Psicológica (2016), 37, 235-247.

A note on improving EAP trait estimation in oblique factor-analytic and item response theory models

Pere J. Ferrando* and Urbano Lorenzo-Seva

Research Center for Behavior Assessment

Universitat Rovira i Virgili

This note is about Bayes EAP scoring in the class of oblique linear and non-linear item response models that can be parameterized as factor analytic models. For these models, we propose an improved implementation approach that (a) provides more detailed and informative output, and (b) uses more prior information from the calibration stage. Overall, we discuss the limitations of EAP estimation as it is implemented in most available programs, provide a technical account of the basic methodology and proposed improvements, implement the proposal in the freely available FACTOR program, and illustrate how it works with FACTOR-based output.

Psychometric applications of Factor-analytic (FA) and Item Response Theory (IRT) models generally use a random-regressors two-stage estimation approach (McDonald, 1982). In the first stage (calibration), the structural item parameters are estimated and the goodness of model-data fit is assessed. In the second stage (scoring), the item parameter estimates are taken as fixed and known, and used to estimate the individual trait levels for each respondent. Of the various existing scoring procedures, this note is about Bayes expected a posteriori (EAP) estimation. According to some authors (e.g. Bartholomew, 1981), Bayes estimation is the only scoring approach that is theoretically justifiable for the models considered here. From a more pragmatic view, EAP estimates have two distinct advantages

* The research was supported by a grant from the Catalan Ministry of Universities, Research and the Information Society (2014 SGR 73) and by a grant from the Spanish Ministry of Education and Science (PSI2014-52884-P). Correspondence should be sent to: Pere Joan Ferrando. Departament de Psicologia, Facultat de Ciències de l'Educació i Psicologia, Universitat Rovira i Virgili. Carretera de Valls s/n 43007 Tarragona (Spain). Tel: +34 977 558079. E-mail: perejoan.ferrando@urv.cat

over existing alternatives. First, they use more information from the calibration stage via the prior distribution of the traits. Second, they have minimum square error and so they have the highest correlations with the ‘true’ traits they measure or predict (e.g. Grice, 2001, Muraki & Engelhard, 1985).

In this note we shall consider EAP estimates for those models that can be parameterized as FA models based on some type of correlation matrix. This class includes the unrestricted (exploratory) and restricted (confirmatory) linear FA model for continuous responses, the one- and two-parameter IRT models for binary responses (1PM, 2PM; Lord & Novick, 1968), and Samejima’s (1969) graded response model (GRM) for ordered-categorical responses (see e.g. Ferrando & Lorenzo-Seva, 2013, or McDonald, 1982). Because they all belong to the FA family, trait estimates obtained from these models can also be referred to as factor scores, which is the usual FA terminology (e.g. Lawey & Maxwell, 1971).

The theory of EAP scoring for the class of models above is well developed, and we make no claim to have obtained any new results here. Rather, the contention of this note is that, as it is implemented in most available programs, EAP estimation is sub-optimal and/or provides less information than would be recommendable. This state of affairs is particularly true for multidimensional models in which the traits are correlated (i.e. oblique models), the main focus of our proposal here. Overall, the main contribution of the note is practical: For the models so far that have been discussed to date we propose an improved EAP scoring approach based on two main points. First, we propose to use a more informative prior that makes use of the inter-factor correlation matrix obtained from the calibration stage. Second, we discuss how a more detailed and informative output in which the EAP point estimates are complemented with indicators of error and reliability can be obtained. Finally, an additional contribution of the note is that the proposals above will be implemented in a free, user-friendly program.

In the remainder of this note, we: *(a)* describe the general results and information that should be provided in EAP scoring, *(b)* discuss limitations when linear and nonlinear models are scored using EAP estimation, and *(c)* make the general proposal discussed above.

General Results

Results will be presented for any number of dimensions, but we shall emphasize the simple bidimensional case for didactic purposes. We first assume that the appropriate model has been fitted and that the model-data fit is acceptable. The structural estimates obtained (item parameters and inter-trait correlations) will next be taken as fixed and known, and used to obtain EAP trait estimates for *(a)* each individual in the group, or *(b)* new individuals belonging to the population in which the model holds. We shall consider a test made up of j=1,2… n items intended to measure p correlated traits distributed multivariate normal, with zero mean, unit variances, and correlation matrix $\mathbf{\Phi}$.

Let $\mathbf{x}_i$ be the full vector of responses given by individual i and $\boldsymbol{\theta}_i = [\theta_{i1}, \ldots \theta_{im}, \ldots \theta_{iq}]$ his/her 'true' trait levels. We shall use the generic expression $P(X_j/\boldsymbol{\theta})$ to denote the conditional probability (discrete case) or conditional density (continuous case) assigned to a specific item score for fixed $\boldsymbol{\theta}$. The likelihood of $\mathbf{x}_i$ can then be written generically as

$$L(\mathbf{x_i} \mid \boldsymbol{\theta}_i) = \prod_{j=1}^{n} P(X_{ij} \mid \boldsymbol{\theta}_i). \quad (1)$$

The EAP point estimate of $\boldsymbol{\theta}_i$ for the m dimension (θ_{im}) is the mean of the posterior distribution of $\boldsymbol{\theta}_i$ for that dimension given $\mathbf{x}_i$

$$EAP(\theta_{im}) = \frac{\int_{\boldsymbol{\theta}} \theta_m L(\mathbf{x_i} \mid \boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}} L(\mathbf{x_i} \mid \boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \quad (2)$$

The term $g(\boldsymbol{\theta})$ in (2) is the joint multivariate density of $\boldsymbol{\theta}$ so it contains information used or obtained in the calibration stage (in particular the inter-traits $\mathbf{\Phi}$ correlation matrix).

The diagonal elements of the posterior (error) covariance matrix are given by

$$PSD^2(\theta_{im}) = \frac{\int_{\boldsymbol{\theta}} (\theta_m - EAP(\theta_{im}))^2 L(\mathbf{x_i} \mid \boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\boldsymbol{\theta}} L(\mathbf{x_i} \mid \boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \quad (3)$$

In the bivariate case in particular, the posterior covariance matrix has the form:

$$\sum(\boldsymbol{\theta} \mid \mathbf{x_i}) = \begin{bmatrix} PSD^2(\theta_{i1}) & PCv(\theta_{i1},\theta_{i2}) \\ PCv(\theta_{i1},\theta_{i2}) & PSD^2(\theta_{i2}) \end{bmatrix} \tag{4}$$

where the *PCv*'s are the posterior covariances and the diagonal elements of $\sum(\boldsymbol{\theta} \mid \mathbf{x_i})$, which are given in (3), are the squares of the posterior standard deviations or PSDs. As the number of items increases, the distribution of the EAP estimates in (2) approaches normality (Chang & Stout, 1993), and the PSDs become equivalent to asymptotic standard errors (Bock & Mislevy, 1982). So, for a test of reasonable length, a normal-based confidence interval approach (strictly speaking, a credibility interval) for the EAP score of individual *i* in trait *m*, can be constructed as

$$\hat{\theta}_{im} \pm z_c PSD(\hat{\theta}_{im}). \tag{5}$$

Interval (5) indicates the degree of uncertainty around the estimated EAP score. So, the narrower the interval is, the more confidence we have that the EAP estimate is close to the true trait level.

Next, by recalling that the traits are distributed *N*(0,1), a conditional reliability coefficient for the EAP estimate of individual *i* in trait *m* can be obtained as (Bock & Mislevy, 1982, Green, Bock, Humphreys, Linn & Reckase, 1984)

$$\rho(\hat{\theta}_{im}) = 1 - \left[PSD(\hat{\theta}_{im})\right]^2. \tag{6}$$

Index (6) is a unitless measure that indicates the precision of the *m* trait estimate for this respondent. More precisely, it can be interpreted as the reliability of the scores corresponding to the sub-population of individuals with trait level θ_{im}.

Finally, an empirical marginal reliability estimate for the EAP scores on trait *m* can be obtained by averaging the squared PSDs in the sample of *N* individuals

$$\rho_m = 1 - \frac{\sum_i^N \left[PSD(\hat{\theta}_{im}) \right]^2}{N} \quad . \tag{7}$$

Coefficient (7) follows the basic definition of reliability in classical test theory (Green et al., 1984). Provided that the standard error of measurement (PSD in this case) remains relatively uniform across trait levels it is representative of the overall precision of the scores as measures of the corresponding dimension (Brown & Croudace, 2015).

In summary, as mentioned above, in our proposal the appropriate implementation of EAP score estimation in an oblique model involves: *(a)* obtaining point estimates that make use of the full prior information (in particular the **Φ** correlation matrix), *(b)* complementing the point estimates with measures of the reliability of these estimates: PSDs, confidence/credibility intervals and individual reliabilities, and *(c)* reporting marginal reliability estimates. We propose to call the new EAP score estimation *Fully-Informative Prior Oblique EAP scores*. In addition, while the expression (7) is not new, in order to make it clear that the expression has been computed on the basis of the new *Fully-Informative Prior Oblique EAP scores*, we propose to call the reliability estimates ***O****verall* ***R****eliability of fully-* ***I****nformative prior* ***O****blique* ***N****-EAP score*s (ORION).

The linear model

The structural correlation matrix implied by the linear oblique FA model is

$$\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{\Psi} \tag{8}$$

where **Λ** is the pattern loading matrix, **Φ** (defined above) is the inter-factor correlation matrix and **Ψ** is the diagonal matrix of the item residual variances. In this model the EAP point estimates in the q dimensions and the posterior error matrix can be obtained in closed form and are respectively (Lawley & Maxwell, 1971)

$$EAP(\boldsymbol{\theta}_i) = \mathbf{\Phi\Lambda}'\mathbf{\Sigma}^{-1}\mathbf{Z} \tag{9}$$

where **Z**, of dimension $n \times N$ is the data matrix containing the standardized item scores of the assessed respondents, and

$$\sum(\boldsymbol{\theta} \mid \mathbf{x}_i) = \left[\boldsymbol{\Phi}^{-1} + \boldsymbol{\Lambda}'\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}\right]^{-1}. \tag{10}$$

The EAP point estimates in (9) are known in FA terminology as "regression factor scores for the oblique model" (Lawley & Maxwell, 1971). As for the PSDs in (10) it is noted that they do not depend on $\boldsymbol{\theta}$. So, in this case, for a given trait the standard errors are the same for all the respondents and the conditional reliability (6) is the same as the marginal reliability (7). The computation of the results proposed here, then, is quite straightforward in this model. However, they are not all usually implemented in the programs available. Thus, regression factor scores can be obtained (e.g. using SPSS) but only the point estimates (9) are provided.

Non-linear models

For binary responses fitted by the 1PM or the 2PM, or graded responses fitted by the GRM, no closed form expressions exist for (2) and (3). In these cases the multiple integrals are approximated as accurately as required using numerical quadrature, generally rectangular quadrature over q equally spaced points (Bock & Mislevy, 1982). This type of implementation is relatively simple but computationally expensive, and the computational demands increase exponentially with the number of traits. For this reason, practical implementations in these cases usually operate as if the traits were uncorrelated (e.g. Muraki & Engelhard, 1985). To illustrate this point we shall again consider the bidimensional case. If θ_1 and θ_2 are independent, then $g(\theta_1,\theta_2)= g(\theta_1)\, g(\theta_2)$, and the integral expression (2) becomes

$$\hat{\theta}_{im} = E(\theta_m \mid \mathbf{x}_i) = \frac{\int\limits_{\theta 1}\int\limits_{\theta 2} \theta_m L(\mathbf{x}_i \mid \theta_1, \theta_2) g(\theta_1) g(\theta_2) d\theta_1 d\theta_2}{\int\limits_{\theta 1}\int\limits_{\theta 2} L(\mathbf{x}_i \mid \theta_1, \theta_2) g(\theta_1) g(\theta_2) d\theta_1 d\theta_2} \tag{11}$$

where m=1 or 2. The quadrature approximation is then

$$\hat{\theta}_{im} \cong \frac{\sum_{k1=1}^{q}\sum_{k2=1}^{q} X_{km} L(\mathbf{x}_i \mid X_{k1,} X_{k2,}) W(X_{k1}) W(X_{k2})}{\sum_{k1=1}^{q}\sum_{k2=1}^{q} L(\mathbf{x}_i \mid X_{k1,} X_{k2,}) W(X_{k1}) W(X_{k2})} \tag{12}$$

where X_{k1} and X_{k2} are the nodes and $W(X_{k1})$ and $W(X_{k2})$ are the weights for the one dimensional quadratures that approximate the distributions of θ_1 and θ_2, respectively. To sum up, estimation is simplified by using unidimensional quadrature, but the information regarding the inter-factor correlations is lost. The same occurs when computing the posterior covariance matrix (3) which is approximated by rectangular quadrature in the same way as (12).

The contention of this note is that, at present, full information EAP estimation is feasible at least up to a reasonable number of factors (say 5) without having to resort to unidimensional quadrature approximations. The approach taken here consists of creating a *p*-dimensional grid with equally-spaced distances, and to compute the volume element corresponding to each mess of the grid. Turning again to the bidimensional case, let *d* be the common inter-point distance, and the mesh of the grid be defined by the points X_{k1} and X_{k2} in (12). So, the volume element $g(\theta_1,\theta_2)d\theta_1 d\theta_2$ is approximated by $g(X_{k1}, X_{k2})d^2$, where $g(X_{k1}, X_{k2})$ is the ordinate of the standard bivariate normal distribution at the point of the plane with coordinates X_{k1} and X_{k2}.

Finally, regarding the additional information recommended here, PSDs are usually provided in the unidimensional case by such standard IRT programs as BILOG and MULTILOG. In the multidimensional case, however, to the best of our knowledge they are only provided by TESTFACT and, for restricted models, by Mplus. In addition, some R packages, mainly *Itm* and *mirt,* provide PSDs for binary and graded multidimensional IRT models.

Implementation in FACTOR and an illustrative example

At the same time as this note was written, all the proposals made were implemented and tested in an experimental version of FACTOR 10. The new implementation will be available in the next release of the program under the term: *Fully-Informative Prior Oblique EAP scores*. The most critical of the technical details is the number of quadrature points. We chose to approximate the multiple integrals by using a number of points which depended on the number of factors: for 2 factors, 21 equally-spaced points were used; for 3 factors, 17 points were used; for 4 factors, 11 points were used; and for 5 or more factors, 7 points were used. In unidimensional implementations, Thissen and Orlando (2001) recommended using 30-50 points. However, they also acknowledge that rectangular quadrature is very robust and they found that if the number of points was increased from 7 to 46, the precision of the estimates only increased slightly. So, our setting seems reasonable.

To illustrate the functioning of the proposal, we now provide the output obtained after the functions were implemented in FACTOR. The illustration is based on the analysis of the responses of 1,042 participants. In particular, we analysed participants' responses to two scales (*Self-reliance* and *Identity*), which are expected to be correlated (Morales-Vives, Camps and Lorenzo-Seva, 2013). Each scale is made up of 7 items, each of which is scored in a 5-point Likert format.

For illustrative purposes, the item responses were treated as both continuous variables (i.e. the linear FA model) and ordered-discrete variables (i.e. the GRM). In the continuous case, the Pearson correlation matrix was factor analysed using Unweighted Least Squares (ULS) FA, and the two factors retained were obliquely rotated using Promin (Lorenzo-Seva, 1999). The observed inter-factor correlation was .348. In the GRM, the polychoric correlation matrix was factor analysed and rotated using the same procedure, and the observed inter-factor correlation was .393. In both cases, factor scores were estimated with standard EAP scoring (i.e., EAP estimates were obtained without using the inter-factor correlation), and Fully-Informative Prior Oblique EAP. The reliabilities of the EAP factor scores in the four different conditions are shown in Table 1. As can be observed in the table, when the inter-factor correlation observed in the data is taken into account, the reliability of the EAP factor scores increases (i.e., the precision of the factor score estimates is higher). In addition, the oblique graded EAP (i.e., when the GRM model is used, and the inter-factor correlation observed in the data is taken into account) is the approach that gave the highest reliabilities. Finally, Table 2 shows the outcome of the first three participants in our sample for the "Identity" factor as examples of how the results on factor scores have been implemented in FACTOR (Lorenzo-Seva & Ferrando, 2013).

A simulation study

The responses of 1,000 individuals were simulated for a number of items and two underlying factors. The simulated data were generated according to a linear common factor model (MacCallum & Tucker, 1991), in which the resulting continuous variables were categorized to yield ordered polytomous observed variables. We manipulated the number of items per factor (5, 10, and 20), and the inter-factor correlation in the population (0, .20, and .50). For each condition in the simulation study, 100 replicas were computed.

Table 1. Reliability of EAP factor scores in the illustrative example.

Method	Self-reliance	Identity
Standard Continuous	.626	.660
Standard Graded	.663	.683
Phi-info Continuous	.676	.702
Phi-info Graded	.810	.861

Table 2. Oblique graded EAP factor score estimates and precision for the Identity factor in the illustrative example.

Participant	Estimated factor score	Approximate 90% confidence interval	Posterior SD (PSD)	Reliability
1	-0.086	(-0.627 ; 0.454)	0.329	.892
2	-1.350	(-1.898 ; -0.802)	0.333	.889
3	1.786	(1.003 ; 2.569)	0.476	.773

For each dataset, factor scores were estimated with standard EAP scoring (i.e., EAP estimates were obtained without using the inter-factor correlation), and Fully-Informative Prior Oblique EAP. The descriptive statistics of the reliabilities of the EAP factor scores in the four different conditions are shown in Table 3. As in the illustrative example, when the inter-factor correlation is taken into account, the reliability of the EAP factor scores increases (even if the inter-factor correlation in the population is expected to be zero). As expected, however, when there is a large number of items per factor, the increment in the precision is smaller.

Table 3. Outcome of simulation study. Averaged reliabilities of EAP factor scores. Standard deviations are shown in brackets.

n/p	phi	S-C	S-G	PHII-C	PHII-G
5	.00	.677 (.012)	.682 (.013)	.723 (.013)	.793 (.010)
	.20	.677 (.011)	.682 (.012)	.725 (.012)	.795 (.009)
	.50	.680 (.013)	.684 (.014)	.741 (.014)	.805 (.009)
10	.00	.804 (.005)	.812 (.006)	.835 (.006)	.892 (.004)
	.20	.804 (.006)	.812 (.007)	.836 (.007)	.892 (.005)
	.50	.803 (.007)	.812 (.008)	.841 (.007)	.895 (.005)
20	.00	.892 (.003)	.896 (.004)	.910 (.003)	.943 (.002)
	.20	.892 (.003)	.896 (.004)	.911 (.003)	.944 (.002)
	.50	.891 (.004)	.895 (.005)	.911 (.004)	.944 (.003)
Overall		.791 (.088)	.797 (.088)	.826 (.075)	.878 (.061)

Notes
S-C: Standard Prior Continuous;
S-G: Standard Prior Graded;
PHII-C: Fully-Informative Prior Continuous;
PHII-G: Fully-Informative Prior Graded

Discussion

This note has proposed an improved approach to EAP scoring within a general class of item response models. The approach is based on results that are theoretically known but which are not generally used in current applications. Overall, we believe that the proposal is of interest for several reasons. First, the class of models we address contains possibly the most used models in substantive applications. Second, the proposal is feasible and provides information that enhances the interpretation of the scoring. Finally, the proposal is implemented in a general, widely used, and non-commercial program.

The feasibility of the "Fully-Informative Prior" procedures which use the inter-factor correlations as prior information raises some points of interest. The present studies (both the empirical study and the simulation) suggest that the additional information contributed by $\mathbf{\Phi}$ improves the accuracy of the estimates, and, therefore, the extent to which the factor scores represent the true trait levels. However, further research is clearly needed. It would also be interesting to assess the extent to which the improved prior improves correlational accuracy (the extent to which the correlated factor scores match the correlations between the true trait levels; e.g. Grice, 2001). Finally, it would also be of interest to assess the extent to which the additional information from $\mathbf{\Phi}$ compensates for a decreasing number of quadrature points with respect to the widely used one-dimensional approximations. All these studies will be more feasible when the procedures are implemented in FACTOR.

A further potential line of research concerns improvements in the quadrature approximations. As mentioned above, the rectangular scheme adopted here is remarkably simple and robust but computationally expensive, and, at present, there are far more efficient approaches. In particular, adaptive multivariate Gauss-Hermite quadrature (e.g. Jackel, 2005) is an approach we intend to implement in the near future.

RESUMEN

Nota sobre un procedimiento mejorado para la estimación de rasgos latentes en los modelos del análisis factorial oblicuo y de la teoría de respuesta a los ítems. En esta nota se discute la estimación Bayesiana EAP en aquella familia de modelos oblicuos de respuesta al ítem que pueden ser parametrizados como modelos lineales o no lineales de análisis factorial. Para estos modelos proponemos una implementación mejorada que (a) proporciona un output más detallado e informativo y (b) utiliza más información a priori obtenida en la etapa de calibración. En conjunto, se

discuten las limitaciones de la estimación EAP tal como está implementada en muchos de los programas actualmente disponibles, se proporciona un resumen técnico de la metodología básica y de las propuestas de mejora, se implementan los procedimientos propuestos en el programa no comercial FACTOR, y, finalmente, se ilustra su funcionamiento con una salida del programa.

REFERENCES

Bartholomew, D. J. (1981). Posterior analysis of the factor model. *British Journal of Mathematical and Statistical Psychology, 34*, 93-99. doi: 10.1111/j.2044-8317.1981.tb00620.x

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*, 431-444.

Brown, A., & Croudace, T. (2015). *Scoring and estimating score precision using multidimensional IRT*. In Reise, S. P. & Revicki, D. A. (Eds.). Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment. New York: Routledge.

Chang, H., & Stout, W. (1993). The asymptotic posterior normality of the latent trait in an IRT model. *Psychometrika,* 58, 37-52. doi: 10.1007/BF02294469.

Ferrando, P. J., & Lorenzo-Seva, U. (2013). *Unrestricted item factor analysis and some relations with item response theory*. Technical Report. Department of Psychology, Universitat Rovira i Virgili, Tarragona. Retrieved from http://psico.fcep.urv.cat/utilitats/factor/Documentation.html

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R.L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptative tests. *Journal of Educational Measurement, 21*, 347–360.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6, 430-450. doi: 10.1037//1082-989X.6.4.430

Jackel, P. (2005). A note on multivariate Gauss-Hermite quadrature. *London: ABN-Amro. Re*.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. London. Butterworths.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading (MA): Addison-Wesley.

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research*, *34*, 347-365. doi: 10.1207/S15327906MBR3403_3

Lorenzo-Seva, U., & Ferrando, P. J. (2013). FACTOR 9.2 A Comprehensive Program for Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models. *Applied Psychological Measurement*, *37*, 497-498. doi: 10.1177/0146621613487794

Morales-Vives, F., Camps, E., & Lorenzo-Seva, U. (2013). Development and validation of the Psychological Maturity Assessment Scale (PSYMAS). *European Journal of Psychological Assessment*, 29, 12-18. doi: 10.1027/1015-5759/a000115

MacCallum, R. C., & Tucker, L. R. (1991). Representing sources of error in the common-factor model: Implications for theory and practice. *Psychological Bulletin*, *109*, 502-511. doi:10.1037/0033-2909.109.3.502

McDonald, R. P. (1982). Linear vs. non linear models in Item Response Theory. *Applied Psychological measurement*, *6*, 379-396.

Muraki, E., & Engelhard, G. (1985). Full-information factor analysis: applications of EAP scores. *Applied Psychological Measurement,* 9, 417-430.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement,* 34. doi:10.1007/BF02290599

Thissen, D., & Orlando, M. (2001). *Item response theory for items scored in two categories.* In D. Thissen & H. Wainer (Eds.), Test scoring (pp. 73–140) Mahwah, NJ: LEA.

(Manuscript received: 11 January 2016; accepted: 21 April 2016)