



A Note on the Eigensystem of the Covariance Matrix of Dichotomous Guttman Items

Clinton P. Davis-Stober^{1*}, Jean-Paul Doignon² and Reinhard Suck³

¹ Department of Psychological Sciences, University of Missouri, Columbia, MO, USA, ² Department of Mathematics, Université Libre de Bruxelles, Brussels, Belgium, ³ Universität Osnabrück, Osnabrück, Germany

We consider the covariance matrix for dichotomous Guttman items under a set of uniformity conditions, and obtain closed-form expressions for the eigenvalues and eigenvectors of the matrix. In particular, we describe the eigenvalues and eigenvectors of the matrix in terms of trigonometric functions of the number of items. Our results parallel those of Zwick (1987) for the correlation matrix under the same uniformity conditions. We provide an explanation for certain properties of principal components under Guttman scalability which have been first reported by Guttman (1950).

Keywords: Guttman scale, dichotomous items, Rasch model, principal component analysis, eigenvalues, eigenvectors

OPEN ACCESS

Edited by:

Pietro Cipresso,
IRCCS Istituto Auxologico Italiano, Italy

Reviewed by:

Juergen Heller,
Universität Tübingen, Germany
Tomer Fekete,
KU Leuven, Belgium

*Correspondence:

Clinton P. Davis-Stober
stoberc@missouri.edu

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 19 May 2015

Accepted: 04 November 2015

Published: 01 December 2015

Citation:

Davis-Stober CP, Doignon J-P and
Suck R (2015) A Note on the
Eigensystem of the Covariance Matrix
of Dichotomous Guttman Items.
Front. Psychol. 6:1767.
doi: 10.3389/fpsyg.2015.01767

1. INTRODUCTION

Guttman scales form the conceptual foundation for modern Item Response Theory (IRT). For example, Guttman scales underlie the Rasch model (e.g., Andrich, 1985) as well as Mokken scales (e.g., van Schuur, 2003),—see Tenenhaus and Young (1985) and Lord and Novick (1968) for classic reviews and discussions of Guttman scaling. Under the auspices of understanding the principal component structure of unidimensional scales, Guttman (1950) derived several important properties relating to the correlation matrix of perfect dichotomous Guttman items. Later work by Zwick (1987) identified that the eigenvalues corresponding to this matrix can be written as simple functions of the number of items, under a set of uniformity conditions.

In this brief note, we extend the results of Zwick (1987) by considering the covariance matrix of dichotomous Guttman items under these same uniformity conditions. We derive closed-form solutions for the eigenvalues and eigenvectors of this matrix, for any number of items. In particular, we provide expressions in terms of simple trigonometric functions of the number of items. These expressions lead to a simple explanation of the signing relationships among principal components for Guttman scales first described by Guttman (1950).

2. MAIN RESULTS

The core idea of a Guttman scale is that the set of items under consideration forms a unidimensional scale, i.e., if a person obtains a correct response to an item then this person would obtain a correct response to all “easier” items. **Table 1** presents a matrix of response patterns conforming to a perfect Guttman scale for five items, with Item 5 being the most “difficult” and Item 1 being the “easiest.”

As in Zwick (1987), we consider the following two assumptions. First, we assume that all items are distinct, i.e., no two items produce identical responses for all possible response patterns. Second,

TABLE 1 | An example of a perfect Guttman scale for five items.

Response pattern	Item 1	Item 2	Item 3	Item 4	Item 5
1	0	0	0	0	0
2	1	0	0	0	0
3	1	1	0	0	0
4	1	1	1	0	0
5	1	1	1	1	0
6	1	1	1	1	1

A value of 0 indicates an incorrect response to the item, a value of 1 indicates a correct response.

we assume that the probability of obtaining each response pattern is $\frac{1}{n+1}$, where n is the number of items, i.e., a uniform distribution over response patterns. This last assumption is rather strong, given that responses are typically modeled using a normal distribution. While we assume uniformly distributed response patterns primarily for mathematical tractability, we demonstrate later via simulations that our results approximate those obtained from a normal distribution under highly discriminating items that are equally spaced by difficulty.

Under our assumptions, the covariance between any items i and j , with $i \leq j$, is equal to the following:

$$\sigma_{i,j}^{cov} = \frac{i(n+1-j)}{(n+1)^2}, \quad \forall i, j \in \{1, 2, \dots, n\} \text{ with } i \leq j. \quad (1)$$

As one would expect, Equation (1) is closely related to the Pearson product-moment correlation, which, as described by Zwick (1987), is equal to:

$$\text{Corr}(i, j) = \sqrt{\frac{i(n+1-j)}{j(n+1-i)}}, \quad \forall i, j \in \{1, 2, \dots, n\} \text{ with } i \leq j. \quad (2)$$

Parallel to Zwick (1987) and Guttman (1950), who handled the correlation matrix, we consider the $n \times n$ covariance matrix defined by Equation (1). We first provide the n distinct eigenvalues.

Proposition 1. *The covariance matrix σ^{cov} , with entries given by Equation (1), has its eigenvalues equal to (in decreasing order)*

$$\lambda_i^{cov} = \frac{1}{(n+1)(2 - 2\cos(\frac{i\pi}{n+1}))}, \quad i = 1, 2, \dots, n. \quad (3)$$

The proof is in the Appendix.

Note how i and n determine the period of the cosine term in the denominator of the right-hand side of Equation (3). From the same equation, the maximal eigenvalue for any fixed number of items n is equal to $\lambda_n^{max} = \frac{1}{(n+1)(2 - 2\cos(\frac{\pi}{n+1}))}$. Note that as $n \rightarrow \infty$, $\lambda_n^{max} \rightarrow \infty$. Also, the eigenvalues of the covariance matrix are very different from the eigenvalues of the Pearson product correlation matrix, which, as described by Zwick (1987), are equal to $\lambda_i^{corr} = \frac{n+1}{i(i+1)}$.

The eigenvectors of the covariance matrix also have an elegant, closed-form expression.

TABLE 2 | This table presents the eigenvector components of the covariance matrix for $n = 5$.

P_1	P_2	P_3	P_4	P_5
$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$
$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$	0	$-\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$
1	0	-1	0	1
$\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$	0	$\frac{\sqrt{3}}{2}$	$-\frac{\sqrt{3}}{2}$
$\frac{1}{2}$	$-\frac{\sqrt{3}}{2}$	1	$-\frac{\sqrt{3}}{2}$	$\frac{1}{2}$

The eigenvectors are arrayed in descending order according to the corresponding eigenvalue, i.e., P_1 is the eigenvector corresponding to the maximal eigenvalue.

Proposition 2. *For the covariance matrix σ^{cov} defined by Equation (1), an eigenvector P_i of eigenvalue λ_i^{cov} , with $i = 1, 2, \dots, n$ (as in Proposition 1), results from setting*

$$P_{i,m} = \sin\left(\frac{im\pi}{n+1}\right), \quad m = 1, 2, \dots, n. \quad (4)$$

The proof is in the Appendix.

Guttman (1950) derived a series of relationships on the eigenvector components of correlation matrices based on perfect, “error free” scales. Let $sgn(x)$ be the sign function of the value x . Define a *sign change* of an eigenvector $P_{i,m}$ as a value j such that $sgn(P_{i,j}) \neq sgn(P_{i,j+1})$, $j \in \{1, 2, \dots, n\}$. As described in Guttman (1950), for n -many items there exists exactly one eigenvector with no sign changes, one eigenvector with a single sign change, one with two sign changes, and so on, with the eigenvector corresponding to the smallest eigenvalue having exactly $n-1$ sign changes. This symmetry can be seen in **Table 2**, which presents the eigenvectors in Equation (4) for $n = 5$. As made explicit by Equation (4), these sign changes result from the symmetry of the sine function as the values of i and m vary.

3. COMPARISON TO IRT DATA

In this section, we illustrate how our analytic results could be used to evaluate responses conforming to modern IRT models. We consider the well-known *two parameter logistic* (2PL) model, where the probability of a correct response to item i is defined as follows:

$$p_i(\theta) = \frac{1}{1 + \exp^{-a_i(\theta - b_i)}}, \quad (5)$$

where $a_i \in \mathbb{R}^+$ is the item discrimination parameter, $b_i \in \mathbb{R}$ is the item difficulty parameter and $\theta \in \mathbb{R}$ is the person-specific ability parameter.

From the perspective of the 2PL model, Guttman items are obtained by letting the a_i (item discrimination) parameter values become arbitrarily large (e.g., van Schuur, 2003), i.e., the probability of a test taker correctly answering an item given that their latent skill is higher (lower, resp.) than the item difficulty is 1 (0 resp.). Our results provide a new perspective on the item covariance and principal component structure of 2PL items under the idealized conditions of a Guttman scale.

Indeed, one could consider the eigenvalues and eigenvectors in Equations (3–4) as an error-free ideal for such response data, under our assumption of a uniform distribution over response patterns.

In the next section, we compare our results to simulated data that relax the assumption of a uniform distribution over response patterns. In the first simulation study, we compare our results to data generated from a Rasch model (Equation 5 with $a_i = 1, i = 1, 2, \dots, n$) where the person specific ability parameter, θ , is randomly drawn from a standard normal distribution. For the second simulation study, we consider a setup nearly identical to the first, with the exception that we consider large values of a_i for each item, i.e., high discrimination among items.

3.1. Simulation Study 1

For this study, we considered six conditions comprised of: 4, 6, 8, 16, 32, and 64 test items. For each condition, the difficulty of the items, b_i , was equally spaced along the interval $[-1, 1]$. For each condition, we randomly sampled 5000 values of θ from a standard normal distribution (e.g., Anderson et al., 2007). We obtained simulated responses to the items by applying the sampled θ values, and item difficulties, b_i , to Equation (5), with $a_i = 1$ for all test items, i.e., a Rasch model. Thus, for each condition, we have 5000 simulated responses to the test items.

For each condition, we computed the covariance matrix of the items using the 5000 simulated responses, i.e., we calculated the sample covariance of the 5000 responses. We then numerically

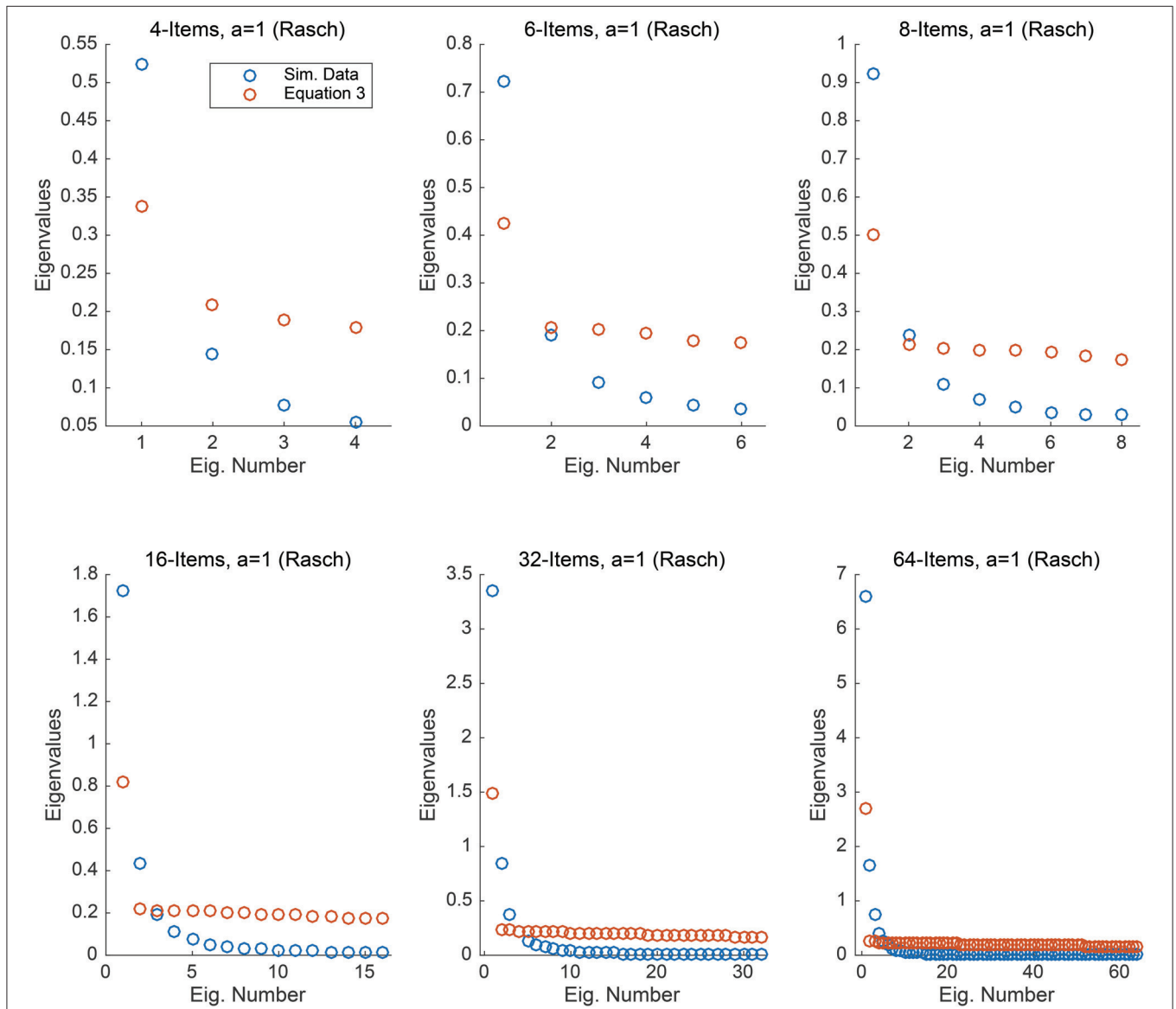
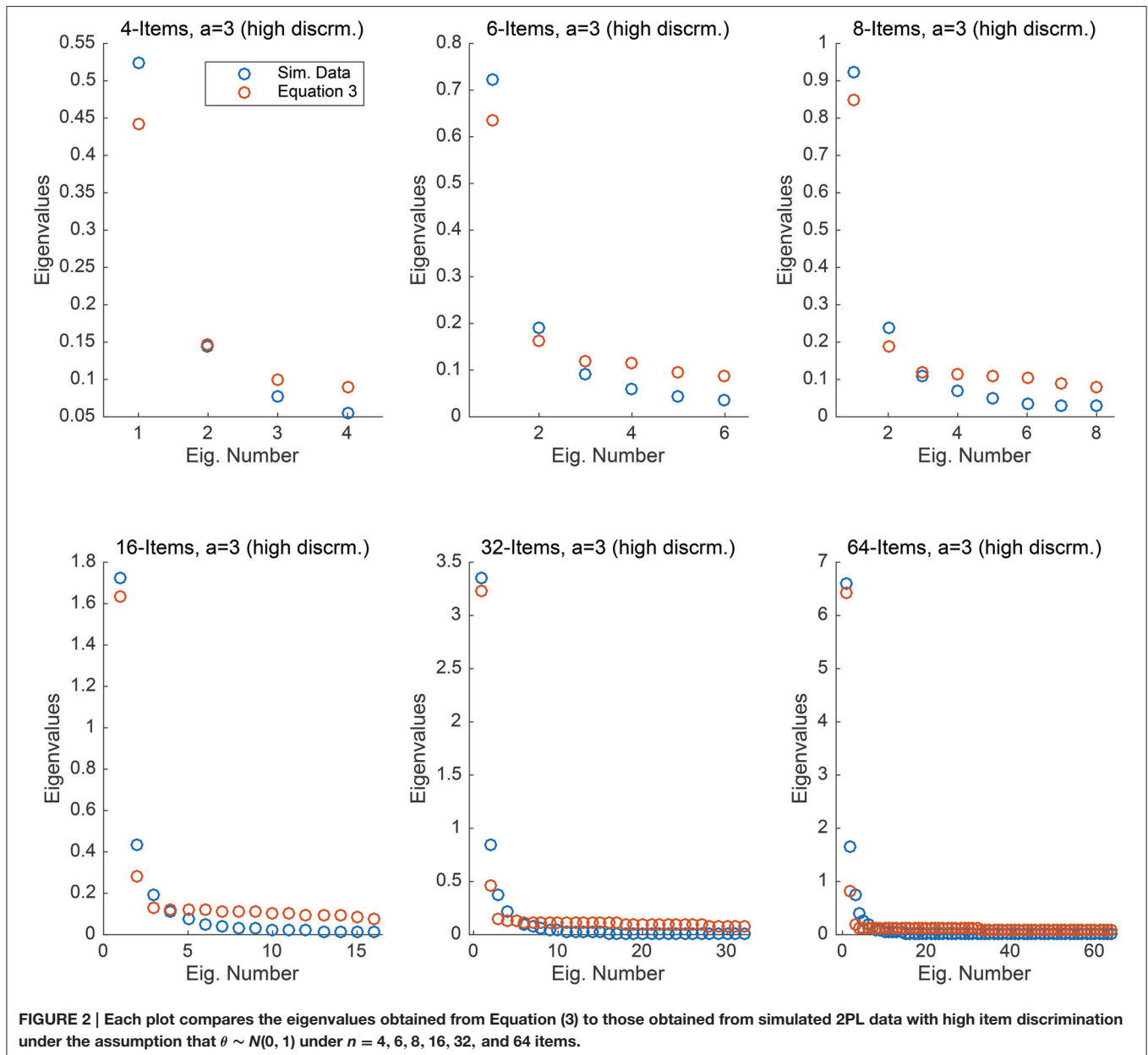


FIGURE 1 | Each plot compares the eigenvalues obtained from Equation (3) to those obtained from simulated Rasch data under the assumption that $\theta \sim N(0, 1)$ under $n = 4, 6, 8, 16, 32,$ and 64 items.

calculated the eigenvalues of this covariance matrix. **Figure 1** compares the eigenvalues obtained from the simulated data to the eigenvalues obtained from Equation (3), for each condition. It is interesting to note that the largest eigenvalue for the simulated data is always larger than the maximal eigenvalue obtained via Equation (3), this is similar to results obtained by Zwick (1987) within the context of the Guttman correlation matrix. In general, moving to a probabilistic response model (the Rasch model) and sampling the θ values from a normal distribution appears to yield covariance eigenvalues that greatly differ from those obtained in Equation (3). As we show in the next study, improving item discrimination will yield different results.

3.2. Simulation Study 2

In this simulation study, we consider nearly identical conditions to the first, with the exception that the item discrimination parameters, a_i , are large in size, indicating excellent item discrimination. As in the previous study, we considered six conditions comprised of: 4, 6, 8, 16, 32, and 64 test items. For each condition, the difficulty of the items, b_i , was equally spaced along the interval $[-1, 1]$. As before, for each condition, we randomly sampled 5000 values of θ from a standard normal distribution. We obtained simulated responses to the items by applying the sampled θ values, and item difficulties, b_i , to Equation (5), with $a_i = 3, i = 1, 2, \dots, n$, indicating excellent item discrimination. As before, for each



condition, we have 5000 simulated responses to the test items.

For each condition, we computed the covariance matrix of the 5000 simulated responses and numerically calculated the eigenvalues of the generated covariance matrix for each condition. **Figure 2** compares the eigenvalues from these simulated data to the eigenvalues obtained via Equation (3), for each condition. It is interesting to note that there is a much closer correspondence between the two sets of eigenvalues under these conditions. Further, this relationship becomes stronger as the number of equally spaced items increases, yielding nearly a perfect match to the maximal eigenvalue as the number of items reaches 32 and 64.

This study illustrates that our analytic results, which are derived under the strong assumption of uniformly distributed response patterns, may be useful as an approximation even when the ability parameter is normally distributed. This approximation is best when the difficulty range of the items are within a single standard deviation of the mean and the items have excellent discriminability. As the range of the item difficulty increases and/or the variance of the ability parameter distribution shrinks, the approximation becomes much poorer. Our Matlab code for generating these graphs and exploring other configurations is available as an online supplement.

REFERENCES

- Anderson, C. J., Li, Z., and Vermunt, J. K. (2007). Estimation of models in a Rasch family for polytomous items and multiple latent variables. *J. Stat Soft.* 20, 1–36. doi: 10.18637/jss.v020.i06
- Andrich, D. (1985). “An elaboration of Guttman scaling with Rasch models for measurement,” in *Sociological Methodology 1985. Jossey-Bass Social and Behavioral Science Series*, ed N. B. Tuma (San Francisco, CA: Jossey-Bass), 33–80.
- Bünger, F. (2014). Inverses, determinants, eigenvalues, and eigenvectors of real symmetric Toeplitz matrices with linearly increasing entries. *Linear Algebra Appl.* 459, 595–619. doi: 10.1016/j.laa.2014.07.023
- Elliott, J. F. (1953). *The Characteristic Roots of Certain Real Symmetric Matrices*. Masters thesis, University of Tennessee.
- Gregory, R. T., and Karney, D. (1969). *A Collection of Matrices for Testing Computational Algorithm*. New York, NY: Wiley-Interscience.
- Guttman, L. (1950). “The principal components of scale analysis,” in *Measurement and Prediction*, eds S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star, and J. A. Clausen (Princeton, NJ: Princeton University Press), 312–361.
- Lord, F. M., and Novick, M. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Tenenhaus, M., and Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scaling, dual scaling, homogeneity analysis

4. CONCLUSION

We derived closed-form solutions for the eigenvalues and eigenvectors of the covariance matrix of dichotomous Guttman items, under a uniform sampling assumption. We demonstrated that these eigenvalues and eigenvectors are simple trigonometric functions of the number of items, n . Our results parallel those of Zwick (1987), who examined the eigenvalues of the correlation matrix of dichotomous Guttman items under the same uniformity assumptions. It remains an open question whether the eigenvectors of the correlation matrix, as investigated by Zwick (1987), can also be solved for explicitly.

ACKNOWLEDGMENTS

We would like to thank Edgar Merkle, Jay Verkuilen, and David Budescu for comments on an earlier draft. Davis-Stober was supported by National Science Foundation grant (SES-1459866, PI: Davis-Stober).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpsyg.2015.01767>

- and other methods for quantifying categorical multivariate data. *Psychometrika* 50, 91–119. doi: 10.1007/BF02294151
- van Schuur, W. H. (2003). Mokken scale analysis: between the Guttman scale and parametric item response theory. *Polit. Anal.* 11, 139–163. doi: 10.1093/pan/mpg002
- Yueh, W.-C. (2005). Eigenvalues of several tridiagonal matrices. *Appl. Math. E-Notes* 5, 66–74.
- Yueh, W.-C., and Cheng, S. S. (2008). Explicit eigenvalues and inverses of tridiagonal Toeplitz matrices with four perturbed corners. *ANZIAM J.* 49, 361–387. doi: 10.1017/S1446181108000102
- Zwick, R. (1987). Some properties of the correlation matrix of dichotomous Guttman items. *Psychometrika* 52, 515–520. doi: 10.1007/BF02294816

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2015 Davis-Stober, Doignon and Suck. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

A. Proofs of Propositions 1 and 2

To prove the main results of the paper, we first derive the general inverse of the covariance matrix of Guttman items under our uniformity assumptions. This inverse has a special tridiagonal form. From this tridiagonal form, we apply known algebraic results to obtain the required eigenvalues and eigenvectors.

Define X as the $(n + 1) \times n$ matrix of perfect Guttman scores under the specified uniformity assumptions. The rows of this matrix correspond to response patterns while the columns correspond to Guttman items, see also **Table 1**. This matrix has zeros on the diagonal and above, and all elements below the diagonal are ones:

$$X = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix}.$$

We denote by $e^{(i)}$ the i -th vector of the canonical basis of \mathbb{R}^{n+1} , and by $v^{(i)}$ the i -th column vector of X . For later use, it is convenient to introduce also $v^{(0)} = (1, 1, \dots, 1)'$ and $v^{(n+1)} = (0, 0, \dots, 0)'$. Thus in \mathbb{R}^{n+1} we have for $i = 1, 2, \dots, n + 1$,

$$v^{(i)} = v^{(i-1)} - e^{(i)},$$

and then also, for $i = 1, 2, \dots, n$,

$$\begin{aligned} v^{(i)} &= \frac{1}{2} (v^{(i-1)} - e^{(i)}) + \frac{1}{2} (v^{(i+1)} + e^{(i+1)}) \\ &= \frac{1}{2} (v^{(i-1)} + v^{(i+1)}) + \frac{1}{2} (e^{(i+1)} - e^{(i)}). \end{aligned} \quad (A1)$$

Obtaining the eigenvalues and eigenvectors of the covariance matrix via the columns $v^{(i)}$ of X can be done in five steps:

1. centering each vector $v^{(i)}$ (for $i = 1, 2, \dots, n$), that is, subtracting the mean of all components of $v^{(i)}$ from each component; let us denote by $\tilde{v}^{(i)}$ the resulting vector;
2. computing the element S_{ij} of the matrix S as the scalar product $\tilde{v}^{(i)} \cdot \tilde{v}^{(j)}$;
3. deriving the inverse of S by taking into account a special property of the rows of S (see below);
4. inferring the eigenvalues and eigenvectors of S^{-1} (then also of S) from the special form of S^{-1} , a tridiagonal matrix;
5. finally, observing that the covariance matrix equals $\sigma^{cov} = \frac{1}{(n+1)}S$, thus obtaining the eigenvalues and eigenvectors of $\hat{\sigma}^{cov}$.

Let us rephrase these steps in a more geometric fashion. In Step 1, $\tilde{v}^{(i)}$ is the image of $v^{(i)}$ by the orthogonal projection from \mathbb{R}^{n+1} to the hyperplane H with equation $\sum_{i=1}^{n+1} x_i = 0$ (indeed, $\tilde{v}^{(i)} \in H$ and furthermore $\tilde{v}^{(i)} - v^{(i)}$, a constant vector, is orthogonal to

H). Moreover, notice that $e^{(i)} - e^{(j)}$ belongs to H and so projects onto itself. Consequently, for $i = 1, 2, \dots, n$, we derive from Equation (A1)

$$\tilde{v}^{(i)} = \frac{1}{2} (\tilde{v}^{(i-1)} + \tilde{v}^{(i+1)}) + \frac{1}{2} (e^{(i+1)} - e^{(i)}). \quad (A2)$$

In Step 2, we compute $S_{i,j}$ as the scalar product of $\tilde{v}^{(i)}$ with $\tilde{v}^{(j)}$. Taking the scalar product of both sides of the previous equation with $\tilde{v}^{(j)}$, we get for $i = 2, 3, \dots, n - 1$ and $j = 1, 2, \dots, n$

$$S_{i,j} = \frac{1}{2} (S_{i-1,j} + S_{i+1,j}) + \frac{1}{2} (\tilde{v}_{i+1}^{(j)} - \tilde{v}_i^{(j)}).$$

Now because

$$\tilde{v}_{i+1}^{(j)} - \tilde{v}_i^{(j)} = v_{i+1}^{(j)} - v_i^{(j)} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

we see that row $S_{i,\bullet}$ is the mean of rows $S_{i-1,\bullet}$ and $S_{i+1,\bullet}$ except for its diagonal element which is $\frac{1}{2}$ more than the mean. This holds for $i = 2, 3, \dots, n - 1$. By considering extraneous rows $S_{0,\bullet} = (0, 0, \dots, 0)$ and $S_{n+1,\bullet} = (0, 0, \dots, 0)$, we can also allow $i = 1$ and $i = n$ [this follows again from (A2), considered now for $i = 1$ and $i = n$, together with $\tilde{v}^{(0)} = \tilde{v}^{(n+1)} = (0, 0, \dots, 0)'$]. This special property of S immediately translates into the following expression for the inverse matrix of S :

$$S^{-1} = \begin{pmatrix} 2 & -1 & 0 & 0 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 2 & -1 & 0 & \dots & 0 \\ 0 & 0 & -1 & 2 & -1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & -1 & 2 & -1 \\ 0 & 0 & \dots & 0 & 0 & -1 & 2 \end{pmatrix}.$$

(indeed, the product of the above matrix with S equals the identity matrix).

The form of S^{-1} follows a particular tridiagonal form that has been extensively studied in the mathematics literature. Elliott (1953) and Gregory and Karney (1969) identified that the eigenvalues λ_i of S^{-1} , and (selected) corresponding eigenvectors P_i , for $i = 1, 2, \dots, n$, are given by $\lambda_i = 2 - 2 \cos(\frac{i\pi}{n+1})$ (in increasing order) and $P_{i,m} = \sin(\frac{im\pi}{n+1})$ respectively (where $m = 1, 2, \dots, n$). These results were later extended to more general tridiagonal matrices by Yueh (2005), see also Yueh and Cheng (2008) and Bunger (2014).

Because the covariance matrix σ^{cov} equals $\frac{1}{(n+1)}S$, the eigenvalues of σ^{cov} are equal to $\frac{1}{n+1}$ times those of S , so they are also $\frac{1}{n+1}$ times the inverses of the eigenvalues of S^{-1} .

Proposition 1 now follows from the fact that the eigenvalues of σ^{cov} are equal to $\frac{1}{n+1}$ times the reciprocals of the eigenvalues of S^{-1} , and Proposition 2 from the fact the matrices σ^{cov} , S , and S^{-1} have the same eigenvectors. This completes the proof. \square