Exploratory Multilevel Redundancy Analysis

Takane, Yoshio
McGill University, Department of Psychology
1205 Dr. Penfield Avenue
Montreal Quebec H3A 1B1, Canada
E-mail: takane@psych.mcgill.ca

Zhou, Lixing
McGill University, Department of Psychology
1205 Dr. Penfield Avenue
Montreal Quebec H3A 1B1, Canada
E-mail: lixing-zhou@hotmail.com

1. Introduction

Hierarchically structured data are commonly encountered in many fields of scientific investigation. In educational assessment studies, for example, students' performance in mathematics is measured in various schools. Such data are called hierarchically structured because students are nested within schools. Another example of hierarchically structured data arise in repeated measurement designs where some attributes of subjects are repeatedly measured over time.

Hierarchical (multilevel) linear models (HLM: Bock, 1989; Bryk and Raudenbush, 1992; Goldstein, 1987; Hox, 1995) are often used to analyze such data, explicitly taking into account the hierarchical nature of the data. Interpretations of parameters in such models, however, become increasingly more difficult as they accommodate more levels, more predictor variables, and more criterion variables. This paper presents a method of multilevel analysis with a dimension reduction feature to facilitate interpretations of model parameters. The proposed method is a multivariate extension of the procedure developed by Takane and Hunter (2002). The method first decomposes variability in the criterion variables into several orthogonal components using predictor variables at different levels, and then applies singular value decomposition (SVD) to the decomposed parts to find more parsimonious representations. An example is given to illustrate the method. Some possible extensions of the proposed method are also suggested.

2. The Model

For illustration, let us consider the following situation. (This situation closely resembles the example given later.) Suppose we are interested in assessing what attributes (factors) of students and their environments affect their performance in mathematics. For this, we measure students' performance in mathematics. We may use multiple tests to obtain reliable test scores and to capture all important aspects of math performance. Students (the first-level units) are usually nested within schools (the second-level units). We also collect relevant information about the students and schools they belong to.

We may ask several questions in this context: 1. How much of the students' math performance can or cannot be explained by school differences. The former is referred to as the between-school effects, and the latter as the within-school effects. 2. How much of the between-school effects can be explained by known school characteristics (the school-level predictor variables) and in what way do the school-level predictor variables affect student performance? 3. How much of the within-school effects can be explained by prescribed subject characteristics (the subject-level predictor variables), and in what way do the subject-level predictor variables affect student performance? 4. Are there any interactions between the school-level and student-level predictor variables that affect student performance? HLM allows us to investigate and answer all of these questions.

Let $\boldsymbol{Y}$ denote the N-subject by p-variable matrix of criterion variables. Let $\boldsymbol{1}_N$ denote the N-component vector of ones. (In this paper, we generally denote the a-component vector of ones by $\boldsymbol{1}_a$.) Each of N subjects is assumed to belong to one of J schools with N_j students in school j $(j = 1, \cdots, J)$. Let $\boldsymbol{G}$ denote the N by J matrix of dummy variables indicating which students belong to which schools. (The matrix $\boldsymbol{G}$ is a block diagonal matrix with $\boldsymbol{1}_{N_j}$ as the jth diagonal block.) Let

(1) $$\boldsymbol{W}_0 = [\boldsymbol{w}_1, \cdots, \boldsymbol{w}_J]'$$

represent the J by q-variable matrix of raw school-level predictor variables. We define $\boldsymbol{W}_0^*$, the matrix of columnwise centered school-level predictor variables, by

(2) $$\boldsymbol{W}_0^* = \boldsymbol{Q}_{1_J/G'G}\boldsymbol{W}_0,$$

where

(3) $$\boldsymbol{Q}_{1_J/G'G} = \boldsymbol{I} - \boldsymbol{1}_J(\boldsymbol{1}_J'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{1}_J)^{-1}\boldsymbol{1}_J'\boldsymbol{G}'\boldsymbol{G}$$

is the orthogonal projector onto the null space of $\boldsymbol{1}_J$ (often denoted as $\mathrm{Ker}(\boldsymbol{1}_J')$) in the metric of $\boldsymbol{G}'\boldsymbol{G}$.

We further let $\boldsymbol{D}_{X^*}$ denote a block diagonal matrix with $\boldsymbol{X}_j^*$ as the jth diagonal block, where $\boldsymbol{X}_j^*$ is the N_j by r-variable matrix of student-level predictor variables. (The matrix $\boldsymbol{D}_{X^*}$ represents interactions between schools and student-level predictor variables.) We assume that $\boldsymbol{X}_j^*$ is columnwise centered within school j. Let

(4) $$\boldsymbol{J}_r = \boldsymbol{1}_J \otimes \boldsymbol{I}_r,$$

where $\otimes$ indicates a Kronecker product, and define

(5) $$\boldsymbol{X}^* = \boldsymbol{D}_{X^*}\boldsymbol{J}_r = \begin{bmatrix} \boldsymbol{X}_1^* \\ \vdots \\ \boldsymbol{X}_J^* \end{bmatrix}.$$

Let

(6) $$\boldsymbol{W}_1 = [\boldsymbol{I}_r \otimes \boldsymbol{w}_1, \cdots, \boldsymbol{I}_r \otimes \boldsymbol{w}_J]'$$

and $\boldsymbol{W}_1^*$ by

(7) $$\boldsymbol{W}_1^* = \boldsymbol{Q}_{J_r/D_{XX}}\boldsymbol{W}_1,$$

where

(8) $$\boldsymbol{Q}_{J_r/D_{XX}} = \boldsymbol{J}_r(\boldsymbol{J}_r'\boldsymbol{D}_{XX}\boldsymbol{J}_r)^{-1}\boldsymbol{J}_r'\boldsymbol{D}_{XX}$$

is the orthogonal projector onto $\mathrm{Ker}(\boldsymbol{J}_r')$ in the metric of

(9) $$\boldsymbol{D}_{XX} = \boldsymbol{D}_{X^*}'\boldsymbol{D}_{X^*}.$$

The full model may then be stated as

(10) $$\boldsymbol{Y} = \boldsymbol{1}_N\boldsymbol{c}_{00}' + \boldsymbol{G}\boldsymbol{W}_0^*\boldsymbol{C}_{01} + \boldsymbol{G}\boldsymbol{Q}_{[1_J,W_0^*]/G'G}\boldsymbol{U}_0 + \boldsymbol{X}^*\boldsymbol{C}_{10} + \boldsymbol{D}_{X^*}\boldsymbol{W}_1^*\boldsymbol{C}_{11} + \boldsymbol{D}_{X^*}\boldsymbol{Q}_{[J_r,W_1^*]/D_{XX}}\boldsymbol{U}_1 + \boldsymbol{E},$$

where $\boldsymbol{c}_{00}'$, $\boldsymbol{C}_{01}$, $\boldsymbol{C}_{10}$, $\boldsymbol{C}_{11}$, $\boldsymbol{U}_0$ and $\boldsymbol{U}_1$ are matrices of regression parameters, $\boldsymbol{Q}_{[1_J,W_0^*]/G'G}$ is the orthogonal projector onto $\mathrm{Ker}([\boldsymbol{1}_J, \boldsymbol{W}_0^*]')$ in the metric of $\boldsymbol{G}'\boldsymbol{G}$, $\boldsymbol{Q}_{[J_r,W_1^*]/D_{XX}}$ is the orthogonal projector onto $\mathrm{Ker}([\boldsymbol{J}_r, \boldsymbol{W}_1^*]')$ in the metric of $\boldsymbol{D}_{XX}$, and $\boldsymbol{E}$ is a matrix of disturbance terms.

The first term in model (10) pertains to the grand means. The second term pertains to the portion of the between-school effects that can be explained by the school-level predictor variables, while the third term to the portion that cannot be explained by the school-level predictor variables. The fourth term represents the portion of the within-school effects that can be explained by the main effects of the student-level predictor variables. (The matrix $\boldsymbol{X}^*$ represents the main effects of the student-level predictor variables.) The fifth term represents the portion of the within-school effects that can be explained by the interactions between the school-level and student-level predictor variables. (The matrix $\boldsymbol{D}_{X^*}\boldsymbol{W}_1^*$ represents the interactions between the two.) The sixth term pertains to the portion of the interactions between schools and the student-level predictor variables that cannot be explained by the fourth and fifth terms. Finally, the last term in the model represents residuals left unaccounted for by any systematic effects in the model.

There are several important special cases of the full model presented above. When no school-level predictor variables exist, neither terms 2 and 3 nor terms 5 and 6 can be isolated. In this case, the model reduces to a simple analysis of covariance model. When no student-level predictor variables exist, terms 4, 5, 6, and 7 cannot be isolated. When neither the school-level nor student-level predictor variables exist, neither terms 2 and 3 nor terms 4, 5, 6, and 7 can be isolated. In this case, we simply have a one-way ANOVA model.

3. Estimation

The seven terms in model (10) are all columnwise orthogonal and so coefficients in each term can be separately estimated by OLS (Ordinary Least Squares). We thus have

$$\text{(11)} \qquad \hat{\boldsymbol{c}}_{00} = \boldsymbol{1}_N'\boldsymbol{Y}/N,$$

$$\text{(12)} \qquad \hat{\boldsymbol{C}}_{01} = (\boldsymbol{W}_0^{*\prime}\boldsymbol{G}'\boldsymbol{G}\boldsymbol{W}_0^*)^{-1}\boldsymbol{W}_0^{*\prime}\boldsymbol{G}'\boldsymbol{Y},$$

$$\text{(13)} \qquad \hat{\boldsymbol{U}}_0 = (\boldsymbol{Q}_{[1_J,W_0^*]/G'G}'\boldsymbol{G}'\boldsymbol{G}\boldsymbol{Q}_{[1_J,W_0^*]/G'G})^{-1}\boldsymbol{Q}_{[1_J,W_0^*]/G'G}'\boldsymbol{G}'\boldsymbol{Y},$$

$$\text{(14)} \qquad \hat{\boldsymbol{C}}_{10} = (\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{Y},$$

$$\text{(15)} \qquad \hat{\boldsymbol{C}}_{11} = (\boldsymbol{W}_1^{*\prime}\boldsymbol{D}_{XX}\boldsymbol{W}_1^*)^{-1}\boldsymbol{W}_1^{*\prime}\boldsymbol{D}_{X^*}'\boldsymbol{Y},$$

$$\text{(16)} \qquad \hat{\boldsymbol{U}}_1 = (\boldsymbol{Q}_{[J_r,W_1^*]/D_{XX}}'\boldsymbol{D}_{XX}\boldsymbol{Q}_{[J_r,W_1^*]/D_{XX}})^{-1}\boldsymbol{Q}_{[J_r,W_1^*]/D_{XX}}'\boldsymbol{D}_{X^*}'\boldsymbol{Y},$$

and finally,

$$\text{(17)} \qquad \hat{\boldsymbol{E}} = \boldsymbol{Q}_{D_{X^*}}\boldsymbol{Q}_G\boldsymbol{Y}, \ \text{ where } \ \boldsymbol{Q}_G = \boldsymbol{I} - \boldsymbol{G}(\boldsymbol{G}'\boldsymbol{G})^{-1}\boldsymbol{G}'.$$

The orthogonality of the three leading terms and the last four terms in (10) may be seen by noting that the former all pertain to subspaces of $\text{Sp}(\boldsymbol{G})$ (note that $\boldsymbol{1}_N = \boldsymbol{G}\boldsymbol{1}_J$), while the latter to $\text{Ker}(\boldsymbol{G}') = \text{Sp}(\boldsymbol{Q}_G)$. (Note in the latter that $\boldsymbol{Q}_G\boldsymbol{D}_{X^*} = \boldsymbol{D}_{X^*}$, and that $\boldsymbol{Q}_{D_{X^*}}\boldsymbol{Q}_G = \boldsymbol{Q}_G\boldsymbol{Q}_{D_{X^*}}$.) The orthogonality among the first three terms may be seen by noting that

$$\text{(18)} \qquad \boldsymbol{G}\boldsymbol{W}_0^* = \boldsymbol{G}\boldsymbol{Q}_{1_J/G'G}\boldsymbol{W}_0 = \boldsymbol{Q}_{1_N}\boldsymbol{G}\boldsymbol{W}_0,$$

where $\boldsymbol{Q}_{1_N} = \boldsymbol{I} - \boldsymbol{1}_N\boldsymbol{1}_N'/N$, and that

$$\text{(19)} \qquad \boldsymbol{G}\boldsymbol{Q}_{[1_J,W_0^*]/G'G} = \boldsymbol{Q}_{[1_N,GW_0^*]}\boldsymbol{G},$$

where $\boldsymbol{Q}_{[1_N,GW_0^*]}$ is the orthogonal projector onto $\text{Ker}([\boldsymbol{1}_N, \boldsymbol{G}\boldsymbol{W}_0^*]')$. The orthogonality among the last four terms may be seen by noting

$$\text{(20)} \qquad \boldsymbol{D}_{X^*}\boldsymbol{W}_1^* = \boldsymbol{D}_{X^*}\boldsymbol{Q}_{J_r/D_{XX}}\boldsymbol{W}_1 = \boldsymbol{Q}_{X^*}\boldsymbol{D}_{X^*}\boldsymbol{W}_1,$$

where $\boldsymbol{Q}_{X^*} = \boldsymbol{I} - \boldsymbol{X}^*(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}$ is the orthogonal projector onto $\text{Ker}(\boldsymbol{X}^{*\prime})$, that

$$\text{(21)} \qquad \boldsymbol{D}_{X^*}\boldsymbol{Q}_{[J_r,W_1^*]/D_{XX}} = \boldsymbol{Q}_{[X^*,D_{X^*}W_1^*]}\boldsymbol{D}_{X^*},$$

where $\boldsymbol{Q}_{[X^*,D_{X^*}W_1^*]}$ is the orthogonal projector onto $\text{Ker}([\boldsymbol{X}^*, \boldsymbol{D}_{X^*}\boldsymbol{W}_1^*]')$, and that $\text{Sp}(\boldsymbol{Q}_{D_{X^*}}) = \text{Ker}(\boldsymbol{D}_{X^*}')$.

Putting the estimates of parameters given above into (10), we obtain the following (orthogonal) decomposition of $\boldsymbol{Y}$:

$$\begin{aligned}(22)\qquad \boldsymbol{Y} = \boldsymbol{P}_{1_N}\boldsymbol{Y} + \boldsymbol{P}_{GW_0^*}\boldsymbol{Y} + \boldsymbol{P}_{GA^*}\boldsymbol{Y} \\ + \boldsymbol{P}_{X^*}\boldsymbol{Y} + \boldsymbol{P}_{D_{X^*}W_1^*}\boldsymbol{Y} + \boldsymbol{P}_{D_{X^*}B^*}\boldsymbol{Y} + \boldsymbol{Q}_{D_{X^*}}\boldsymbol{Q}_G\boldsymbol{Y},\end{aligned}$$

where in general $\boldsymbol{P}_Z = \boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'$ is the orthogonal projector onto $\mathrm{Sp}(\boldsymbol{Z})$, $\boldsymbol{A}^* = \boldsymbol{Q}_{[1_J,W_0^*]/G'G}$, and $\boldsymbol{B}^* = \boldsymbol{Q}_{[J_r,W_1^*]/D_{XX}}$. This decomposition of $\boldsymbol{Y}$ entails a more generic decomposition of E^N, the N-dimensional Euclidean space, which is split into the orthogonal direct-sum of the seven subspaces spanned by the orthogonal projectors preceding $\boldsymbol{Y}$'s in (22). This generic decomposition is depicted in the following table.

Table 1. The decomposition of $E^N = $ ***Sp***$(\boldsymbol{I}_N)$.

<table>
<tr><td>(1) $\boldsymbol{P}_{1_N}$</td><td>(2) $\boldsymbol{P}_{GW_0^*}$</td><td>(3) $\boldsymbol{P}_{GA^*}$</td><td>(4) $\boldsymbol{P}_{X^*}$</td><td>(5) $\boldsymbol{P}_{D_{X^*}W_1^*}$</td><td>(6) $\boldsymbol{P}_{D_{X^*}B^*}$</td><td>(7) $\boldsymbol{Q}_{D_{X^*}}\boldsymbol{Q}_G$</td></tr>
<tr><td></td><td colspan="2">(8) $\boldsymbol{P}_G\boldsymbol{Q}_{1_N}$</td><td></td><td colspan="2">(9) $\boldsymbol{P}_{D_{X^*}Q_{J_r/D_{XX}}}$</td><td></td></tr>
<tr><td></td><td colspan="2"></td><td colspan="3">(10) $\boldsymbol{P}_{D_{X^*}}$</td><td></td></tr>
<tr><td></td><td colspan="2"></td><td colspan="4">(11) $\boldsymbol{Q}_G$</td></tr>
<tr><td></td><td colspan="6">(12) $\boldsymbol{Q}_{1_N}$</td></tr>
<tr><td colspan="7">(13) $\boldsymbol{I}_N$</td></tr>
</table>

In the table, the symbol Sp(.) is omitted around the projectors to avoid clutters in notation.

The table shows six binary decompositions that together lead to the seven-term decomposition of E^N: (13) = (1) + (12), (12) = (8) + (11), (8) = (2) + (3), (11) = (7) + (10), (10) = (4) + (9), and (9) = (5) + (6), where the parenthesized numbers indicate the subspaces spanned by the projectors indicated by those numbers, and the plus sign indicates the orthogonal direct-sum of two subspaces. Of course, the orthogonality of subspaces implies the additivity of projectors associated with them.

Due to the orthogonality of the terms in the model, the sum of squares of the elements of $\boldsymbol{Y}$ ($\mathrm{SS}(\boldsymbol{Y}) = \mathrm{tr}(\boldsymbol{Y}'\boldsymbol{Y})$) can also be uniquely decomposed into the sum of termwise sums of squares. Let SS_i denote the sum of squares of the elements of the ith term in (22). The SS_i ($i = 1, \cdots, 7$) indicates the size of the effects of the ith term. In addition, we define

$$(23)\qquad \mathrm{SS}_T = \mathrm{SS}(\boldsymbol{Y}) - \mathrm{SS}_1,$$

$$(24)\qquad \mathrm{SS}_B = \mathrm{SS}_2 + \mathrm{SS}_3,$$

and

$$(25)\qquad \mathrm{SS}_W = \mathrm{SS}_4 + \mathrm{SS}_5 + \mathrm{SS}_6 + \mathrm{SS}_7.$$

The SS_T, SS_B, and SS_W stand for the total SS, between-group (school) SS, within-group (school) SS, respectively, in analogy to the univariate ANOVA situation. It holds that $\mathrm{SS}_T = \mathrm{SS}_B + \mathrm{SS}_W$.

4. Dimension Reduction

The columns of $\boldsymbol{Y}$ are often highly correlated and their variabilities can be more succinctly summarized by fewer components of $\boldsymbol{Y}$ than the number of variables in $\boldsymbol{Y}$. Each of the estimated terms in the model (except term 1, which is a rank-one matrix and no further rank reduction is possible) may be subjected to rank reduction in line with the spirit of redundancy analysis (Van den Wollenberg, 1977). Again, due to the orthogonality of the estimated terms in the model, this can be done separately for each term. (It is possible to recombine some of the terms in the decomposition into one, which is then subjected to a rank reduction.) Each rank reduction can be carried out by singular value decomposition (SVD) of the relevant term, or generalized SVD of the estimated regression coefficients related to that term under a special nonidentity metric matrix. The former is denoted as

SVD($\boldsymbol{GW}_0^*\hat{\boldsymbol{C}}_{01}$) for the second term in (22), for example, while the latter by GSVD$(\hat{\boldsymbol{C}}_{01})_{W_0^{*\prime}G'GW_0^*}$, where the subscript on GSVD indicates the row side (left-hand side) metric matrix.

These two SVDs are related in a simple manner. To show this relationship, however, we have to know how we find GSVD$(\hat{\boldsymbol{C}}_{01})_{W_0^{*\prime}G'GW_0^*}$. We first obtain SVD$((\boldsymbol{W}_0^{*\prime}\boldsymbol{G}'\boldsymbol{GW}_0^*)^{1/2}\hat{\boldsymbol{C}}_{01})$. Let this SVD be denoted as

$$(26) \qquad (\boldsymbol{W}_0^{*\prime}\boldsymbol{G}'\boldsymbol{GW}_0^*)^{1/2}\hat{\boldsymbol{C}}_{01} = \boldsymbol{H}^*\boldsymbol{\Delta}^*\boldsymbol{K}^{*\prime},$$

where $\boldsymbol{H}^*$ is the matrix of left singular vectors, $\boldsymbol{K}^*$ the matrix of right singular vectors, and $\boldsymbol{\Delta}$ the positive-definite diagonal matrix of singular values. Let GSVD$(\hat{\boldsymbol{C}}_{01})_{W_0^{*\prime}G'GW_0^*}$ be denoted as

$$(27) \qquad \hat{\boldsymbol{C}}_{01} = \boldsymbol{H}\boldsymbol{\Delta}\boldsymbol{K}'.$$

Then, $\boldsymbol{H}$, $\boldsymbol{K}$, and $\boldsymbol{\Delta}$ in (27) can be obtained from (26) by

$$(28) \qquad \boldsymbol{H} = (\boldsymbol{W}_0^{*\prime}\boldsymbol{G}'\boldsymbol{GW}_0^*)^{-1/2}\boldsymbol{H}^*,$$

$\boldsymbol{K} = \boldsymbol{K}^*$, and $\boldsymbol{\Delta} = \boldsymbol{\Delta}^*$. Note that $\boldsymbol{H}'\boldsymbol{W}_0^{*\prime}\boldsymbol{G}'\boldsymbol{GW}_0^*\boldsymbol{H} = \boldsymbol{I}$. (That is, $\boldsymbol{H}$ is orthogonal with respect to the metric matrix $\boldsymbol{W}_0^{*\prime}\boldsymbol{G}'\boldsymbol{GW}_0^*$.)

Let SVD($\boldsymbol{GW}_0^*\hat{\boldsymbol{C}}_{01}$) be denoted as

$$(29) \qquad \boldsymbol{GW}_0^*\hat{\boldsymbol{C}}_{01} = \tilde{\boldsymbol{H}}\tilde{\boldsymbol{\Delta}}\tilde{\boldsymbol{K}}'.$$

Then,

$$(30) \qquad \tilde{\boldsymbol{H}} = \boldsymbol{GW}_0^*\boldsymbol{H},$$

$\tilde{\boldsymbol{K}} = \boldsymbol{K}$, and $\tilde{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$. Thus, (29) is essentially equivalent to (27). (One can be derived from the other.) The former, however, has some computational advantage. In general, the matrix whose SVD is computed in (27) is much smaller in size than that in (29). Although this has been shown only for the second term in (22), essentially the same holds for all other terms.

We employ the following two quantities (both of which are simple derivatives of the quantities obtained in the GSVD or SVD above) in the example to be reported in the next section. One is the weight matrix $\boldsymbol{T}$ applied to predictor variables, e.g., $\boldsymbol{GW}_0^*$, to obtain the matrix $\boldsymbol{F}$ of component scores. This can be obtained by simple scaling of the matrix $\boldsymbol{H}$ in (27) by $\sqrt{N}$, namely $\boldsymbol{T} = \sqrt{N}\boldsymbol{H}$, so that $\boldsymbol{F} = \boldsymbol{GW}_0^*(\sqrt{N}\boldsymbol{H})$. The other is the covariance matrix $\boldsymbol{V}$ between component scores $\boldsymbol{F}$ and observed scores on the criterion variables $\boldsymbol{Y}$. This matrix is often called the matrix of component (criterion) loadings and is obtained by $\boldsymbol{V} = \boldsymbol{F}'\boldsymbol{Y}/N$.

The sum of squares explained by each component is indicated by the squared singular values, whose sum over all components is equal to the SS of that term. This fact may be exploited to calculate the percentage SS contributed by each component.

5. An Example

In this section, we demonstrate the use of our proposed method. We utilize part of the data from the Education Longitudinal Study of 2002 (ELS: 2002). This study measured students' achievements in mathematics and also gathered information about students' attitudes. In addition, data were collected from their teachers regarding their views of students as well as their own credentials and educational background (Ingels, Planty, Bozick, and Owings, 2005).

The survey was conducted in the spring term of 2002. The subjects were high school sophomores (10th graders). Schools were first selected, and then students were sampled randomly within schools. Students took cognitive tests on reading and mathematics, and questionnaires were administered to parents, Math and English teachers, school principals, and heads of the school library media center. There was a total of 750 schools and over 15,000 students and their parents in the original data set, with one Mathematics teacher and one English teacher for each student.

Criterion referenced NELS-equated proficiency scores were calculated in the form of probabilities based on a cluster of items that mark certain proficiency levels. There are five levels of proficiency in math which are hierarchically ordered in the sense that mastery of a higher level typically implies proficiency at the lower levels. The NELS-equated proficiency probabilities were computed using IRT-estimated item parameters calibrated in NELS: 88. We use the five proficiency probabilities as our criterion variables. Each proficiency probability represents the probability that a student would pass a given proficiency level. Proficiency at level 1 corresponds to simple arithmetical operations on whole numbers. Level 2 pertains to simple operations with decimals, fractions, powers, and roots. Level 3 represents simple problem solving, requiring the understanding of low level mathematical concepts. Level 4 pertains to understanding of intermediate level mathematical concepts and/or multi-step solutions to word problems. Level 5 concerns complex multi-step word problems and/or advanced mathematics material.

We eliminated students with missing data from our analysis. We also eliminated schools with fewer than 20 students in the data set. This left us with $N = 10,939$ students nested within $J = 562$ schools. The school-level predictor variables used are given in Table 3. Each of the statements was rated on a 5-point scale with respect to how accurate the statement was as a description of the school (1. not accurate at all to 5. very accurate). The student-level predictor variables used are shown in Table 4. There were three categorical variables. They were coded into 8 dummy variables altogether prior to the analysis.

Table 2 gives a breakdown of the total SS (SS_T) explained by the different terms in model (10).

Table 2. A breakdown of the total SS.

Between-School SS (SS_B) 17.9%		Within-School SS (SS_W) 82.1%			
SS_2	SS_3	SS_4	SS_5	SS_6	SS_7
2.7%	15.2%	2.3%	0.2%	19.7%	60.0%

Only 18% of SS_T can be explained by the differences between schools. The remaining 82% is left unaccounted for by the school differences. Of SS_B, SS_2 represents the SS that can be accounted for by the main effects of school-level predictor variables. This amounts to only 2.6% of SS_T. The SS_3, the remaining part of SS_B, represents the SS that cannot be accounted for by the school-level predictor variables. A majority of SS_B belongs to this category. Of SS_W, SS_4 represents the SS that can be accounted for by the main effects of student-level predictor variables. This amounts to only 2.3% of SS_T. The SS_5 represents the SS that can be accounted for by the interactions between school-level and student-level predictor variables. Only 0.2% of SS_T can be accounted for by the interactions. The SS_6 represents the portions of interactions between schools and student-level predictor variables that cannot be explained by the previous two (SS_4 and SS_5). This accounts for approximately 20% of SS_T. The SS_7 indicates the residual SS representing the SS that is not predictable by any systematic effects in the model. A majority (60%) of SS_T is left unaccounted for by any systematic effects.

Although the predictor variables at either level account for only small proportions of SS_T, their effects are deemed fairly stable because of the large sample size (N over 10,000) in this data set. (We analyzed parts of the data prior to the analysis of the entire data set and found consistent results in all cases.) So we next look at them a little more closely. For the effects of of school-level predictor variables, we estimated $\boldsymbol{C}_{01}$ in the second term in model (10) and then applied GSVD. Singular values were found to be .36, .05, and .03. So the first component accounted for almost 99% of the SS_2. Table 3 reports the estimated component weights. Teachers pressing students to achieve and students being expected to do homework seem to have positive effects on students' performance. (Recall, however, SS_2 represents only a small portion of SS_T.) Teacher's morale being high, on the other hand, has only a negligible effect. (Covariances between this component and the criterion variables were .14, .19, .20, .17, .08, so this component seems to represent students' overall performance in mathematics.

Table 3. The effects of school-level predictor variables.

Variables	Component weights ($\boldsymbol{T}$)
Teachers press students to achieve	.69
Teachers' morale is high	-.00
Students expected to do homework	.42

We next look at the effects of student-level predictor variables, which are summarized in Table 4. We estimated $\boldsymbol{C}_{10}$ in (10) and then applied GSVD. Singular values were found to be .33, .08, .01, .01, and .00, so the first component again accounted for a majority (over 97%) of the SS_4. It may be observed that male students did slightly better than female students. There are larger race differences among the three racial groups. White students performed better than black and Hispanic students. (Again, recall that only 2.3% of the SS_T can be accounted for by SS_4.) Contrary to people's common sense, hours spent on homework had relatively small effects on students' mathematical proficiency. A moderate amount of time spent on homework has a small positive effect, while no hours or too many hours have small negative effects. Covariances between this component and the criterion variables (component loadings) were .12, .18, .19, .15, .04, so again the component seems to represent students' overall performance in mathematics.

Table 4. The effects of student-level predictors.

Variables	Categories	Component weights ($\boldsymbol{T}$)
1.Gender	male	.28
	female	-.28
2. Race	Black	-1.90
	Hispanic	.18
	White	1.71
3. Homework	0 hours	-.13
	1-4 hours	.28
	5 or more hours	-.15

6. Discussion and Future Work

In this paper, we proposed a method for multilevel redundancy analysis. This method is particularly attractive since OLS estimates of regression parameters can be obtained in closed form. The estimated regression parameters are then subjected to rank reduction by GSVD. Reduced-rank approximations of regression parameters are useful, particularly when the dimensionality of the parameter space is high. An application of the proposed method was empirically demonstrated through a real example.

There are a number of possible extensions that can make the proposed method even more useful:

1. Although only the two-level model has been discussed in this paper, similar methods can be developed for higher-level multivariate data. The number of terms in the model, however, grows very quickly. For example, a full three-level HLM with predictor variables at all levels, there are 15 terms altogether.

2. Bootstrap (e.g., Efron, 1982) or other resampling techniques could be used to assess the stability of individual parameters, which may in turn be used to test their significance. Since the normality assumption is almost always in suspect in survey data, the bootstrap methods may also be useful to benchmark the distribution of the conventional statistics used in HLM.

3. The number of components to be retained in dimension reduction may be determined by permutation tests in a manner similar to Takane and Hwang (2002), who developed a permutation procedure

for testing the number of significant canonical correlations.

4. Additional (linear) constraints can be readily incorporated in the OLS estimation procedure. This allows the statistical tests of the hypotheses represented by the constraints.

5. When the $\boldsymbol{U}$ parameters are assumed to be random rather than fixed, observations obtained from subjects in the same schools are no longer statistically independent. The dependence structure among the observations may be estimated from the initial estimates of parameters (obtained under the independence assumption), which may then be used to re-estimate the parameters, and so on. This leads to an iterative estimation procedure for full maximum likelihood estimation (MLE) of parameters (Goldstein, 1987). A simpler method called REML (REstricted Maximum Likelihood: e.g., LaMotte, 2007) may also be of interest in this context.

6. The ridge type of regularized LS (RLS) estimation may be used instead of OLS. The RLS is easy to apply and is known to provide estimates of regression parameters which are on average closer to population parameters (Takane and Hwang, 2007; Takane and Jung, 2008).

7. Interesting special cases arise when we set $\boldsymbol{Y} = \boldsymbol{Q}_{1_N}\boldsymbol{X} = \tilde{\boldsymbol{X}}$ and or $\boldsymbol{Y} = \boldsymbol{D}_{X^*}$. The former leads to

$$\tilde{\boldsymbol{X}} = \boldsymbol{P}_G\tilde{X} + \boldsymbol{Q}_G\tilde{\boldsymbol{X}} = \boldsymbol{P}_{GW_0^*}\tilde{\boldsymbol{X}} + \boldsymbol{P}_{GA^*}\tilde{\boldsymbol{X}} + \boldsymbol{Q}_G\tilde{\boldsymbol{X}}, \tag{31}$$

and the latter to

$$\boldsymbol{D}_{X^*} = \boldsymbol{P}_{X^*}\boldsymbol{D}_{X^*} + \boldsymbol{P}_{D_{X^*}Q_{J_r/D_{XX}}}\boldsymbol{D}_{X^*} = \boldsymbol{P}_{X^*}\boldsymbol{D}_{X^*} + \boldsymbol{P}_{D_{X^*}W_1^*}\boldsymbol{D}_{X^*} + \boldsymbol{P}_{D_{X^*}B^*}\boldsymbol{D}_{X^*}, \tag{32}$$

where $\boldsymbol{A}^*$ and $\boldsymbol{B}^*$ were introduced shortly after (22) above Table 1. The SVD of terms in these decompositions may be called multilevel PCAs (Principal Component Analyses).

7. References

Bock, R. D. (1989). *Multilevel Analysis of Educational Data.* San Diego, CA: Academic Press.

Bryk, A. S., and Raudenbush, S. W. (1992). *Hierarchical Linear Models.* Newbury Park, CA: Sage Publications.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia: SIAM.

Goldstein, H. I. (1987) *Multilevel Models in Educational and Social Research.* London: Oxford University Press.

Hox, J. J. (1995). *Applied Multilevel Analysis.* Amsterdam: TT-Publikaties.

Ingels, S. J., Planty, M., and Bozick, R. (2005). *A Profile of the American High School Senior in 2004: A First Look-Initial Results from the First Follow-up of the Education Longitudinal Study of 2002 (ELS:2002) (NCES 2006348).* U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

LaMotte, L. R. (2007). A direct derivation of the REML likelihood function. *Statistical Papers*, **48**, 321–327.

Takane, Y., and Hunter, M. A. (2002). Dimension reduction in hierarchical linear models. In S. Nishisato, Y. Baba, H. Bozdogan, and K. Kanefuji (Eds.), *Measurement and Multivariate Analysis* (pp. 145–154). Tokyo: Springer Verlag.

Takane, Y., and Hwang, H. (2002). Generalized constrained canonical correlation analysis. *Multivariate Behavioral Research*, **37**, 163–195.

Takane, Y., and Hwang, H. (2007). Regularized linear and kernel redundancy analysis. *Computational Statistics and Data Analysis*, **52**, 394–405.

Takane, Y., and Jung, S. (2008). Regularized partial and/or constrained redundancy analysis. *Psychometrika*, **73**, 671–690.

Van den Wollenberg, A. L. (1977). Redundancy analysis: An alternative for canonical correlation analysis. *Psychometrika*, **42**, 207–219.