor score estimates to their use. *Psychometrika, 36,* 427–436.
Wackwitz, J. H., & Horn, J. L. (1971). On obtaining the best estimates of factor scores within an ideal simple structure. *Multivariate Behavioral Research, 6,* 389–408.
Wechsler, D. (1991). *WISC–III manual.* San Antonio, TX: Psychological Corporation.
Wilson, E. B. (1928). On hierarchical correlational systems. *Proceedings of the National Academy of Sciences, 14,* 283–291.

Appendix A

Output for Four Orthogonal Factors From Refined Factor Score Evaluation Program (Equation 5)

**** BEGIN OUTPUT FROM PROC IML ****

Factor Score Coefficients for Items/Factors

ITEM	FSCOEF			
PICCOM	0.000	0.219	−0.021	−0.037
INFORM	0.262	−0.062	−0.018	0.053
CODING	−0.063	−0.054	0.382	−0.008
SIMILA	0.269	0.001	−0.062	−0.107
PICARA	0.018	0.086	0.144	−0.051
ARITHM	−0.012	−0.091	0.091	0.382
BLOCKD	−0.149	0.465	−0.037	0.275
VOCABU	0.464	−0.047	−0.036	−0.100
OBJECT	−0.046	0.393	−0.080	−0.167
COMPRE	0.144	−0.050	0.005	0.156
SYMBOL	0.002	−0.068	0.502	−0.102
DIGITS	−0.044	−0.010	−0.059	0.334

Indeterminacy/Determinacy Indices
(Multiple R, R-Squared, and Minimum Correlation)

FACTOR	MULTR	RSQR	MINCOR
1	0.886	0.786	0.572
2	0.865	0.748	0.495
3	0.829	0.687	0.373
4	0.683	0.467	−0.066

Validity Coefficients

FACTOR	VALID		MULTR
1	0.886	compare to MULTR →	0.886
2	0.865		0.865
3	0.829		0.829
4	0.683		0.683

Univocality
(Rows = Factor Scores/Columns = Factors)

UNIV					FACTCOR			
—	0.082	0.017	0.170	compare to FACTCOR →	—	0.000	0.000	0.000
0.080	—	0.110	0.042		0.000	—	0.000	0.000
0.016	0.105	—	0.091		0.000	0.000	—	0.000
0.131	0.033	0.075	—		0.000	0.000	0.000	—

Correlational Accuracy

SCORECOR

1.000	—	—	—
0.093	1.000	—	—
0.019	0.127	1.000	—
0.191	0.048	0.109	1.000

compare to FACTCOR →

FACTCOR

1.000	—	—	—
0.000	1.000	—	—
0.000	0.000	1.000	—
0.000	0.000	0.000	1.000

Appendix B

Output for Four Oblique Factors From Refined Factor Score Evaluation Program (Equation 10)

**** BEGIN OUTPUT FROM PROC IML ****

Factor Score Coefficients for Items/Factors

ITEM	FSCOEF			
PICCOM	0.042	0.348	0.001	−0.023
INFORM	0.324	−0.029	0.010	0.068
CODING	−0.049	−0.021	0.539	0.041
SIMILA	0.342	0.027	−0.053	−0.165
PICARA	0.067	0.179	0.250	−0.048
ARITHM	0.043	−0.069	0.180	0.718
BLOCKD	−0.048	0.433	0.017	0.322
VOCABU	0.381	0.001	0.003	−0.098
OBJECT	−0.018	0.467	−0.066	−0.182
COMPRE	0.239	−0.021	0.042	0.266
SYMBOL	0.039	−0.015	0.567	−0.076
DIGITS	−0.033	0.027	−0.084	0.783

Indeterminacy/Determinacy Indices

(Multiple R, R-Squared, and Minimum Correlation)

FACTOR	MULTR	RSQR	MINCOR
1	0.929	0.862	0.724
2	0.911	0.830	0.660
3	0.876	0.768	0.536
4	0.792	0.627	0.253

Validity Coefficients

FACTOR	VALID		MULTR
1	0.920	compare to MULTR →	0.929
2	0.898		0.911
3	0.862		0.876
4	0.734		0.792

Univocality

(Rows = Factor Scores/Columns = Factors)

UNIV

—	0.453	0.322	0.408
0.464	—	0.438	0.236
0.343	0.456	—	0.325
0.511	0.288	0.382	—

compare to FACTCOR →

FACTCOR

—	0.505	0.373	0.556
0.505	—	0.508	0.321
0.373	0.508	—	0.443
0.556	0.321	0.443	—

Correlational Accuracy

SCORECOR

1.000	—	—	—
0.418	1.000	—	—
0.304	0.390	1.000	—
0.374	0.196	0.299	1.000

compare to FACTCOR →

FACTCOR

1.000	—	—	—
0.505	1.000	—	—
0.373	0.508	1.000	—
0.556	0.321	0.443	1.000

Appendix C

Output for Four Orthogonal Factors From Coarse Factor Score Evaluation Program

**** BEGIN OUTPUT FROM PROC IML ****

Simplified Weights for Items/Factors

ITEM	FSCOEF			
PICCOM	0	1	0	0
INFORM	1	0	0	0
CODING	0	0	1	0
SIMILA	1	0	0	0
PICARA	0	1	0	0
ARITHM	0	0	0	1
BLOCKD	0	1	0	0
VOCABU	1	0	0	0
OBJECT	0	1	0	0
COMPRE	1	0	0	0
SYMBOL	0	0	1	0
DIGITS	0	0	0	1

Indeterminacy/Determinacy Indices
(Multiple R, R-Squared, and Minimum Correlation)

FACTOR	MULTR	RSQR	MINCOR
1	0.886	0.786	0.572
2	0.865	0.748	0.495
3	0.829	0.687	0.373
4	0.683	0.467	–0.066

Validity Coefficients

FACTOR	VALID		MULTR
1	0.852	compare to MULTR →	0.886
2	0.801		0.865
3	0.805		0.829
4	0.627		0.683

Univocality
(Rows = Factor Scores/Columns = Factors)

UNIV					FACTCOR			
—	0.263	0.070	0.254	compare to FACTCOR →	—	0.000	0.000	0.000
0.220	—	0.192	0.056		0.000	—	0.000	0.000
0.102	0.293	—	0.170		0.000	0.000	—	0.000
0.224	0.111	0.070	—		0.000	0.000	0.000	—

Correlational Accuracy

SCORECOR					FACTCOR			
1.000	—	—	—	compare to FACTCOR →	1.000	—	—	—
0.457	1.000	—	—		0.000	1.000	—	—
0.200	0.414	1.000	—		0.000	0.000	1.000	—
0.388	0.227	0.211	1.000		0.000	0.000	0.000	1.000

Appendix D

Output for Four Orthogonal Factors From Program for Creating Coarse Factor Scores (Equation 5)

**** BEGIN OUTPUT FROM PROC IML ****

Ranked Factor Score Coefficients: Factors 1

ITEM	RANKCOEF
VOCABU	0.464
SIMILA	
INFORM	0.262
COMPRE	0.144
PICARA	0.018
SYMBOL	0.002
PICCOM	0.000
ARITHM	−0.012
DIGITS	−0.044
OBJECT	−0.046
CODING	−0.063
BLOCKD	−0.149

Standard Errors and t-values for Factor Score Coefficients

ITEM	ERRORS				T_VALUES			
PICCOM	0.036	0.039	0.044	0.057	0.004	5.574	−0.484	−0.650
INFORM	0.039	0.043	0.048	0.062	6.662	−1.447	−0.383	0.850
CODING	0.034	0.037	0.042	0.054	−1.828	−1.452	9.178	−0.154
SIMILA	0.038	0.041	0.046	0.060	7.117	0.032	−1.365	−1.792
PICARA	0.036	0.039	0.043	0.056	0.508	2.225	3.323	−0.911
ARITHM	0.033	0.036	0.040	0.053	−0.351	−2.514	2.270	7.271
BLOCKD	0.041	0.044	0.049	0.064	−3.658	10.535	−0.743	4.287
VOCABU	0.042	0.046	0.051	0.067	10.952	−1.021	−0.697	−1.502
OBJECT	0.037	0.040	0.044	0.058	−1.243	9.843	−1.810	−2.888
COMPRE	0.036	0.040	0.044	0.058	3.951	−1.269	0.115	2.715
SYMBOL	0.036	0.039	0.044	0.057	0.063	−1.743	11.528	−1.794
DIGITS	0.030	0.033	0.037	0.048	−1.452	−0.292	−1.605	6.962

Total Item Contribution to Squared Multiple Correlation

ITEM	CONTRIB			
PICCOM	0.000	0.133	−0.003	−0.003
INFORM	0.184	−0.009	−0.002	0.011
CODING	0.001	−0.007	0.259	−0.001
SIMILA	0.187	0.000	−0.001	−0.008
PICARA	0.005	0.037	0.060	−0.004
ARITHM	−0.003	−0.003	0.027	0.195
BLOCKD	−0.028	0.340	−0.009	0.073
VOCABU	0.368	−0.010	−0.003	−0.014
OBJECT	−0.006	0.289	−0.008	0.010
COMPRE	0.083	−0.007	0.001	0.048
SYMBOL	0.000−	0.013	0.365	−0.005
DIGITS	−0.006	−0.001	0.001	−0.164
—	—	—	—	—
TOTALS	0.786	0.748	0.687	0.467

Appendix E

Output for Four Modified Orthogonal Factors From Coarse Factor Score Evaluation Program

**** BEGIN OUTPUT FROM PROC IML ****

Simplified Weights for Items/Factors

ITEM	FSCOEF			
PICCOM	0	1	0	0
INFORM	1	0	0	0
CODING	0	0	1	0
SIMILA	1	0	0	0

PICARA	0	0	0	0
ARITHM	0	0	0	1
BLOCKD	−1	1	0	1
VOCABU	1	0	0	0
OBJECT	0	1	0	−1
COMPRE	1	0	0	1
SYMBOL	0	0	1	0
DIGITS	0	0	0	1

Indeterminacy/Determinacy Indices
(Multiple R, R-Squared, and Minimum Correlation)

FACTOR	MULTR	RSQR	MINCOR
1	0.886	0.786	0.572
2	0.865	0.748	0.495
3	0.829	0.687	0.373
4	0.683	0.467	−0.066

Validity Coefficients

FACTOR	VALID		MULTR
1	0.864	compare to MULTR →	0.886
2	0.831		0.865
3	0.805		0.829
4	0.629		0.683

Univocality
(Rows = Factor Scores/Columns = Factors)

UNIV					FACTCOR			
—	0.226	0.070	0.404	compare to FACTCOR →	—	0.000	0.000	0.000
−0.005	—	0.192	0.087		0.000	—	0.000	0.000
0.032	0.200	—	0.211		0.000	0.000	—	0.000
0.155	0.111	0.070	—		0.000	0.000	0.000	—

Correlational Accuracy

SCORECOR					FACTCOR			
1.000	—	—	—	compare to FACTCOR →	1.000	—	—	—
0.176	1.000	—	—		0.000	1.000	—	—
0.096	0.345	1.000	—		0.000	0.000	1.000	—
0.484	0.252	0.260	1.000		0.000	0.000	0.000	1.000

Received March 5, 2001
Revision received August 10, 2001
Accepted August 15, 2001 ■