VisualGSCA 1.0 – A Graphical User Interface Software Program for Generalized Structured Component Analysis

Heungsun Hwang
Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, H3A 1B1, Canada

Abstract

Generalized Structured Component Analysis (GSCA) has been proposed for structural equation modeling. The method was recently implemented into a software program. *VisualGSCA 1.0* provides a graphical user interface whereby path analytic models are easily drawn in the program window and their GSCA parameter estimates are subsequently displayed in the same window. A brief introduction to GSCA is first presented. Next, the use of *VisualGSCA 1.0* is demonstrated step by step through an empirical application.

1. Introduction

VisualGSCA 1.0 (Hwang, 2007) is a software program that implements the estimation procedure of Generalized Structured Component Analysis (GSCA) (Hwang & Takane, 2004). The software program is downloadable free of charge from the following site http://www.psych.mcgill.ca/perpg/fac/hwang/software.html.

For purposes of Structural Equation Modeling (SEM), GSCA represents an alternative to Partial Least Squares (PLS) (Lohmöller, 1989; Wold, 1975). Like PLS, GSCA can be viewed as a component-based approach to SEM in the sense that latent variables are defined as components or weighted composites of observed variables as follows:

$$\boldsymbol{\gamma}_i = \mathbf{W}\mathbf{z}_i, \tag{1}$$

where $\mathbf{z}_i$ and $\boldsymbol{\gamma}_i$ are a J by 1 vector of observed variables/indicators and a d by 1 vector of latent variables/constructs for an observation i (i = 1, ⋯, N), respectively, and $\mathbf{W}$ is a d by J matrix consisting of component weights. GSCA also involves a measurement model that specifies the relationships between indicators and constructs, and a structural model that expresses the relationships among constructs. Specifically, in GSCA, the measurement model is given by

$$\mathbf{z}_i = \mathbf{C}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, \tag{2}$$

where $\mathbf{C}$ is a J by d matrix of loadings, and $\boldsymbol{\varepsilon}_i$ is a J by 1 vector of residuals for $\mathbf{z}_i$.

The structural model is defined by

$$\boldsymbol{\gamma}_i = \mathbf{B}\boldsymbol{\gamma}_i + \boldsymbol{\xi}_i, \tag{3}$$

where $\mathbf{B}$ is a d by d matrix of path coefficients, and $\boldsymbol{\xi}_i$ is a d by 1 vector of residuals

This paper was submitted as a book chapter for "New Trends in Psychometrics" (Tokyo: Universal Academy Press).

for $\boldsymbol{\gamma}_i$.

Then, GSCA integrates the above three equations into a single one as follows:

$$\begin{bmatrix}\mathbf{z}_i\\ \boldsymbol{\gamma}_i\end{bmatrix}=\begin{bmatrix}\mathbf{C}\\ \mathbf{B}\end{bmatrix}\boldsymbol{\gamma}_i+\begin{bmatrix}\boldsymbol{\varepsilon}_i\\ \boldsymbol{\xi}_i\end{bmatrix}$$
$$\begin{bmatrix}\mathbf{I}\\ \mathbf{W}\end{bmatrix}\mathbf{z}_i=\begin{bmatrix}\mathbf{0} & \mathbf{C}\\ \mathbf{0} & \mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{I}\\ \mathbf{W}\end{bmatrix}\mathbf{z}_i+\begin{bmatrix}\boldsymbol{\varepsilon}_i\\ \boldsymbol{\xi}_i\end{bmatrix} \quad (4)$$
$$\mathbf{u}_i=\mathbf{A}\mathbf{u}_i+\mathbf{e}_i$$

where $\mathbf{I}_J$ is an identity matrix of order J, $\mathbf{u}_i=\begin{bmatrix}\mathbf{I}\\ \mathbf{W}\end{bmatrix}\mathbf{z}_i$, $\mathbf{A}=\begin{bmatrix}\mathbf{0} & \mathbf{C}\\ \mathbf{0} & \mathbf{B}\end{bmatrix}$, and $\mathbf{e}_i=\begin{bmatrix}\boldsymbol{\varepsilon}_i\\ \boldsymbol{\xi}_i\end{bmatrix}$. As shown above, in the GSCA model, all indicators and constructs are included in $\mathbf{u}_i$ and their inter-dependencies are expressed by **A**. Similarly, PLS also entails the specifications of measurement and structural models. However, it does not combine the two models into a unified algebraic formulation such as (4). Instead, it addresses them separately.

The parameters of GSCA (**W** and **A**) are estimated so that the sum of squares of all residuals ($\mathbf{e}_i$) is as small as possible across observations. This is equivalent to minimizing the following least-squares criterion:

$$\phi=\sum_{i=1}^{N}(\mathbf{u}_i-\mathbf{A}\mathbf{u}_i)'(\mathbf{u}_i-\mathbf{A}\mathbf{u}_i), \quad (5)$$

with respect to **W** and **A** under an identification constraint $\sum_{i=1}^{N}\gamma_{ik}^2=1$, where γ_{ik} is the *k*-th element of $\boldsymbol{\gamma}_i$. An Alternating Least Squares (ALS) algorithm (de Leeuw, Young & Takane, 1976) was developed to minimize (5). The ALS algorithm repeats two main steps until convergence: In the first step, for fixed **W**, **A** is updated in the least-squares sense; and in the second, **W** is updated in the least-squares sense for fixed **A** (for a detailed description of the ALS algorithm, refer to Hwang & Takane, 2004; Hwang et al., 2007).

GSCA provides an overall measure of model fit, called *FIT*, given by $FIT=1-\left[\phi/\sum_{i=1}^{N}\mathbf{u}_i'\mathbf{u}_i\right]$. The values of FIT range from 0 to 1. The larger the value of *FIT*, the more variance in the data is accounted for by the specified model. However, FIT is affected by model complexity. Thus, another index of fit, called *AFIT*, was developed which takes into account this complexity: $AFIT = 1-(1-\text{FIT})d_0/d_1$, where d_0 =

NJ is the degrees of freedom for the null model ($\mathbf{W} = \mathbf{0}$ and $\mathbf{A} = \mathbf{0}$) and $d_1 = NJ - P$ is the degrees of freedom for the model being tested, where P is the number of free parameters. Furthermore, *VisualGSCA1.0* offers two additional measures of overall model fit: (1) (unweighted least-squares) GFI (Jöreskog & Sörbom, 1986) and (2) SRMR (standardized root mean square residual). Both are proportional to the difference between the sample covariances and the covariances reproduced by the parameter estimates of GSCA. The reproduced covariance matrix is derived by $\mathbf{G}(\mathbf{I}-\mathbf{A})^{-1}\boldsymbol{\Sigma}_e(\mathbf{I}-\mathbf{A})^{-1\prime}\mathbf{G}'$, where $\mathbf{G}=[\mathbf{I}_J,\mathbf{0}]$ and $E(\mathbf{e}_i\mathbf{e}_i')=\boldsymbol{\Sigma}_e$. The GFI values close to 1 and the SRMR values close to 0 may be taken as indicative of good fit (e.g., Hu & Bentler, 1999).

2. Basic features of *VisualGSCA 1.0*

VisualGSCA 1.0 provides a Graphical User Interface (GUI) which allows users to easily express their model as a path diagram and to view the GSCA estimates of model parameters in the same diagram. The main codes for GSCA estimation were written in MATLAB while the GUI of the program was written in C++. The design of the current GUI was developed with reference to that of VisualPLS (Fu, 2006). Specifically, *VisualGSCA 1.0* enables users to:

- Directly draw a path diagram in the window of the program.
- Specify formative and reflective indicators in a straightforward manner.
- Use the bootstrap method (Efron, 1982) to estimate the standard errors of parameter estimates.
- Apply Covariance Structure Analysis (CSA) (Jöreskog, 1973) to fit the same model.

3. Demonstration of *VisualGSCA 1.0*

In this section, we demonstrate how to run the program to fit a specified model to data. For this purpose, we use part of the organization identification data (Bergami & Bagozzi, 2000) which were also analyzed in Hwang and Takane (2004). The number of observations is equal to 305. The model specified for these data is displayed in Figure 1. (No residual terms are displayed in the figure). As shown in the figure, this model consists of four constructs and 21 reflective indicators: Organizational Prestige (Org_Pres) is measured by 8 indicators (cei1 – cei8), Organizational Identification (Org_Iden) by 6 indicators (ma1 – ma6), Affective Commitment-Joy (AC_Joy) by 4 indicators (orgcmt1, 2, 3 and 7), and Affective Commitment – Love (AC_Love) by 3 indicators (orgcmt5, 6, and 8). The three indicators for AC_Love were reverse coded.

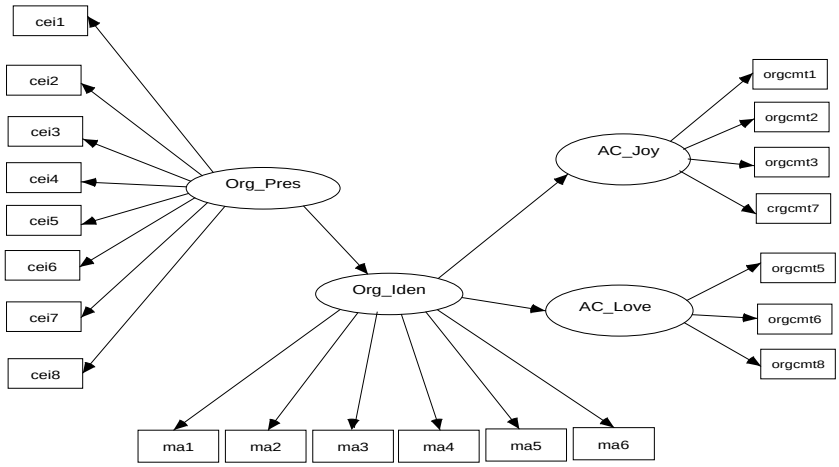

Figure 1. The specified structural equation model for the example data.

3.1. Preparation of a raw date file

VisualGSCA 1.0 is run on individual-level raw data. The raw data file is to be prepared in text format (.txt or .dat). The specific data format required for *VisualGSCA 1.0* is as follows:

- The first row contains the names of indicators (maximum 8 characters). The name of each indicator needs to be separated by a space.
- The data input begins on the second row. Data from an observation, or responses by an individual on each indicator, need to be separated by a space.
- Data for each observation should appear on a single row.

The following shows the first five rows of the data set created for the present example:

cei1	cei2	cei3	cei4	cei5	cei6	cei7	cei8	ma1	ma2	ma3	ma4	ma5	ma6	orgcmt1	orgcmt2	orgcmt3	orgcmt7	orgcmt5	orgcmt6	orgcmt8
3	3	3	4	3	3	3	3	4	4	4	4	3	4	3	2	3	4	3	3	3
4	5	3	4	4	3	3	3	2	2	3	3	2	3	3	1	3	3	3	3	3
5	5	4	4	5	4	4	5	3	2	4	3	2	4	1	3	3	4	4	3	2
4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	3	2	3	3	3	3

3.2. Creation of a new project

Users can create a new GSCA project by clicking on the [New Model] icon 

. Then, the *Configuration* dialog window will pop up. An exemplar of this window is displayed in Figure 2. Users are then prompted to:

- Type the name of their new project in *Name of Project*. The default extension of the project file is .vg (e.g., demo.vg).
- Assign an input data file to *Raw Data File* by clicking the [Open] button. The data file will be displayed in a separate window. Click on [OK] to close the data file window, and again click on [OK] to close the Configuration dialog window.

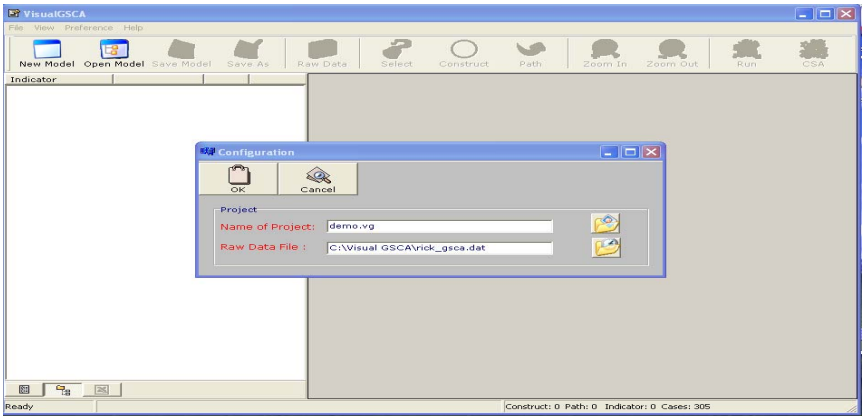

Figure 2. The Configuration dialog window for creating a new project.

3.3. Model Specification

The list of indicators in the input data file will subsequently appear in the left window of the main program under the label *Indicator* (see Figure 3). Users can then specify the structural equation model with the following steps.

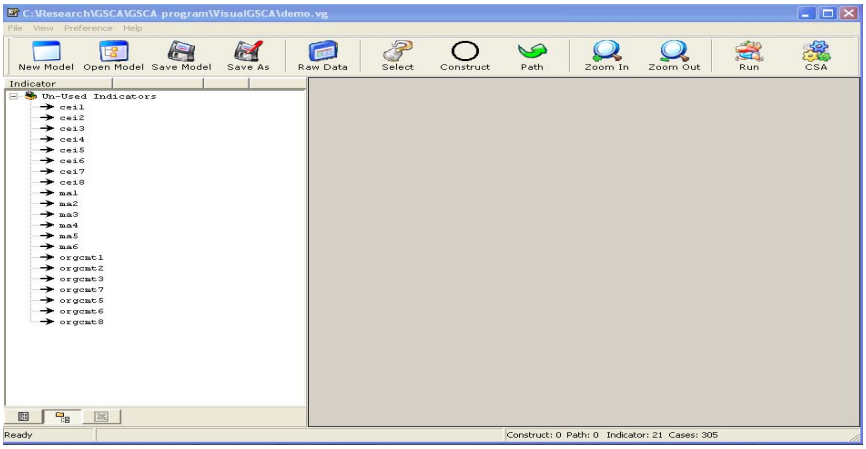

Figure 3. The main program window with the list of indicators found in the input data file.

Step 1: Draw constructs

As shown in Figure 4, users are to draw constructs before assigning indicators to them. This is done as follows:

- Click once on the [Construct] icon  to activate it.
- Click the left mouse button with the cursor placed in the right window of the program as many times as the number of constructs. In the present example, four clicks resulted in the creation of four constructs. By default, the four constructs were initially named *Con_0* to *Con_3*.
- Then, click on the [Stop New Construct] box.

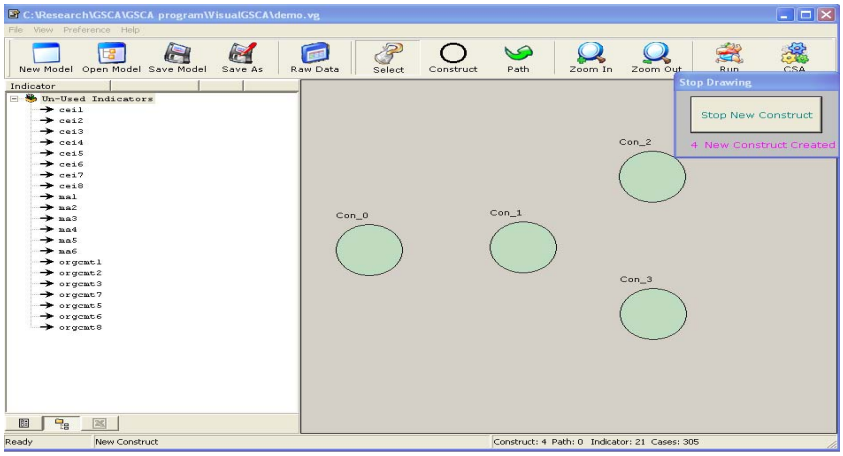

Figure 4. The main program window with all constructs drawn.

Step 2: Specify the measurement model

After drawing constructs, users are to specify the measurement model as follows:

- Double-click on a construct. The *Indicators and Constructs* dialog window will pop up (see Figure 5).
- The construct can be renamed by typing its label in the *Construct Label* dialog box.
- Select the appropriate indicators among the un-used indicators in the list which appears on the left hand side of the dialog window, and then click the [Assign] button to assign the selected indicators to their respective construct.
- Choose whether the selected indicators are to be specified as *Reflective* (default) or *Formative*.
- Repeat the above steps for the remaining constructs.

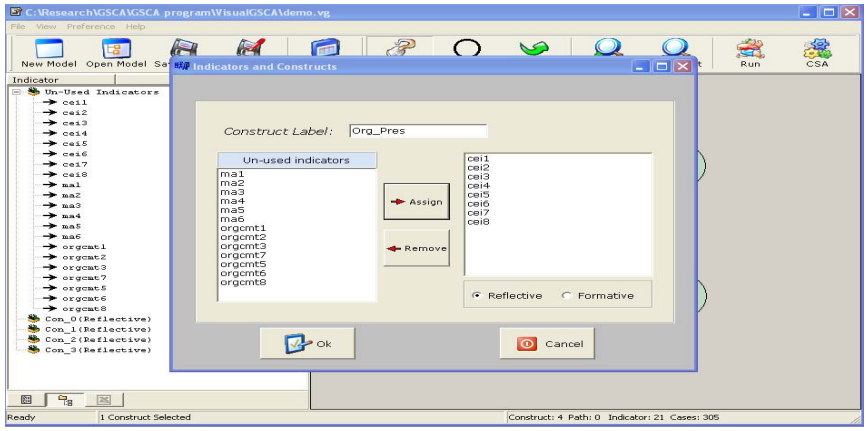

Figure 5. The Indicators and Constructs dialog window to be used for specifying the measurement model.

Step 3: Specify the structural model

As displayed in Figure 6, the structural model can be specified by adding paths among constructs in the main window. The paths are to be drawn as follows:

- Press the [Path] icon  to activate it.
- Click on one exogenous construct and then drag the path to the corresponding endogenous construct. Repeat the above steps until all paths are drawn, and then click on the [Stop New Path] box.

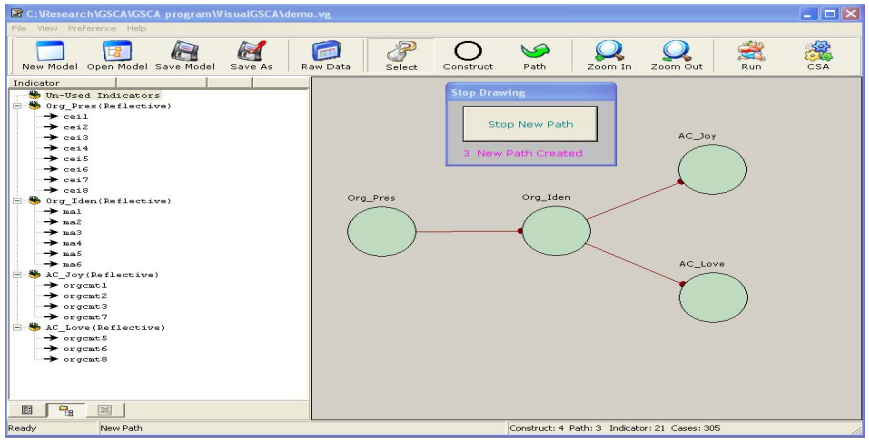

Figure 6. The main program window with all paths drawn.

3.4. Run GSCA and View results

Once the steps presented above are completed, one can run GSCA for fitting the specified model to the data. This is done by clicking on the [Run] icon 

. By default, the program employs the bootstrap method with 100 bootstrap samples for providing the standard errors of parameter estimates. As shown in Figure 7, once the estimation procedure is completed, the estimates of loadings and weights for all indicators are displayed in the left *Indicator* window, along with their standard errors (SE) in parentheses. The estimates of path coefficients are also displayed graphically over/next to each of the previously drawn paths. Additionally, measures of overall model fit are provided in the *Model Fit* pop-up window which appears below the model diagram in Figure 7.

In this application, GSCA estimation indicated that the specified model accounted for about 61% of the variance in the data (FIT = .606, AFIT = .537). Moreover, both GFI and SRMR values suggested a good level of overall model fit (GFI = .993, SRMR = .079). The weight estimates of all indictors were significant, suggesting that all indicators contributed well to determining their corresponding constructs. Also, the loading estimates of all indicators were significant and of appreciable magnitude, indicating that the constructs explained a large portion of the variances in their corresponding indicators. Furthermore, with respect to the structural model, organizational prestige had a significant and positive effect on organizational identification. In turn, organizational identification significantly and positively impacted both forms of affective commitment.

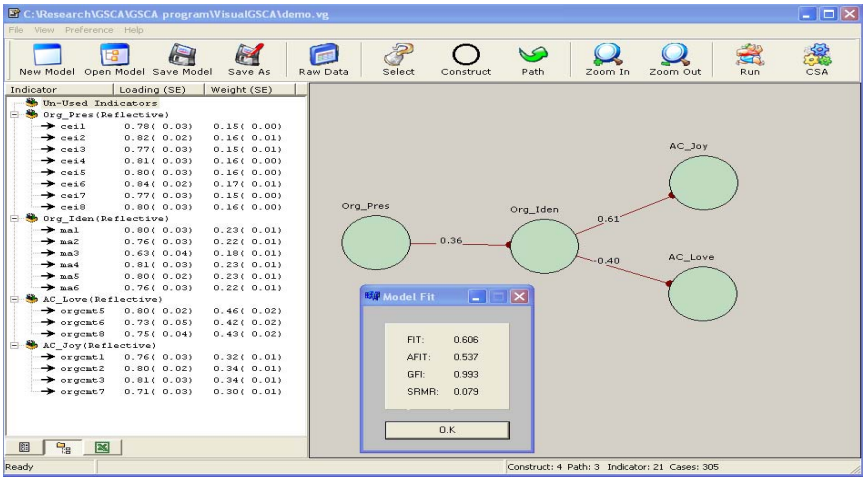

Figure 7. The main program window with GSCA results displayed.

3.5. Application of covariance structure analysis to the same model

In addition to GSCA estimation, users can also apply CSA (based on maximum likelihood) for fitting the same model to the data by clicking on the [CSA] icon 

. The CSA estimates of unstandardized loadings (*Loading_US*) and their standard errors (SE) are displayed in the left *Indicator* window. Their standardized counterparts (*Loading_S*) are also presented in the same window. The CSA estimates of path coefficients are displayed graphically over/next to their respective paths and measures of overall model fit are, as before, provided in the *Model Fit* pop-up window.

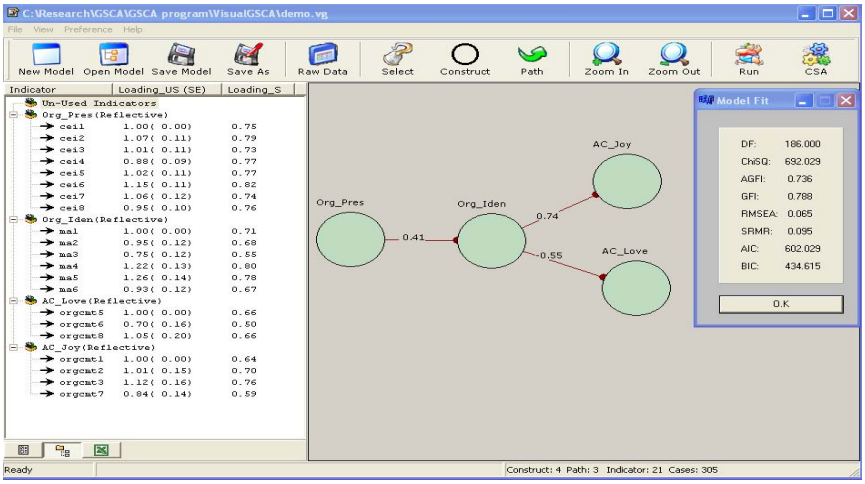

Figure 7. The main program window with CSA results displayed.

4. Concluding Remarks

As demonstrated in this brief overview, *VisualGSCA 1.0* is quite user-friendly, whereby users can directly draw their model as a path diagram and then view model estimation results in a program window. The availability of this software is likely to contribute to the wide adoption of GSCA for SEM applications in psychology and various other fields of inquiry.

Although the core program is efficient in terms of model specification and estimation, the software requires additional work so that it may provide users with a greater number of analytical capabilities and incorporate recent developments in GSCA. For instance, as discussed in Hwang and Takane (2004), one major advantage of GSCA over PLS is that it can easily accommodate multiple group comparisons through the use of constraints imposed across groups. However, this feature is not yet incorporated into the current program. Moreover, an extension of GSCA has recently been proposed that combines GSCA with fuzzy clustering in a unified framework

which, in turn, accounts for group-level heterogeneity (Hwang et al., 2007). This represents another desirable feature which needs to be incorporated in subsequent versions of the software.

References

Arbuckle, J. (1994). AMOS – Analysis of moment structures. *Psychometrika, 59*, 135-137.

Bergami, M., & Bagozzi, R. P. (2000). Self-categorization, affective commitment and group self-esteem as distinct aspects of social identify in the organization. *British Journal of Social Psychology, 39*, 555-577.

de Leeuw, J., Young, F. W., & Takane, Y. (1976). Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika, 41*, 471-503.

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

Fu, J.-R. (2005). *Visual PLS 1.04* [available at http://fs.mis.kuas.edu.tw/~fred/vpls/]

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Hwang, H. (2007), *VisualGSCA* [available at http://www.psych.mcgill.ca/perpg/fac/hwang/VisualGSCA.html]

Hwang, H., & Takane, Y. (2004). Generalized structured component analysis. *Psychometrika, 69*, 81-99.

Hwang, H., DeSarbo, W. S., & Takane, Y. (2007). Fuzzy clusterwise generalized structured component analysis. *Psychometrika, 72*, 181-198.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger, & O. D. Duncan (Eds.), *Structural Equation Models in the Social Sciences*. New York: Academic Press.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Methods*. Mooresville, IN: Scientific Software, Inc.

Lohmöller, J.-B. (1989). *Latent Variable Path Modeling with Partial Least Squares*. New York: Springer-Verlag.

Wold, H. (1975). Path models with latent variables: The NIPALS approach. In H. M. Blalock et al. (Eds.), *Quantitative Sociology: International Perspectives on Mathematical Statistical Model Building*. New York: Academic Press.