8-1-2005

A procedure for imposing a dichotomous incidence variable on the weighting of items in a dual scaling analysis of successive-categories (rating) data

Adam Wyse

A Procedure for Imposing a Dichotomous Incidence Variable on the Weighting of Items in a Dual Scaling Analysis of Successive-Categories (Rating) Data

Adam E. Wyse

August, 2005

A Thesis Submitted to the Faculty of the Center for Quality and Applied Statistics in Partial Fulfillment of the Requirements for the Degree of Master of Science in Applied Statistics

Prof. Daniel R. Lawrence
(Thesis Advisor)

Prof. Joseph G. Voelkel

Prof. Dr. Steven M. LaLonde

Prof. Dr. Donald D. Baker
(Department Head)

THESIS RELEASE PERMISSION FORM

ROCHESTER INSTITUTE OF TECHNOLOGY

COLLEGE OF EGINEERING

Title of Thesis:

A Procedure for Imposing a Dichotomous Incidence Variable
on the Weighting of Items in a Dual Scaling Analysis
of Successive-Categories (Rating) Data

I, Adam E. Wyse, hereby grant permission to the Wallace Memorial Library of R.I.T. to reproduce my thesis in whole or in part. Any reproduction should not be for commercial use or profit.

Signature ________________________________

Date: August 17, 2005

Abstract

Demographic information collected on surveys is often of certain interest to those who use these surveys to measure customer satisfaction or do market research. This thesis presents a method for imposing a dichotomous incidence variable, possibly a demographic variable, on the weighting of items in a set of successive-categories (rating) data using dual scaling. The idea is to augment the matrix of rating data with the "criterion variable" containing the dichotomous information so that this item determines one of the initial solutions of the dual scaling analysis. In conjunction with the augmentation of the criterion item, the original data are "centered" between two numbers that represent the two criterion groups. The resulting modified data matrix is then subjected to a dual scaling analysis. The procedure is discussed with practical guidelines for its use and interpretation of results. Examples of application involve both fabricated and actual data.

Table of Contents

I. Introduction....5

II. Literature Review....7

III. Successive-Categories Data....13

IV. Matrix Notation....18

V. Principles Behind the Procedure....22

VI. An Expositional Application....26

VII. A Real-Data Application....35

VIII. Other Possible Formulations....46

IX. Conclusions and Topics for Future Research....49

References....51

Appendix A....54

Appendix B....61

Appendix C....67

Appendix D....72

I. Introduction

It is often of particular interest to those who develop surveys to solicit not only respondents' preferences and opinions, but also to study commonalities or differences in the respondents that are being surveyed. Many of these differences or commonalities are captured in the form of demographic information. For example, it might be of particular interest to those who are conducting the survey to study the effect that gender has on how respondents answered a specific item or set of items. Further, it might be of importance to identify the questions whose responses are most consistent with gender, since products and services are often designed to target segments of the population that have certain characteristics.

The items on which such demographic information is collected are often categorical in nature and can usually be classified as incidence data, since membership in a specific demographic category is generally absolute and does not indicate a degree of "preference" of one category over another. If the data of the survey are also in the form of incidence data, then this particular demographic characteristic can be included with the incidence preference data of the survey and submitted to dual scaling for a forced classification analysis (See Nishisato, 1984.). In this way the demographic variable can be made to determine the solution. The survey items or questions whose responses are most influenced by the particular demographic will have the highest absolute weights. (Nishisato (1994) and Day (1989) both carried out this type of analysis.) Unfortunately, though, preference data are not always collected in the form of incidence data. What happens when these data have been collected in the form of dominance data (i.e., paired-comparison, rank-order, or successive-categories data) and the user still wants to determine the effect of a demographic variable whose information has been collected as incidence data?

Lawrence (2000) discussed this problem as it pertains to rank-order data. He described a method for imposing a dichotomous (male vs. female, user vs. non-user, etc.) criterion variable on the weighting of items in a dual scaling analysis of a set of rank-order data. Lawrence (manuscript in preparation) has also formulated a procedure for handling paired-comparison data that is similar to his method for analyzing rank-order data. In both cases, the criterion variable is "transformed" to dominance data and included as part of the data matrix, which is then subjected to a dual scaling analysis. Presently, there is no such method for handling successive-categories data.

This thesis proposes a method for the case of successive-categorical data that capitalizes on both the definition of the format and the structure of the data. This procedure is applied to contrived data with practical guidelines for use of the method and interpretation of the solutions. Finally, a subset of "real" data obtained from a health survey given at the Rochester Institute of Technology is analyzed, followed by a discussion of the results.

II. Literature Review

The study and application of surveys in market research and educational testing in recent decades have necessitated the development of multivariate models and methods to handle a wide range of categorical data. Several distinct categorical data types have been popularized and studied in that regard. Among them are contingency tables, multiple-choice questions, sorting formats, paired comparisons, ranking data, and successive-categories (or rating) data. Many of the multivariate methods that have been developed to handle these particular nominal and ordinal data types have origins that trace back to the 1930's, 1940's and 1950's (Richarson & Kuder, 1933; Hirshfield, 1935; Horst, 1935; Fisher 1940; Guttman, 1941; Burt, 1950; Hayashi, 1950) and are similar in many respects. These methods have been introduced under a variety of names, including American Optimal Scaling, Optimal Scoring, Appropriate Scoring, Canadian Dual Scaling, Dutch Homogeneity Analysis, French Multiple Correspondence Analysis, Israeli Scalogram Analysis, and Japanese Quantification. In most instances, the main difference, it would seem, is the name itself, which is primarily a function of where a particular method was developed (Tenenhaus & Young, 1985). All of these methods rely of the principle of singular-value decomposition and are usually formulated along the lines of the Method of Reciprocal Averages, the Analysis-of-Variance Approach, the Principal-Components-Analysis Approach, or the Generalized-Canonical-Analysis Approach (Tenenhaus & Young, 1985). The objective, of course, is to determine a set of optimal weights for stimuli/items (columns) and scores for subjects/respondents (rows) given some initial constraints or conditions—that is, some criterion that needs to be maximized and constraints that must be satisfied based on the formulation (normalization, sum-of-squares-equal-to-a-prescribed-total, sum-of-weighted-responses-equal-to-zero, and so on).

Recently, correspondence analysis has become increasingly popular and Greenacre (1984) and Lebart, Morineau and Warwick (1984) have published texts devoted to its exposition. These texts have given rise to the development and implementation of statistical software to handle categorical data via correspondence analysis (CA) or multiple correspondence analysis (MCA). Introduced by Benzecri (1973), CA was originally developed to handle contingency tables. Its extension, MCA, is applied to multiway multiple-choice data. Although related to dual scaling, CA and MCA, as original formulated, could only be used on incidence data while their Canadian counterpart dual scaling (DS), developed by Nishisato (1978, 1980b & 1994), can be used on both incidence data (i.e., contingency/frequency, multiple-choice, and sorting data) and dominance data (i.e., paired-comparison, rank-order, and successive-categories data) (Nishisato & Gaul, 1988).

Greenacre and Torres (2002) have since proposed a method called "doubling" that attempts, at least in part, to bridge the gaps between CA and DS. Their method uses each subject's preference and dispreference counts, one data set atop the other, and was shown to produce results that are identical to those of DS. Greenacre and Torres also showed that while DS analyzes "centered" data, CA analyzes "uncentered" data and that the only differences in the methods lay in the so-called "scaling factors".

van de Velden (2000) also discussed this method and noted the equivalence of dual scaling and correspondence analysis using this strategy. He further proposed that a dual scaling analysis of Greenacre and Torres' "doubled" matrix would be equivalent to analyzing the dominance matrix of DS. To the user, the most obvious difference between the two methods would be the content and appearance of the output. Obviously, because CA works with a "doubled" data matrix, it would have to be judged less computationally efficient than DS in the analysis of dominance data.

Dual scaling, which is a discrete analogue of principal components analysis, also allows for a unique option known as forced classification. This procedure is based the *Principle of Internal Consistency* (Guttman, 1950) and the *Principle of Equivalent Partitioning* (Nishisato, 1984). The *Principle of Internal Consistency* states that if a given response pattern is repeated in the data matrix, it will become the primary factor in determining a solution. The *Principle of Equivalent Partitioning* states that if an item is repeated a certain number of times *k*, this is equivalent to introducing the item once (in a (1,0) format) and multiplying it by *k*. (This principle is very similar to the *Principle of Distributional Equivalence* of correspondence analysis (Benzecri, 1973)). Together, these two properties make forced classification possible.

Basically, forced classification augments an original matrix by repeating a given response pattern a certain number of times *k*. This means that the original number of items *n* is altered so that the new number of items is $(n+k-1)$ and a $N \times (n+k-1)$ matrix is formed, where *N* is the number of respondents (Nishisato, 1984). As the value of *k* is increased, theoretically to infinity, the particular item that has been weighted by *k* becomes the principal factor in determining the first solution of the dual scaling analysis. In fact, the repeated item and the new dimension have a correlation that approaches one, since the repeated item effectively defines the dimension (Nishisato & Gaul, 1990). Nishisato (1986) generalized his forced classification procedure to handle not only multiple-choice data but also sorting data and, in a specific way, rank-order data and paired-comparison data (1986). He further demonstrated that the value of *k* may be any real number and can be chosen not only to get a particular solution to *dominate* the analysis but also to cause a particular solution to be *suppressed.* Forced classification has also been adapted for use on contingency tables, and this is commonly referred to as *conditional forced classification* (Nishisato & Baba, 1999). The mathematical aspects of the procedure for each of these particular data types, excluding conditional forced classification, are discussed by Nishisato (1988 & 1994). Applications of forced classification have not reached their full potential for several reasons, including software limitations; few software packages include dual scaling which offers forced classification, but many do include correspondence analysis, which does *not* (as of yet) offer it. (Actually, correspondence analysis does offer a procedure known as partial multiple correspondence analysis, which can be used to eliminate the effect of a particular item from the other items of the analysis, and while no proof of equivalence has been established, this procedure is believed to be similar to conditional forced classification (Yanai & Maeda, 2002).)

Several different applications of the method of forced classification have been considered in recent literature. Nishisato and Gaul (1988 & 1990) and Day (1989) discussed applications of the method utilizing a "criterion" variable in the case of multiple-choice data. (Nishisato and Gaul (1990) also applied it (a bit differently) in the analysis of paired-comparison data.) Day (1989) specifically looked at using an "ideal subject" as the criterion variable in a forced classification. That a particular subject, classification or demographic item could be used as the criterion item in a forced classification analysis is an interesting idea. Nishisato (1994) further discussed this notion, stating that an "unrelated" item could possibly be added to a questionnaire or data set and then be used as the criterion item for forced classification.

If demographic information on the respondents is captured, the criterion item might very well be a demographic characteristic. Since the responses to the preference items would themselves constitute incidence data, it is readily apparent that a demographic variable could be used as the criterion item in the subsequent forced classification analysis. The analyst can then determine which other items are highly correlated with the dimension defined by the criterion item. Along these lines, Lawrence (1997), in a study of some customer satisfaction data collected on a hospital survey, used overall satisfaction as the criterion item in a forced classification analysis. In this and all the other expositions noted, except for the case involving paired-comparison data, the data were either incidence data or treated as incidence data, having a (1,0) format, and a value *k* was chosen to weight the criterion item for "forcing" the solution.

The study of dominance data (i.e., paired-comparison, rank-order, and successive-categories data) has been somewhat limited, especially as it relates to forced classification. Nishisato (1984, 1986, & 1994) did study forced classification of dominance data, but he did it in the framework of choosing a *pair* of items to be used as the criteria (Nishisato, 1994). He stated that *two* items were needed to "drive" the solution and that each of these arbitrarily chosen criterion items must be multiplied by *k* within the so-call "dominance matrix" in order to force the solution. He described how this could be done in the cases of paired-comparison and rank-order data but neglected the case of successive-categories data. Lawrence (2000) noted a two-fold limitation of this type of analysis, observing that these two items must be part of the original data set and that they might not necessarily be complementary.

Lawrence (2000) proposed a method for imposing a dichotomous (incidence) variable, the criterion item, on the weighting of items in a dual scaling analysis of rank-order (dominance) data. Essentially, Lawrence's method combined the incidence item with dominance data by, in effect, converting the incidence data to rankings and combining them with the original body of rank data. The subsequent matrix of rankings was then subjected to a standard dual scaling analysis and was shown to "force" the dimension of interest to be the first solution of the analysis—that is, the criterion item was made to determine the first solution.

More recently, Lawrence (manuscript in preparation) has applied a similar strategy in the forced classification analysis of paired-comparison data, here again subject to a dichotomous incidence variable. The case involving successive-categories data is the focus of this thesis.

III. Successive-Categories Data

The study and implementation of methods for the analysis of successive-categories (rating) data have been somewhat scarce with results that are sometimes debatable. Successive-categories data, also know as ordered-choice or Likert data, are based on a set of strictly ordered ratings (e.g., *Poor*, *Fair*, *Good*) that are assigned to a given set of stimuli by each respondent. Many analysts, in their multivariate analyses, treat this type of data as if it were continuous, often subjecting the data to principal components analysis or factor analysis. Neither of these techniques is really appropriate for analyzing such data, especially when the response categories are few in number. Nishisato's dual scaling approach best analyzes this type of data for what it is, namely categorical, and generally should be used for analyses of this sort.

Nishisato and Sheu (1984) discussed several different methods that might be an improvement on a method initially proposed by Nishisato (1980a) that was based on a Thurstonian formulation. Nishisato and Sheu outlined three methods, termed *Method D* (design-matrix), *Method R* (ranking), and *Method P* (paired-comparison), with the intent of correcting several problems (including imbalance of stimuli and category boundaries) that were discovered in the earlier version. *Method R* is the simplest of the three and was used by Nishisato in his 1994 study of successive-categories data using dual scaling. It is for this purpose that *Method R* is used in the formulation and procedure proposed in this project.

Method R assigns ranks to both the stimuli and the (implicit) category boundaries that exist in the data. For example, a set of successive-categories responses might be *Poor*, *Fair*, *Good* and *Excellent*, where 1 is made to represent *Poor*, 2 is made to represent *Fair*, 3 is made to represent *Good*, and 4 is made to represent *Excellent*. (*Table 1* is an example of a typical successive-categories response pattern with four subjects rating five items using any one of four possible ratings (or categories).) There would be three category boundaries, one between each pair of successive categories—that is, between *Poor* and *Fair*, *Fair* and *Good*, and *Good* and *Excellent*. Ranks are then assigned to both the items that have been rated by the respondents and the category boundaries between these ratings.

Table 1:

Subject	Item 1	Item 2	Item 3	Item 4	Item 5
1	3	4	2	3	1
2	4	4	2	1	3
3	1	2	4	3	2
4	2	2	3	4	2

Nishisato (1994) describes a systematic procedure for assigning these ranks: Starting with the first category, the average rank $(k_0 + 1)/2$, where k_0 is the total number of stimuli classified into Category 1, is given to all the stimuli in that category. The category boundary τ_1, between Categories 1 and 2, is given rank $k_0 + 1$. The next set of stimuli, k_1, where k_1 is the total number of stimuli classified into Category 2, are assigned the rank $(k_1 + 1)/2 + (k_0 + 1)$. The second category boundary τ_2, between Categories 2 and 3, is given the rank $(k_0 + 1) + (k_1 + 1)$. This process continues until all the categories and category boundaries have received the proper rankings. (The data in *Table 1* would have the rankings shown in *Table 2* on the following page.) These rankings are then converted to dominance numbers e_{ij} by the formula

$e_{ij} = 2K_{ij} - (n + m + 1)$, where K_{ij} is the rank assigned by subject *i* to item *j*, *n* is the number of items, and *m* is the number of category boundaries (Nishisato, 1994). The resulting matrix of dominance numbers is then subjected to a dual scaling analysis.

Table 2:

Subject	τ_1	τ_2	T_3	Item 1	Item 2	Item 3	Item 4	Item 5
1	2	4	7	5.5	8	3	5.5	1
2	2	4	6	7.5	7.5	3	1	5
3	3	5	7	1.5	4	8	6	1.5
4	1	5	7	3	3	6	8	3

A major source of debate and reason for limited analysis of successive-categories data (as categorical data) revolves around the aspect of multidimensionality. It is often the case that respondents answering surveys respond in a multidimensional sense and that they are often using several different criteria in selecting their response to a survey question. It is with this in mind that Nishisato (1994) raised questions with regard to the analysis of successive-categories data. He pointed out that due to the empirical nature of the analysis, solutions beyond the first one commonly do not have strictly ordered category boundaries. This often makes it rather difficult to interpret any solution other than the first. Extensive work has been done by Odondi (1997) to try to remedy this. Both Odondi (1997) and Nishisato (1994) proposed methods based on multi-step procedures using both clustering methods and dual scaling in which subjects were grouped homogenously with each subsequent grouping then subjected to a separate dual scaling analysis. These methods, notably that of Odondi, were shown to produce results with category boundaries ordered properly. These methods are not very user-friendly, however, and often require more than one pass through the data.

In the initial formulation of the dual scaling procedure for handling successive-categories data, the constraint of ordered category boundaries is not in effect (Nishisato, 1994). The idea of extracting only one solution or having to use a clustering procedure with more than one pass through the data is usually not very attractive to those doing the data analysis, especially in view of the fact that there are usually multiple dimensions to be extracted. Extracting only the first solution, which has ordered category boundaries, appears to be the most popular analysis strategy since this solution would seem to be the only one that can be reasonably easily interpreted. However, this issue takes on greater importance and generates more debate when respondents from different subpopulations are represented in the same data set. In fact, it would often make sense, at least intuitively, that these people would answer questions in a different manner. Males and females, for instance, might be subconsciously using a different continuum or set of boundaries when answering questions. In this sense, the category boundaries of a dimension so determined might not be ordered due to these differences. In fact, one would expect that the differences that exist between different subpopulations would be reflected in the category boundaries and that the category boundary most affected by group dissimilarities would have the highest absolute weight and the category boundary least affected by dissimilarities would have the lowest absolute weight. These boundaries would not necessarily be in order since the category boundary that might do most to separate the two groups would not necessarily be at the extremes of the data. It would seem that the problem would become compounded as the number subgroups responding to the survey increases.

In fact, though, the analyst is at least as, if not more, concerned with the weights of items of a survey than he is those of the category boundaries—especially given that the category boundaries have been arbitrarily introduced into the analysis anyway. With that in mind, the procedure of this thesis is based on extracting more than one solution in the dual scaling analysis and then identifying the items whose weights most reflect the different demographic characteristics of the subgroups, ignoring the category boundaries and their order. The first solution is still of certain importance, since it has the ordered category boundaries and explains most of the variation in the data, but the focus is placed almost entirely on subsequent solutions (primarily the second) and the weights assigned to the items in those solutions. (A study of the category boundaries would be of interest only with regard to the ones for the original data and then only from the standpoint of identifying differences in the continuum due to subgroups.)

Finally, there are two additional attributes or characteristics of successive-categories data that are worth mentioning. *Table 1* illustrates a very common but notable situation that occurs when people are asked to respond using rating scales. Subject 4 of *Table 1* has chosen not to give any of the five items a *Poor* rating. It is quite possible that a subject would not use all of the categories in assigning ratings to a set of items, or that a given item would receive the same rating from all subjects or none of a particular rating from all subjects. This will be important in dual scaling analyses of successive-categories data to follow.

A second important attribute of successive-categories data is that each response is generally independent of any other. The rating that a respondent gives to each item is based on how he defines the categories in relation to that particular item. In essence, his responses are based on his definition of the continuum of choices he has to choose from based on that item. There is not a strict inter-dependency of items—as there is in the case of rank-order data, say, where once one particular item is given the top ranking, no other item can receive the same rank (assuming no ties). With these concepts in mind, the method of this thesis is formulated.

IV. Matrix Notation

Let $\mathbf{F}_{N\times n}$ be a $N \times n$ matrix of successive-categories data, where N is the number of subjects and n is the number of items. Also, let $\mathbf{x}_{N\times 1}$ be a column vector representing some grouping or demographic characteristic. In most cases, membership in a demographic (demo) group is generally coded as "1", "2", "3", …, where "1" represents membership in the first group, "2" represents membership in the second group, and so on.

The goal is to find a procedure for imposing the effects of the demographic variable on the weighting of the items in the original data set $\mathbf{F}$ in such a way that this variable determines the axis of the "forced" dimension. In essence, we want to force the analysis to have a solution reflecting the effects of the demographic characteristic. The aim is to get this solution to be one of the first solutions of the analysis.

Suppose a survey is given to two groups of respondents—males and females, or some other dichotomous demo grouping—and we are interested in the effect of the male/female dichotomy, say, in the way these people respond to the items on a survey. Membership in the two groups represented in the demo variable is recorded on the survey. A column vector $\mathbf{x}_{N\times 1}$ for the demographic variable, indicating the group membership of each respondent, is constructed with attribute A (e.g., male) coded as a "1" and attribute B (e.g., female) coded as a "2". Somehow we want to include the information of this incidence demo variable with the dominance successive-categories data, collected as ratings on the survey, so that in a subsequent dual scaling analysis the weighting of the items in the data set will reflect the effect of the demographic variable in one of the initial solutions of the analysis.

We will begin by introducing the column vector $\mathbf{x}*_{N\times1}$, a sort of redefinition of $\mathbf{x}_{N\times1}$. Let all of those that belong to the first group in $\mathbf{x}_{N\times1}$ be given a value of 1 in $\mathbf{x}*_{N\times1}$. Then, let all those that belong to the second group in $\mathbf{x}_{N\times1}$ be assigned the value $c+2+2\cdot c_1$ in $\mathbf{x}*_{N\times1}$, where c is the number of categories in the original data set and c_1 is the number of new categories to be included between the original data and the two extreme values in $\mathbf{x}*_{N\times1}$. (Note that the variables c_1 and $c+2+2\cdot c_1$ are set simultaneously.) More simply, $\mathbf{x}*_{N\times1}$ is defined so that Demo Group 1 = 1 and Demo Group 2 = $c+2+2\cdot c_1$.

By way of example, suppose that $N=5$ respondents, or subjects, in two demographic groups rate $n=3$ items on a scale of 1 to $c=3$ (i.e., into any one of three categories). The demo vector **x** and the data matrix **F** might, respectively, look like

$$\mathbf{x}_{5\times1} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \\ 1 \end{bmatrix} \text{ and } \mathbf{F}_{5\times3} = \begin{bmatrix} 1 & 3 & 2 \\ 3 & 2 & 1 \\ 3 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 3 & 2 \end{bmatrix}.$$

If $c_1 = 5$ new categories were to be included on either side of the categories of the original data matrix **F**, then we would have $c+2+2\cdot c_1 = 3+2+10 = 15$ and

$$\mathbf{x}*_{5\times1} = \begin{bmatrix} 1 \\ 15 \\ 15 \\ 1 \\ 1 \end{bmatrix}.$$

The original data (consisting of ratings) should also be modified so that 1 and $c+2+2\cdot c_1$ are the extreme choices of the successive categories, and the ratings in the original data are "centered" in between 1 and $c+2+2\cdot c_1$. (The purpose of "centering" the data matrix $\mathbf{F}_{N\times n}$ and altering the demo vector $\mathbf{x}_{N\times 1}$ will be explained in the following chapter.) The modification of $\mathbf{F}$ is done in the following manner: Let $\mathbf{G}_{N\times n}$ be a "centering" matrix to be added. Define $\mathbf{G}$ such that $\mathbf{G} = (c_1+1)\cdot \mathbf{1}_{N\times n}$, where c_1 is again the number of new categories included and $\mathbf{1}_{N\times n}$ is the unit matrix. We now modify the original data matrix $\mathbf{F}$ by adding $\mathbf{G}$ and augmenting $\mathbf{x}^*$ to that sum. Expressed symbolically,

$$\mathbf{F}^*_{N\times(n+1)} = \left[(\mathbf{F}+\mathbf{G})_{N\times n} \middle| \mathbf{x}^*\right].$$

The new $N\times(n+1)$ matrix $\mathbf{F}^*$ is then submitted to dual scaling for analysis.

Again, by way of example, if $\mathbf{F}$, c_1 and $\mathbf{x}^*$ are defined as above, the centering matrix would be

$$\mathbf{G}_{5\times 3} = (5+1)\cdot \mathbf{1}_{5\times 3} = \begin{bmatrix} 6 & 6 & 6 \\ 6 & 6 & 6 \\ 6 & 6 & 6 \\ 6 & 6 & 6 \\ 6 & 6 & 6 \end{bmatrix},$$

so that

$$\mathbf{F}^*_{5\times 4} = \left[\left[\begin{pmatrix} 1 & 3 & 2 \\ 3 & 2 & 1 \\ 3 & 1 & 2 \\ 1 & 2 & 2 \\ 1 & 3 & 2 \end{pmatrix} + \begin{pmatrix} 6 & 6 & 6 \\ 6 & 6 & 6 \\ 6 & 6 & 6 \\ 6 & 6 & 6 \\ 6 & 6 & 6 \end{pmatrix}\right] \middle| \begin{bmatrix} 1 \\ 15 \\ 15 \\ 1 \\ 1 \end{bmatrix}\right] = \left[\begin{array}{ccc|c} 7 & 9 & 8 & 1 \\ 9 & 8 & 7 & 15 \\ 9 & 7 & 8 & 15 \\ 7 & 8 & 8 & 1 \\ 7 & 9 & 8 & 1 \end{array}\right].$$

Note that the categories of the original data have been "centered" in $\mathbf{F}^*$ between 1 and 15. There are ten new categories in $\mathbf{F}^*$, five (2, 3, 4, 5 and 6) between 1 (the category representing Demo Group 1 in $\mathbf{x}^*$) and 7 (the "adjusted" lowest category of **F**) and five (10, 11, 12, 13 and 14) between 9 (the "adjusted" highest category of **F**) and 15 (the category representing Demo Group 2 in $\mathbf{x}^*$).

V. Principles Behind the Procedure

The procedure described in this thesis is based on several principles that allow for its formulation and the interpretation of its results. First of all, the values in the demo vector $\mathbf{x}^*$ must be assigned in such a way that they are at the extreme categories of the "new" data matrix $\mathbf{F}^*$ and away from the categories of the original data matrix $\mathbf{F}$. The definition of *Method R* for analyzing successive-categories data and the way that dual scaling handles the data makes the reason for the definition of $\mathbf{F}^*$ apparent. If the dichotomous elements that comprise $\mathbf{x}^*$ were simply the extreme categories of original the data, then the groupings of the respondent scores would be dependent on how many other items had been classified by respondents in those extreme categories. This would make it difficult to identify the two demographic groups by way of the respondent scores, since each score within a subgroup would not necessarily have the same, or even approximately the same, magnitude. This necessitates the assignment of values in $\mathbf{x}^*$ to be such that the two groups are represented by the extreme categories of $\mathbf{F}^*$, away from the categories of the original data.

The definition of $\mathbf{x}^*$ also depends on the number of new categories c_1 that are included between the original data and the two extreme categories representing the two subgroups. (Including additional categories can between the extreme categories of the data and the categories represented in the demo vector $\mathbf{x}^*$ doesn't violate the "integrity" of the original data; the added categories could simply be thought of as categories that are unused when a respondent rates the items.) In most cases, adding ten new categories to each side of the data matrix is sufficient.

An examination of the proposed procedure points to differences between this method and other methods for doing forced classification. One major difference is that in this procedure the effect of the demographic variable appears in the second solution of the dual scaling output instead of the first solution, as it does in the forced classification procedures that have been developed to handle of other types of data. In this case, the first solution will have the category boundaries ordered and represent the original data as free of the effect of the demo variable as possible, while the second solution will not necessarily have ordered category boundaries but will reflect the effect of the demo variable. The reason for this is probably directly related to the additional categories introduced into the matrix $\mathbf{F}^*$. These new categories are directly related to the values (categories/group number) in the vector $\mathbf{x}^*$, and the corresponding category boundaries are linear combinations of each other in the dominance matrix. The dominance numbers for the category boundaries between the category corresponding to Demo Group 1 and the categories for the original data are "complementary" to those between the categories for the original data and the category corresponding to Demo Group 2. The difference between the dominance numbers for Demo Group 1 and Demo Group 2 in each of these newly-introduced category boundaries is only two, which follows directly from the way the dominance matrix is constructed.

The ratio of the dominance number for Demo Group 1 to that for Demo Group 2 approaches one as more new categories are introduced into the matrix $\mathbf{F}^*$. In fact, if an infinite number of new categories (and category boundaries) were introduced, the difference in the dominance numbers would be so minimal that it would be impossible to distinguish between the two demographic groups based on the dominance numbers for the category boundaries in the dominance matrix. Under these circumstances, the first solution of the dual scaling analysis would account for virtually 100% of the variation in the data with categories boundaries appearing to determine that first solution.

Since the ratio of the dominance numbers for the two demographic groups approaches one as new categories are added, the respondent scores for the two groups approach one and the weights for the items in $\mathbf{F}^*$ will approach those for the items in $\mathbf{F}$. (An exception to this occurs when item responses in the original data are similar—in the same or opposite direction—to the groupings in the demo vector. In this case, the correlation between the weights for the items in $\mathbf{F}$ and the weights for items in $\mathbf{F}^*$ will not be as high as it would be if none of the items of where similar to the groupings in the demo vector.) It is important to note that $\mathbf{F}^*$ contains the demo vector as an additional item; hence, there will never be a perfect correlation between weights for the items in $\mathbf{F}$ and $\mathbf{F}^*$ unless the demographic variable had no effect on the original responses comprising the data.

Since the first solution of the dual scaling analysis accounts for most of the variation in the original data and category boundaries, the second solution will account for the demographic variable. (The elements of the vector in the dominance matrix corresponding to the demographic variable are the most positive and most negative in value, hence it is not surprising that this would constitute the next most variation to be explained.) Accordingly, the assignment of weights to the items will be influenced by the demographic variable. Item weights that are directionally the same as the weight for the demo item will have the highest signed value. Moreover, the respondent scores will fall into two groups (consistent with grouping in the demo item) that are opposite in sign but similar in magnitude; *within* the demographic groups, the respondent scores are very nearly the same. These are desired results of a forced classification analysis.

Another major difference between this forced classification procedure and the procedures for handling other types of dominance data is the need for only a one-column formulation of the demographic variable to force the solution. The apparent reason that only one column is needed in this case is linked to the nature of successive-categories data; the items in successive-categories are not inter-dependent. In fact, as was previously mentioned, the response to each item is based on the respondent's definition of the rating categories in relation to each individual item. Therefore, it would seem to make sense that a single-column formulation of the criterion item could be used.

VI. Expositional Application

Suppose ten respondents are asked six questions that involve ratings based on a rating scale where 1 represents "poor", 2 represents "fair", 3 represents "good", and 4 represents "excellent", leading to a 10×6 data matrix **F**. Also, suppose that information on a dichotomous demographic variable for the ten respondents is recorded along with the rating data. The demographic information comprises a one-column vector **x**, where among its elements a 1 represents membership in Group 1 and a 2 represents membership in Group 2. The proposed modification to **F** is shown below with ten new categories included between the centered data and the two "adjusted" values of the demo variable, which become the two extreme categories of $\mathbf{F}^*$. The centering matrix **G** is added to **F** and the new demo vector $\mathbf{x}^*$ is then augmented to form the 10×7 matrix $\mathbf{F}^*$. An analysis of both the original data matrix **F** and the new matrix $\mathbf{F}^*$ is carried out using dual scaling. The results for the two analyses are shown, including the dominance matrix, item weights, category boundary weights, respondent scores, and percent of total variation explained by each solution. The two sets of results are then interpreted and compared.

Let $\mathbf{F} = \begin{bmatrix} 4 & 1 & 3 & 4 & 2 & 3 \\ 3 & 3 & 2 & 1 & 2 & 3 \\ 3 & 2 & 3 & 2 & 4 & 2 \\ 4 & 1 & 4 & 3 & 2 & 4 \\ 3 & 3 & 3 & 4 & 3 & 1 \\ 4 & 4 & 3 & 4 & 2 & 2 \\ 2 & 2 & 2 & 4 & 3 & 1 \\ 4 & 3 & 2 & 3 & 2 & 4 \\ 1 & 2 & 3 & 2 & 4 & 3 \\ 2 & 1 & 1 & 1 & 3 & 3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 1 \\ 2 \\ 1 \\ 2 \end{bmatrix}$.

Then, $\mathbf{G} = \begin{bmatrix} 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \end{bmatrix}$ and $\mathbf{x}^* = \begin{bmatrix} 26 \\ 1 \\ 1 \\ 26 \\ 26 \\ 26 \\ 1 \\ 26 \\ 1 \\ 26 \end{bmatrix}$, so that

$$\mathbf{F}+\mathbf{G} = \begin{bmatrix} 4 & 1 & 3 & 4 & 2 & 3 \\ 3 & 3 & 2 & 1 & 2 & 3 \\ 3 & 2 & 3 & 2 & 4 & 2 \\ 4 & 1 & 4 & 3 & 2 & 4 \\ 3 & 3 & 3 & 4 & 3 & 1 \\ 4 & 4 & 3 & 4 & 2 & 2 \\ 2 & 2 & 2 & 4 & 3 & 1 \\ 4 & 3 & 2 & 3 & 2 & 4 \\ 1 & 2 & 3 & 2 & 4 & 3 \\ 2 & 1 & 1 & 1 & 3 & 3 \end{bmatrix} + \begin{bmatrix} 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \\ 11 & 11 & 11 & 11 & 11 & 11 \end{bmatrix} = \begin{bmatrix} 15 & 12 & 14 & 15 & 13 & 14 \\ 14 & 14 & 13 & 12 & 13 & 14 \\ 14 & 13 & 14 & 13 & 15 & 13 \\ 15 & 12 & 15 & 14 & 13 & 15 \\ 14 & 14 & 14 & 15 & 14 & 12 \\ 15 & 15 & 14 & 15 & 13 & 13 \\ 13 & 13 & 13 & 15 & 14 & 12 \\ 15 & 14 & 13 & 14 & 13 & 15 \\ 12 & 13 & 14 & 13 & 15 & 14 \\ 13 & 12 & 12 & 12 & 14 & 14 \end{bmatrix},$$

$$\text{and } \mathbf{F}^* = \left[\begin{array}{cccccc|c} 15 & 12 & 14 & 15 & 13 & 14 & 26 \\ 14 & 14 & 13 & 12 & 13 & 14 & 1 \\ 14 & 13 & 14 & 13 & 15 & 13 & 1 \\ 15 & 12 & 15 & 14 & 13 & 15 & 26 \\ 14 & 14 & 14 & 15 & 14 & 12 & 26 \\ 15 & 15 & 14 & 15 & 13 & 13 & 26 \\ 13 & 13 & 13 & 15 & 14 & 12 & 1 \\ 15 & 14 & 13 & 14 & 13 & 15 & 26 \\ 12 & 13 & 14 & 13 & 15 & 14 & 1 \\ 13 & 12 & 12 & 12 & 14 & 14 & 26 \end{array}\right].$$

The dominance matrix for **F** works out to be

$$\mathbf{E} = \begin{bmatrix} -6 & -2 & 4 & 7 & -8 & 1 & 7 & -4 & 1 \\ -6 & 0 & 8 & 4 & 4 & -3 & -8 & -3 & 4 \\ -8 & 0 & 6 & 3 & -4 & 3 & -4 & 8 & -4 \\ -6 & -2 & 2 & 6 & -8 & 6 & 0 & -4 & 6 \\ -6 & -4 & 6 & 1 & 1 & 1 & 8 & 1 & -8 \\ -8 & -2 & 2 & 6 & 6 & 0 & 6 & -5 & -5 \\ -6 & 2 & 6 & -2 & -2 & -2 & 8 & 4 & -8 \\ -8 & -2 & 4 & 7 & 1 & -5 & 1 & -5 & 7 \\ -6 & 0 & 6 & -8 & -3 & 3 & -3 & 8 & 3 \\ -2 & 2 & 8 & 0 & -6 & -6 & -6 & 5 & 5 \end{bmatrix}.$$

(Note that the dominance numbers for the category boundaries comprise the first three columns of dominance matrix **E**, while the dominance numbers for the items make up the last six columns.)

Finally, the dominance matrix for $\mathbf{F}^*$ is computed to be

$$
\mathbf{E}^* = \begin{bmatrix}
-31 & -29 & -27 & -25 & -23 & -21 & -19 & -17 & -15 & -13 & -11 & -7 & -3 & 3 & \cdots \\
-29 & -27 & -25 & -23 & -21 & -19 & -17 & -15 & -13 & -11 & -9 & -5 & 1 & 9 & \cdots \\
-29 & -27 & -25 & -23 & -21 & -19 & -17 & -15 & -13 & -11 & -9 & -7 & 1 & 7 & \cdots \\
-31 & -29 & -27 & -25 & -23 & -21 & -19 & -17 & -15 & -13 & -11 & -7 & -3 & 1 & \cdots \\
-31 & -29 & -27 & -25 & -23 & -21 & -19 & -17 & -15 & -13 & -11 & -7 & -5 & 5 & \cdots \\
-31 & -29 & -27 & -25 & -23 & -21 & -19 & -17 & -15 & -13 & -11 & -9 & -3 & 1 & \cdots \\
-29 & -27 & -25 & -23 & -21 & -19 & -17 & -15 & -13 & -11 & -9 & -5 & 3 & 7 & \cdots \\
-31 & -29 & -27 & -25 & -23 & -21 & -19 & -17 & -15 & -13 & -11 & -9 & -3 & 3 & \cdots \\
-29 & -27 & -25 & -23 & -21 & -19 & -17 & -15 & -13 & -11 & -9 & -5 & 1 & 7 & \cdots \\
-31 & -29 & -27 & -25 & -23 & -21 & -19 & -17 & -15 & -13 & -11 & -3 & 1 & 7 & \cdots
\end{matrix}
$$

$$
\begin{matrix}
\cdots & 9 & 11 & 13 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 29 & 6 & -9 & 0 & 6 & -5 & 0 & 31 \\
\cdots & 11 & 13 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 29 & 31 & 5 & 5 & -2 & -7 & -2 & 5 & -31 \\
\cdots & 11 & 13 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 29 & 31 & 4 & -3 & 4 & -3 & 9 & -3 & -31 \\
\cdots & 9 & 11 & 13 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 29 & 5 & -9 & 5 & -1 & -5 & 5 & 31 \\
\cdots & 9 & 11 & 13 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 29 & 0 & 0 & 0 & 7 & 0 & -9 & 31 \\
\cdots & 9 & 11 & 13 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 29 & 5 & 5 & -1 & 5 & -6 & -6 & 31 \\
\cdots & 11 & 13 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 29 & 31 & -1 & -1 & -1 & 9 & 5 & -7 & -31 \\
\cdots & 9 & 11 & 13 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 29 & 6 & 0 & -6 & 0 & -6 & 6 & 31 \\
\cdots & 11 & 13 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 29 & 31 & -7 & -2 & 4 & -2 & 9 & 4 & -31 \\
\cdots & 9 & 11 & 13 & 15 & 17 & 19 & 21 & 23 & 25 & 27 & 29 & -1 & -7 & -7 & -7 & 4 & 4 & 31
\end{bmatrix}
$$

(Note that the dominance numbers for category boundaries comprise the first 25 columns of dominance matrix $\mathbf{E}^*$, while the dominance numbers for the items make up the last seven columns.)

The output of the dual analysis of **F** follows:

Variation by Solution for DSA of **F** (Original Data)

Variation	1	2	3	4	5
Pct Total	34.414	24.224	21.105	10.901	5.022
Cum Pct	34.414	58.638	79.653	90.554	95.576

Item Weights by Solution for DSA of **F** (Original Data)

Item	1	2	3	4	5
1	1.092	0.482	-1.443	-0.092	0.246
2	-0.626	0.825	-0.270	2.324	0.919
3	-0.015	0.224	0.297	-1.376	2.276
4	0.583	2.009	0.457	-0.825	-1.217
5	-0.100	-0.990	1.884	-0.047	0.091
6	-0.080	-1.492	-1.599	-0.509	-0.398

Category Boundaries by Solution for DSA of **F** (Original Data)

Cat Bound	1	2	3	4	5
1	-2.095	0.111	-0.170	-0.391	-0.810
2	-0.355	-0.359	0.254	0.158	-0.562
3	1.595	-0.810	0.589	0.758	-0.545

Respondent Scores by Solution for DSA of **F** (Original Data)

Resp	1	2	3	4	5
1	1.372	0.461	-0.512	-1.438	-0.864
2	0.843	-0.965	-0.847	1.739	0.608
3	1.093	-0.754	1.061	-0.071	1.398
4	1.042	-0.398	-0.964	-1.775	0.971
5	1.092	1.090	0.979	0.302	0.080
6	1.041	1.360	-0.331	1.033	0.758
7	0.951	0.657	1.525	0.185	-1.233
8	1.176	-0.170	-1.378	0.681	-0.760
9	0.465	-1.319	1.278	-0.192	0.673
10	0.586	-1.701	0.195	0.242	-1.690

The output of the dual scaling analysis of $\mathbf{F}^*$ follows:

Variation by Solution for DSA of $\mathbf{F}^*$ (Demo Included)

Variation	1	2	3	4	5
Pct Total	89.889	8.926	0.498	0.311	0.192
Cum Pct	89.889	98.815	99.313	99.624	99.816

Item Weights by Solution for DSA of $\mathbf{F}^*$ (Demo Included)

Item	1	2	3	4	5
1	0.128	0.293	0.115	-2.671	-1.915
2	-0.122	-2.600	-1.170	-3.142	3.023
3	-0.025	-0.241	-0.618	0.902	-3.377
4	0.042	0.220	-3.792	0.316	-1.108
5	0.012	-0.751	0.116	3.476	1.506
6	-0.006	0.023	3.892	-0.753	-1.043
7	0.393	5.480	0.133	0.562	0.503

Category Boundaries by Solution for DSA of **F** * (Demo Included)

Cat Bound	1	2	3	4	5
1	-1.726	-0.056	-0.007	-0.014	-0.030
2	-1.612	-0.064	-0.007	-0.015	-0.029
3	-1.498	-0.072	-0.007	-0.015	-0.028
4	-1.384	-0.080	-0.006	-0.015	-0.027
5	-1.269	-0.088	-0.006	-0.015	-0.026
6	-1.155	-0.096	-0.006	-0.016	-0.025
7	-1.041	-0.104	-0.006	-0.016	-0.024
8	-0.927	-0.112	-0.006	-0.016	-0.023
9	-0.812	-0.120	-0.006	-0.016	-0.023
10	-0.698	-0.128	-0.005	-0.017	-0.022
11	-0.584	-0.136	-0.005	-0.017	-0.021
12	-0.367	-0.112	0.534	0.872	0.724
13	-0.060	-0.369	0.447	0.438	0.617
14	0.283	-0.395	0.438	0.399	1.426
15	0.559	-0.217	-0.003	-0.019	-0.012
16	0.673	-0.225	-0.003	-0.020	-0.011
17	0.787	-0.233	-0.003	-0.020	-0.010
18	0.901	-0.241	-0.003	-0.020	-0.009
19	1.016	-0.249	-0.003	-0.020	-0.008
20	1.130	-0.257	-0.002	-0.021	-0.007
21	1.244	-0.265	-0.002	-0.021	-0.006
22	1.358	-0.273	-0.002	-0.021	-0.005
23	1.473	-0.282	-0.002	-0.021	-0.005
24	1.587	-0.290	-0.002	-0.022	-0.004
25	1.701	-0.298	-0.002	-0.022	-0.003

Respondent Scores by Solution for DSA of **F** * (Demo Included)

Resp	Group	1	2	3	4	5
1	2	1.020	0.820	-0.258	0.216	-1.357
2	1	0.972	-1.221	0.963	-1.920	0.388
3	1	0.975	-1.256	-0.032	0.395	-0.628
4	2	1.018	0.808	0.748	0.229	-1.946
5	2	1.018	0.775	-1.446	0.585	0.808
6	2	1.017	0.807	-1.158	-1.121	0.540
7	1	0.973	-1.234	-1.457	0.336	0.145
8	2	1.019	0.813	0.656	-1.160	0.375
9	1	0.971	-1.275	0.498	1.094	0.032
10	2	1.016	0.741	1.488	1.343	1.649

A look at the first five solutions of the dual scaling analysis of **F** indicates that the demographic characteristic does not define *any* of these dimensions. This does not mean that the ratings of the respondents were not influenced by the demo in the analysis, only that the latent effect of the demo does not show up in the first five solutions. With the demo included as the criterion item, however, the dual scaling analysis (of **F***) clearly shows that the demo item, not unexpectedly, determines the second solution of the analysis. In effect, the demo variable *fixes the axis* of the second solution. It appears that Items 1 and 4 "load" most positively on this demo-defined dimension, while Item 2 "loads" very negatively. An examination of the original data and the demo item reveals that the ratings for Items 1 and 4 seem to align with the 1's and the 2's of the demo, and the ratings for Item 2 seem to align in the opposite direction. The weights of the first six items in the first solution of the DSA of **F*** are similar to the weights of the same items in the DSA of the original data matrix **F**. It is important to note that the item weights would increasingly differ in the DSAs of **F** and **F*** as the ratings for one or more items "lined up" with the elements of the demo item.

The category boundaries of the first solution of both analyses are ordered, as one would expect. The category boundaries of the second solution of the DSA of **F*** are *not* ordered, but the category boundaries for the items of the original data (Category Boundaries 12, 13, and 14) *are* ordered. This is not always the case using this procedure, but as was previously mentioned, in a forced classification analysis the analyst would generally be more concerned with the item weights than the category boundaries. Category Boundary 14, having the highest magnitude, appears to be most affected by the demo item, albeit negatively. As for the respondent scores in the second solution, they settle into two groups determined by the pattern in the demo vector, with Group 1 scores taking on values around −1.2 and Group 2 scores assuming values around 0.8.

VII. Real-Data Application

A health survey was administered by the Student Health Center at Rochester Institute of Technology (RIT) asking students about their stress level (Appendix C). The rating scaling on the survey had 1 representing "never", 2 representing "seldom", 3 representing "occasionally", 4 representing "often", and 5 representing "most of the time". Students were also asked several demographic questions on the survey (gender, employment status, health appointment vs. no health appointment, etc.). A convenient subset of 20 respondents who answered the first section of the survey (concerning frustrations) was selected for inclusion in the analyses to follow, which in the case of the forced classification analyses, separately included the demo items "gender" (male vs. female), "employment status" (employed vs. unemployed), and "had a health appointment vs. no health appointment". The output from the DSA of the original data is compared to the output from the forced classification analyses, one for each individual demo item. The original rating data, the demo vectors, and the modified matrices are shown below. The percent of variation accounted for by solution, item weights and corresponding respondent scores for each analysis follow.

From the survey responses, we have

$$\mathbf{F} = \begin{bmatrix} 3 & 3 & 2 & 3 & 2 & 4 & 2 \\ 5 & 3 & 2 & 2 & 2 & 1 & 1 \\ 3 & 3 & 2 & 2 & 2 & 2 & 3 \\ 3 & 3 & 4 & 3 & 2 & 3 & 3 \\ 4 & 3 & 3 & 3 & 3 & 1 & 1 \\ 4 & 3 & 4 & 5 & 2 & 3 & 3 \\ 3 & 3 & 4 & 3 & 2 & 2 & 4 \\ 3 & 3 & 4 & 3 & 2 & 2 & 2 \\ 4 & 3 & 2 & 2 & 1 & 1 & 2 \\ 3 & 3 & 3 & 4 & 3 & 2 & 3 \\ 2 & 2 & 3 & 2 & 2 & 3 & 3 \\ 3 & 3 & 2 & 2 & 3 & 1 & 2 \\ 4 & 4 & 3 & 2 & 2 & 2 & 2 \\ 3 & 3 & 5 & 2 & 3 & 3 & 2 \\ 3 & 3 & 4 & 3 & 1 & 1 & 2 \\ 3 & 3 & 4 & 3 & 4 & 5 & 2 \\ 4 & 3 & 3 & 3 & 4 & 3 & 3 \\ 5 & 4 & 5 & 3 & 4 & 5 & 4 \\ 3 & 4 & 5 & 4 & 2 & 1 & 3 \\ 4 & 4 & 3 & 2 & 2 & 5 & 2 \end{bmatrix}, \ \mathbf{x}_1 = \begin{bmatrix} 2 \\ 2 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 1 \\ 2 \\ 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 2 \\ 1 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 1 \\ 2 \\ 2 \\ 2 \\ 1 \\ 2 \end{bmatrix}, \text{ and } \mathbf{x}_3 = \begin{bmatrix} 2 \\ 1 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 1 \\ 1 \\ 1 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 1 \\ 2 \end{bmatrix},$$

where $\mathbf{F}$ is the matrix of original rating data for the 20 respondents, $\mathbf{x}_1$ the *Gender* demographic with 1 representing "male" and 2 representing "female", $\mathbf{x}_2$ is the *Employment* demo with 1 representing "unemployed" and 2 representing "employed", and $\mathbf{x}_3$ is the *Health-Appointment* demo with 1 representing "had a health appointment" and 2 representing "did not have a health appointment".

"Centered" and augmented by the modified demo $\mathbf{x}_1^*$, $\mathbf{F}$ becomes

$$\mathbf{F}_1^* = \left[\begin{array}{ccccccc|c}
14 & 14 & 13 & 14 & 13 & 15 & 13 & 27 \\
16 & 14 & 13 & 13 & 13 & 12 & 12 & 27 \\
14 & 14 & 13 & 13 & 13 & 13 & 14 & 1 \\
14 & 14 & 15 & 14 & 13 & 14 & 14 & 1 \\
15 & 14 & 14 & 14 & 14 & 12 & 12 & 27 \\
15 & 14 & 15 & 16 & 13 & 14 & 14 & 27 \\
14 & 14 & 15 & 14 & 13 & 13 & 15 & 27 \\
14 & 14 & 15 & 14 & 13 & 13 & 13 & 1 \\
15 & 14 & 13 & 13 & 12 & 12 & 13 & 27 \\
14 & 14 & 14 & 15 & 14 & 13 & 14 & 1 \\
13 & 13 & 14 & 13 & 13 & 14 & 14 & 1 \\
14 & 14 & 13 & 13 & 14 & 12 & 13 & 1 \\
15 & 15 & 14 & 13 & 13 & 13 & 13 & 27 \\
14 & 14 & 16 & 13 & 14 & 14 & 13 & 27 \\
14 & 14 & 15 & 14 & 12 & 12 & 13 & 27 \\
14 & 14 & 15 & 14 & 15 & 16 & 13 & 1 \\
15 & 14 & 14 & 14 & 15 & 15 & 14 & 1 \\
16 & 15 & 16 & 14 & 15 & 16 & 15 & 1 \\
14 & 15 & 16 & 15 & 13 & 12 & 14 & 27 \\
15 & 15 & 14 & 13 & 13 & 16 & 13 & 1
\end{array}\right],$$

where $\mathbf{F}_1^*$ includes ten new categories between the "centered" data and extreme values of the demo item *Gender*. The augmentation of the modified demo item followed the addition of the centering matrix to the original data matrix $\mathbf{F}$.

"Centered" and augmented by the modified demo $\mathbf{x}_2$*, $\mathbf{F}$ becomes

$$\mathbf{F}_2* = \left[\begin{array}{ccccccc|c}
14 & 14 & 13 & 14 & 13 & 15 & 13 & 27 \\
16 & 14 & 13 & 13 & 13 & 12 & 12 & 1 \\
14 & 14 & 13 & 13 & 13 & 13 & 14 & 1 \\
14 & 14 & 15 & 14 & 13 & 14 & 14 & 1 \\
15 & 14 & 14 & 14 & 14 & 12 & 12 & 1 \\
15 & 14 & 15 & 16 & 13 & 14 & 14 & 1 \\
14 & 14 & 15 & 14 & 13 & 13 & 15 & 1 \\
14 & 14 & 15 & 14 & 13 & 13 & 13 & 27 \\
15 & 14 & 13 & 13 & 12 & 12 & 13 & 1 \\
14 & 14 & 14 & 15 & 14 & 13 & 14 & 27 \\
13 & 13 & 14 & 13 & 13 & 14 & 14 & 27 \\
14 & 14 & 13 & 13 & 14 & 12 & 13 & 27 \\
15 & 15 & 14 & 13 & 13 & 13 & 13 & 27 \\
14 & 14 & 16 & 13 & 14 & 14 & 13 & 27 \\
14 & 14 & 15 & 14 & 12 & 12 & 13 & 1 \\
14 & 14 & 15 & 14 & 15 & 16 & 13 & 27 \\
15 & 14 & 14 & 14 & 15 & 15 & 14 & 27 \\
16 & 15 & 16 & 14 & 15 & 16 & 15 & 27 \\
14 & 15 & 16 & 15 & 13 & 12 & 14 & 1 \\
15 & 15 & 14 & 13 & 13 & 16 & 13 & 27
\end{array}\right],$$

where $\mathbf{F}_2$* is the modified data matrix, the modified demo vector for *Employment* having been augmented after the centering matrix was added to $\mathbf{F}$. Ten new categories were included between the "centered" data and the extreme values of the demo item.

"Centered" and augmented by the modified demo $\mathbf{x}_3^*$, $\mathbf{F}$ becomes

$$
\mathbf{F}_3^* = \left[\begin{array}{ccccccc|c}
14 & 14 & 13 & 14 & 13 & 15 & 13 & 27 \\
16 & 14 & 13 & 13 & 13 & 12 & 12 & 1 \\
14 & 14 & 13 & 13 & 13 & 13 & 14 & 27 \\
14 & 14 & 15 & 14 & 13 & 14 & 14 & 27 \\
15 & 14 & 14 & 14 & 14 & 12 & 12 & 27 \\
15 & 14 & 15 & 16 & 13 & 14 & 14 & 27 \\
14 & 14 & 15 & 14 & 13 & 13 & 15 & 27 \\
14 & 14 & 15 & 14 & 13 & 13 & 13 & 27 \\
15 & 14 & 13 & 13 & 12 & 12 & 13 & 27 \\
14 & 14 & 14 & 15 & 14 & 13 & 14 & 1 \\
13 & 13 & 14 & 13 & 13 & 14 & 14 & 1 \\
14 & 14 & 13 & 13 & 14 & 12 & 13 & 1 \\
15 & 15 & 14 & 13 & 13 & 13 & 13 & 27 \\
14 & 14 & 16 & 13 & 14 & 14 & 13 & 27 \\
14 & 14 & 15 & 14 & 12 & 12 & 13 & 27 \\
14 & 14 & 15 & 14 & 15 & 16 & 13 & 27 \\
15 & 14 & 14 & 14 & 15 & 15 & 14 & 27 \\
16 & 15 & 16 & 14 & 15 & 16 & 15 & 27 \\
14 & 15 & 16 & 15 & 13 & 12 & 14 & 1 \\
15 & 15 & 14 & 13 & 13 & 16 & 13 & 27
\end{array}\right],
$$

where $\mathbf{F}_3^*$ is the "centered" data matrix augmented by the modified demo vector for *Health-Appointments* Once again, ten new categories have been included between the centered data and the extreme values of the demo item.

The output of the dual scaling analysis of **F** follows:

Variation by Solution for DSA of **F** (Original Data)

Variation	1	2	3	4	5
Pct Total	57.443	14.782	9.733	5.544	4.464
Cum Pct	57.443	72.225	81.958	87.502	91.966

Item Weights by Solution for DSA of **F** (Original Data)

Item	1	2	3	4	5
1	0.681	-0.132	-1.356	-0.634	0.793
2	0.350	-0.521	-0.678	0.109	1.111
3	0.595	0.274	2.091	-1.199	1.688
4	-0.132	-0.857	1.111	-1.024	-1.737
5	-0.768	0.650	-0.985	-1.568	-1.127
6	-0.609	2.799	0.029	0.666	0.029
7	-0.608	-0.344	1.204	1.872	-0.550

Respondent Scores by Solution for DSA of **F** (Original Data)

Resp	1	2	3	4	5
1	0.914	0.966	-0.759	0.999	-1.155
2	0.999	-0.883	-1.677	-0.657	0.585
3	1.021	-0.430	-0.815	2.049	-0.409
4	1.140	0.365	1.258	0.460	0.174
5	1.079	-0.613	-0.961	-1.638	0.402
6	1.025	-0.120	1.199	-0.703	-0.904
7	1.051	-0.492	1.469	0.999	-0.107
8	1.241	-0.384	0.600	-0.630	0.459
9	1.070	-1.127	-0.845	0.967	0.658
10	1.003	-0.638	0.508	-0.567	-2.374
11	0.838	0.851	0.945	2.044	-0.526
12	0.976	-0.724	-1.559	-0.145	-0.759
13	1.188	-0.254	-0.855	0.198	1.345
14	1.039	0.940	0.132	-0.849	1.149
15	1.151	-0.936	0.813	-0.295	0.738
16	0.699	1.993	0.092	-1.133	-0.605
17	0.780	1.525	-0.945	-0.556	-1.536
18	0.647	1.746	0.359	-0.621	1.398
19	1.021	-0.957	1.459	-0.793	0.450
20	0.886	1.477	-0.787	0.825	1.205

The output of the dual scaling analysis of $\mathbf{F}_1$* follows:

Variation by Solution for DSA of $\mathbf{F}_1$* (Male vs. Female)

Variation	1	2	3	4	5
Pct Total	90.009	8.727	0.391	0.315	0.174
Cum Pct	90.009	98.736	99.127	99.442	99.616

Item Weights by Solution for DSA of $\mathbf{F}_1$* (Male vs. Female)

Item	1	2	3	4	5
1	0.172	0.002	-0.059	2.353	0.610
2	0.078	-0.004	-0.833	1.169	-0.587
3	0.145	-0.070	0.679	-3.705	1.571
4	-0.046	0.050	-1.173	-1.918	2.058
5	-0.175	-0.481	0.971	1.781	3.477
6	-0.114	-0.714	5.006	-0.018	-1.718
7	-0.153	-0.334	-1.071	-2.092	-2.627
8	0.002	5.696	0.594	0.022	-0.112

Respondent Scores by Solution for DSA of $\mathbf{F}_1$* (Male vs. Female)

Resp	Group	1	2	3	4	5
1	2	0.999	0.960	1.409	0.762	-1.263
2	2	1.000	1.015	-0.463	1.684	0.448
3	1	1.000	-0.983	-1.304	0.762	-1.663
4	1	1.003	-0.988	-0.286	-1.313	-0.401
5	2	1.002	1.006	-0.236	0.961	1.610
6	2	1.001	0.997	0.362	-1.189	0.513
7	2	1.001	0.991	-0.317	-1.476	-0.907
8	1	1.005	-0.961	-1.012	-0.658	0.665
9	2	1.001	1.018	-0.834	0.840	-1.159
10	1	1.000	-0.974	-1.415	-0.546	1.232
11	1	0.997	-1.019	0.115	-0.983	-1.718
12	1	0.999	-0.974	-1.527	1.517	0.684
13	2	1.004	0.999	0.080	0.837	-0.565
14	2	1.002	0.964	1.377	-0.138	0.537
15	2	1.003	1.020	-0.577	-0.820	0.071
16	1	0.995	-1.038	1.637	-0.104	1.196
17	1	0.997	-1.033	0.981	0.923	0.873
18	1	0.994	-1.026	1.500	-0.357	0.371
19	2	1.000	1.016	-0.566	-1.450	0.690
20	1	0.999	-1.014	1.104	0.756	-1.209

The output of the dual scaling analysis of $\mathbf{F}_2$* follows:

Variation by Solution for DSA of $\mathbf{F}_2$* (Employed vs. Unemployed)

Variation	1	2	3	4	5
Pct Total	90.102	8.718	0.350	0.259	0.179
Cum Pct	90.102	98.820	99.170	99.429	99.608

Item Weights by Solution for DSA of $\mathbf{F}_2$* (Employed vs. Unemployed)

Item	1	2	3	4	5
1	0.166	-0.301	-1.575	2.264	1.394
2	0.072	-0.266	-1.549	0.216	-0.155
3	0.139	-0.276	2.715	-2.609	2.024
4	-0.052	-0.467	0.373	-2.234	1.696
5	-0.180	0.252	-0.595	1.193	2.649
6	-0.117	0.570	3.718	3.547	-0.839
7	-0.160	-0.336	1.325	-1.349	-3.364
8	0.194	5.664	-0.450	-0.686	-0.073

Respondent Scores by Solution for DSA of $\mathbf{F}_2$* (Employed vs. Unemployed)

Resp	Group	1	2	3	4	5
1	2	1.008	0.906	-0.900	0.830	-0.955
2	1	0.988	-1.110	-1.458	1.248	0.769
3	1	0.989	-1.100	-0.529	0.979	-1.984
4	1	0.992	-1.090	1.441	-0.046	-0.396
5	1	0.990	-1.105	-0.835	0.728	1.693
6	1	0.989	-1.109	1.112	-0.175	0.804
7	1	0.989	-1.120	0.912	-0.894	-1.026
8	2	1.014	0.846	-0.447	-1.707	0.469
9	1	0.989	-1.130	-1.064	0.568	-0.876
10	2	1.009	0.846	-0.685	-1.885	0.326
11	2	1.007	0.896	0.876	-0.803	-2.177
12	2	1.009	0.860	-2.138	-0.415	-0.037
13	2	1.013	0.859	-1.230	-0.272	-0.243
14	2	1.011	0.903	0.320	-0.310	0.772
15	1	0.991	-1.133	0.088	-0.812	0.290
16	2	1.005	0.948	1.236	0.723	1.141
17	2	1.006	0.932	0.232	1.318	0.598
18	2	1.003	0.925	1.422	0.897	0.766
19	1	0.988	-1.133	0.500	-1.373	0.746
20	2	1.008	0.921	0.357	1.430	-0.678

The output of the dual scaling analysis of F_3* follows:

Variation by Solution for DSA of F_3* (Health Appointment vs. No Health Appointment)

Variation	1	2	3	4	5
Pct Total	92.319	6.361	0.447	0.314	0.179
Cum Pct	92.319	98.680	99.127	99.441	99.620

Item Weights by Solution for DSA of F_3* (Health Appointment vs. No Health Appointment)

Item	1	2	3	4	5
1	0.143	0.015	-0.748	2.426	1.152
2	0.048	0.207	-1.131	1.245	-0.166
3	0.116	0.063	-0.096	-3.707	2.121
4	-0.075	0.317	-1.399	-1.870	1.799
5	-0.203	0.303	2.005	1.626	2.707
6	-0.135	-0.406	4.829	-0.253	-1.174
7	-0.181	0.246	-0.247	-2.064	-3.298
8	0.941	-5.626	-0.482	0.057	0.009

Respondent Scores for DSA of F_3* (Health Appointment vs. No Health Appointment)

Resp	Group	1	2	3	4	5
1	2	1.023	-0.551	0.698	0.720	-0.999
2	1	0.924	1.774	-0.411	1.649	0.653
3	2	1.024	-0.505	-0.672	0.848	-2.048
4	2	1.027	-0.535	-0.085	-1.271	-0.458
5	2	1.025	-0.501	-0.888	1.003	1.641
6	2	1.025	-0.530	-0.530	-1.163	0.715
7	2	1.025	-0.507	-0.840	-1.428	-0.996
8	2	1.029	-0.512	-0.848	-0.575	0.637
9	2	1.025	-0.501	-1.487	0.940	-0.942
10	1	0.924	1.786	-0.089	-0.579	0.529
11	1	0.922	1.740	1.398	-1.094	-2.083
12	1	0.923	1.791	-0.094	1.491	0.113
13	2	1.028	-0.521	-0.684	0.884	-0.180
14	2	1.025	-0.542	0.616	-0.188	0.844
15	2	1.026	-0.503	-1.382	-0.740	0.312
16	2	1.019	-0.567	1.881	-0.200	1.114
17	2	1.021	-0.554	1.429	0.863	0.537
18	2	1.018	-0.587	1.502	-0.409	0.636
19	1	0.924	1.776	-0.583	-1.484	0.785
20	2	1.023	-0.577	1.127	0.736	-0.809

A look at the weights assigned to the items in the dual scaling analysis of the original data matrix **F** suggests that students tend to have experienced more frustration in the areas addressed by the first three questions than in the areas addressed by the last four questions. It appears that students seem to be experiencing higher frustrations in professionally and economically related areas than they are in socially related areas. In this "standard" analysis, there is no clear evidence of the effect of any of the demographic characteristics.

The dual scaling analysis of $\mathbf{F}_1$* forces the second solution to reflect the effect of "gender" in the analysis (with the first solution paralleling the first solution in the analysis of **F**). The respondents in Group 1 are males, while those in Group 2 are females. All respondents in each group are assigned the same score in the second solution, approximately 1 for males and approximately -1 for females. Among the items, there are none that strongly "load" on this "gender" dimension in a positive way. Item 4 loads *most* positively, indicating that females might be slightly more inclined than males to experience frustration at not meeting their goals. Item 6 loads in a highly negative way, indicating that at least among the 20 students selected males experience a very high degree of frustration in dating. Of course, this makes sense since the student population at RIT disproportionately favors males.

Similarly, as expected, the dual scaling analysis of $\mathbf{F}_2$* forces the effect of "employment status" to appear in the second solution. (The first solution again parallels the first solution of the DSA of **F**.) As with the analysis involving "gender", the two groups of respondents in this analysis were assigned two distinct scores, -1.1 for employed and 0.9 for unemployed. Items 5 and 6 load most positively on this "employment" dimension, while Items 4 and 7 load most negatively. This seems to indicate that among the selected subset of 20 respondents employed students experience more frustration pertaining to social issues, while unemployed students experience more frustration pertaining to finding a job and accomplishing their goals. This also makes sense since unemployed students have more time to be with their friends, but at the same time, these same students might not feel they have reached their goals not being gainfully employed.

Finally, the demo variable "health appointment vs. no health appointment" was made to be the criterion item in the formulation of $\mathbf{F}_3$*, which was then submitted to a DSA. The two groups of respondents, as determined by the make-up of the demo item, once again take on two different sets of scores; those in Group 1 (i.e., those who had a health appointment) were assigned scores ranging from 1.7 to 1.8, while respondents in Group 2 (i.e., those who did not have a health appointment) were assigned scores in the neighborhood of −0.5. Item 6 loads most positively on the dimension defined by this demographic variable, indicating that among the 20 students in the selected subset those who feel frustration in accomplishing their goals are the more likely to have had a health appointment than those who do not feel such frustration. (As in the DSA of $\mathbf{F}_1$* and $\mathbf{F}_2$*, the first solution of the DSA of $\mathbf{F}_3$* parallels the first solution of the DSA of **F**.)

VIII. Other Possible Formulations

There are several formulations that might also work in forcing the effect of a dichotomous criterion item on the assignment of weights and scores in a DSA of successive-categories data. This section of the thesis presents three other possible formulations that were considered in the investigations of this thesis.

One alternative method that was considered used a single-column criterion item $\mathbf{x}^*$ (with Group 1 = 1 and Group 2 = $c+2$ as its elements), repeated a sufficient number of times to force $\mathbf{x}^*$ to define the initial dimension (per Guttman's (1950) *Principle of Internal Consistency*). It is important to note that the unit matrix $\mathbf{1}_{N \times n}$ must be added to the original data matrix before the criterion item (i.e., the demo vector) is augmented; if the data matrix $\mathbf{F}$ is not "centered" (by adding the unit matrix to $\mathbf{F}$), the respondent scores of the first demo group will be affected by the number of items in the original data that received the lowest rating. (Under these circumstances, the value (1) representing Demo Group 1 and the lowest rating (1) would be the same, confounding the analysis.) Several problems arose using this strategy, the first of which was that the first solution, which was supposed to reflect effect of the criterion (or demo) item, was sometimes shown to have all negative item weights. This happened in spite of the fact that there was at least one item that "moved" ordinally in the same direction as the demo. Obviously, this made it difficult to identify items whose responses were influenced by the demo variable. The item that was the most ordinally similar to the demo generally had the *smallest* negative weight, but this did little to make it and others like it easy to identify in the output. As might be expected, the second solution of this analysis paralleled the first solution of the analysis of the original data matrix $\mathbf{F}$, except when several items were ordinally similar to the demo.

Another formulation involved augmenting the demo item in the form of a two-column matrix $\mathbf{x}^*_{N\times2}$ to the original data matrix **F**. This two-column matrix was constructed in a manner similar to the one in Lawrence's (2000) method for handling rank-order data—that is, **x*** was created so that Demo Group 1 rows were made to be $[1,\ c+2]$ and Demo Group 2 rows were made to be $[c+2,\ 1]$, where again c is the number of categories in the rating scale. Before **x*** is augmented, the unit matrix $\mathbf{1}_{N\times n}$ is added to **F** in order to "center" the ratings between 1 and $c+2$, the two values that made up **x***. This formulation was usually able to force the first solution of the analysis to be a dimension defined by the demo variable and produce a second solution reflecting the original (unaltered) data, but a problem came up when the number of items got to be more than five. Under these circumstances, the demographic item represented in **x*** will not always dominate the analysis. Capitalizing on Guttman's (1950) *Principle of Internal Consistency*, **x*** would have to be repeated several times in **F*** in order to get it to be the determining "force" in the analysis. Obviously, this strategy is not very computationally efficient, which was a deterrent to its being selected as the formulation of choice. It did perform fairly well for small data sets, but more work is needed to determine if it is able to handle *any* small data set, regardless of the composition. It was also apparent using this formulation that the first solution, whether **x*** is augmented only once or multiple times depending on the number of items in **F**, did somewhat parallel the second solution of the DSA using the formulation proposed in this thesis. In other words, using the two-column expression for **x*** (as defined above), led to weights for the items and scores for the respondents that were fairly similar to in sign but slightly different than in magnitude those based on the proposed single-column expression for **x*** that involved adding new categories.

The last formulation that was considered involved a two-column expression for the demo item that was similar in definition to the one-column expression $\mathbf{x}*_{N\times 1}$ used in the thesis formulation. This two-column matrix $\mathbf{x}*_{N\times 2}$ was defined so that the rows for Demo Group 1 were made to be $[1,\ c+2+2\cdot c_1]$ and the rows for Demo Group 2 were made to be $[c+2+2\cdot c_1,\ 1]$. (In effect, $\mathbf{x}*_{N\times 1}$ was the first column of $\mathbf{x}*_{N\times 2}$, and its additive complement was the second column.) Recall that *c* is the number of original categories and c_1 is the number of new categories introduced into the data set in the formulation. The matrix $\mathbf{G} = (c_1 + 1)\cdot \mathbf{1}_{N\times n}$ is used to "center" the original data matrix $\mathbf{F}$ before $\mathbf{x}*_{N\times 2}$ is augmented to get $\mathbf{F}*$. The results of the DSA of this formulation of $\mathbf{F}*$ showed no improvement over the results using the formulation that involved the single-column expression $\mathbf{x}*_{N\times 1}$. In fact, the results of the two analyses were quite similar—except that the percent of variation explained by the first solution was actually *lower* when the two-column expression $\mathbf{x}*_{N\times 2}$ was used in the formulation of $\mathbf{F}*$. As for the item weights, they were similar in sign and magnitude. As in the case of the first alternative formulation, the second solution here paralleled the first solution of the DSA of the original data.

IX. Conclusions and Topics for Future Research

The proposed procedure of including new categories between the "adjusted" (centered) original data and the extreme values of the modified criterion item allows the user to determine the second solution of a dual scaling analysis so that the effect of the particular demographic characteristic of the criterion item defines that solution. (That the "forced" dimension shows up in the *second* solution sets this procedure apart from other forced classification methods that have the forced dimension showing up in the *first* solution.) Robust results were obtained when at least ten new categories were included. These results are easily interpretable and allow the analyst to identify in the scores the two groups into which the respondents were demographically classified and, more importantly the specific items that strongly load on the dimension defined by the demo variable. In this forced dimension, item weights that have the same sign as the weight assigned to the demo item are positively influenced by the demo, while item weights that have the opposite sign are negatively influenced. Except in situations where there are several items whose ratings strongly reflect the influence of the demographic characteristic, the first solution using the proposed forced classification procedure "parallels" (i.e., is quite similar) the first solution of a standard DSA of the original data.

There are several areas of work that merit further study. One such area involves the use of the "centering" matrix $\mathbf{G}$. Although $\mathbf{G}$ was specifically defined in this study, it is entirely possible that a matrix other than $\mathbf{G}$, as defined, could be used—that is, a matrix which is some multiple of the unit matrix, not just the one that "centers" the ratings the data matrix $\mathbf{F}$ between the two numbers that make up the modified demo vector $\mathbf{x}^*$. (As was discussed earlier in this thesis, "centering" the data produces robust results that are not dependent on the homogeneity of the items.) For example, one possibility might be simply to add the unit matrix itself to the $\mathbf{F}$

matrix. It seems obvious that adding different multiples of the unit matrix to the original data would variously affect the item weighting, but more work is needed to see just what that effect would be.

Another area for future research could focus on extending the proposed procedure for handling a two-group demographic variable to the point where it can handle variables comprised of three or more demo groups. Like the procedure for handling the dichotomous demo, this extension might require only a single-column criterion item to force the solution (which is quite possible given the structure of successive-categories data), but this, too, must be investigated.

Finally, it might also be worthwhile to compare the items with high weights using the proposed method with the high-weight items based on Odondi's (1997) clustering method to see if the two sets of items that are identified would be the same. There are most certainly many possibilities for future research related to the procedure presented in this thesis.

References

Benzecri, J. P. (1973) *L'analysis des donnees T. 2, L'analyse des correspondence* [Data analysis: T. 2, Correspondence analysis]. Paris: Dunod.

Burt, C. (1950). The factor analysis of qualitative data. *British Journal of Psychology*, 3, 166-185.

Day, D. A. (1989). Investigating the validity of the English and French versions of the Myers-Briggs Type Indicator. Unpublished doctoral dissertation, University of Toronto, Canada.

Fisher, R. A. (1940). The precision of discriminant function, *Annals of Eugenics*, 10, 422-429.

Greenacre, M. J. (1984). *Theory and Application of Correspondence Analysis*. London: Academic Press.

Greenacre, M. J. and Torres, A. (2002) Note: Dual scaling and correspondence analysis of preferences, paired comparison and ratings. *International Journal of Research in Marketing,* 19, 401-405.

Guttman, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst et al. (Eds.). *The Prediction of Personal Adjustment.* (pp. 319-348). New York: Social Science Research Council.

Guttman, L. (1950). The principal components of scale analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfield, S. A. Star, & J. A. Clausen. *Measurement and Prediction.* Princeton: Princeton University Press.

Hayashi, C. (1950). On the quantification of qualitative data from the mathematico-statistical point of view. *Annals of the Institute of Statistical Mathematics.* 2 (No. 1), 35-47.

Hirshfield, H. O. (1935). A connection between correlation and contingency. *Cambridge Philosophical Society Proceedings*, 31, 520-524.

Horst, P. (1935). Measuring complex attitudes. *Journal of Social Psychology*, 6, 6 369-374.

Lawrence, D. R. (1997). Analyzing customer-satisfaction data using dual scaling. Paper presented at the 41st Annual Fall Technical Conference of ASA & ASQ. Baltimore, Maryland.

Lawrence, D. R. (2000). Imposing a dichotomous incidence variable on the weighting of rank-order items in dual scaling. Published in the *Proceedings of The 2000 International Conference on Measurement and Multivariate Analysis.* Banff, Alberta, Canada.

Lawrence, D. R. (manuscript in preparation). Imposing a dichotomous incidence variable on the weighting of paired-comparison data in dual scaling. Rochester Institute of Technology, New York.

Lebart, L. , Morineau, A., & Warwick, K. M. (1984) *Multivariate descriptive analysis: Correspondence analysis and related techniques for large matrices.* New York: Wiley-Interscience

Nishisato, S. (1978). Optimal scoring of paired comparison and rank order data: An alternative to Guttman's formulation. *Psychometrika*, 43, 263-271.

Nishisato, S. (1980a). Dual scaling of successive categories data. *Japanese Psychological Research*, 22,134-143.

Nishisato, S. (1980b). *Analysis of Categorical Data: Dual Scaling and Its Application.* Toronto: University of Toronto Press.

Nishisato, S. (1984). Forced classification: A simple application of a quantification technique. *Psychometrika*, 49, 25-36

Nishisato, S. (1986). Generalized forced classification for quantifying categorical data. In E. Diday, Y. Escoufier, L. Lebart, J. Pages, Y. Schektman, & R. Tomassone (Eds.), *Data analysis and informatics, IV* (pp.351-362). Amsterdam, the Netherlands: North-Holland

Nishisato, S. (1988). Forced classification procedure of dual scaling: Its mathematical roperties. In H. H. Boch (Eds.), *Classification and Related Methods of Data Analysis* (pp. 523-532). North-Holland: Amsterdam, The Netherlands.

Nishisato, S. (1994). *Elements of Dual Scaling: An Introduction to Practical Data Analysis.* Hillsdale, NJ: Lawrence Erblaum.

Nishisato, S & Baba, Y. (1999). On contingency, projection and forced classification of dual scaling. *Behaviormetrika*, 26, 207-219.

Nishisato, S. & Gaul W. (1988). Marketing data analysis by dual scaling. *International Journal of Research in Marketing*, 5, 151-170.

Nishisato, S. & Gaul W. (1990). An approach to marketing data analysis: The forced classification procedure of dual scaling. *Journal of Marketing Research* 27, 354-360

Nishisato, S. & Sheu, W. J. (1984). A note on dual scaling of successive categories data. *Psychometrika*, 49, 493-500.

Odondi, M. (1997). Multidimensional analysis of successive categories (rating) data by dual scaling. Unpublished doctoral dissertation, University of Toronto.

Richardson, M., & Kuder, G. F. (1933). Making a rating scale that measures. *Personnel Journal*, 12, 36-40.

Tenehaus, M. & Young, F. W. (1985). An analysis and synthesis of multiple correspondence analysis, optimal scoring, dual scaling, homogeneity analysis and other methods for quantifying categorical data. *Pyschometrika*, 50, 91-119.

van de Velden, Michel (2000). Dual scaling and correspondence analysis of rank order data. *Innovation in Multivariate Statistical Analysis: A Festschrift for Heinz Neudecker*, 36, 87-99.

Yanai, H. & Maeda, T. (2002). Partial multiple correspondence analysis, In S. Nishisato, Y. Baba, H. Bozdogan and K. Kanefuji (Eds.). *Measurement and Multivariate Analysis* (pp. 57-68). Tokyo: Springer-Verlag.

Appendix A

Each Computer Program in Appendix A was written in the R programming language and should be used with R Version 2.0.1 or higher.

Successive-Categories Format Program

```
SuccFormat<-function(data1,NumStim,NumCat)
#Written by Adam E. Wyse, 2005
#takes in data1 (original data), NumStim (Number of Stimulus), and Number of Categories from the User
#uses notation and methods that are consistent with Nishisato (1994)

{
#Note this algorithm needs to be altered if you want to handle data with missing responses

CatBound<-NumCat-1
nr1<-nrow(data1) #extracts the number of rows from data1
nc1<-ncol(data1) #extracts the number of columns from data1
TotalDim<-NumStim+CatBound #Adds Number of Stimilus and Category Boundary #Together which is my Total
Dimension

Rating<-matrix(0,nr1,(NumCat)) #finds number of stimulus that each subject put at each specific rating

SuccData<-matrix(0,nr1,TotalDim) #This will store and output my Successive-Categories Data in Rank Order
#Format

Rank<-0 #Variable that keeps track of ranks
Holder<-0 #Used to Hold the Rank when we have more than one ranking for a particular respondent

F<-0 #Used to Hold the Final Rank, this variable is necessary due to the multiple responses in a given category

DMatrix<-matrix(0,nr1,TotalDim) #This will store and output the Dominance Matrix

for (i in 1:nr1) #runs through the matrix to figure out how many are selected at each rating

{
for (j in 1:nc1)
{
Rating[i,(data1[i,j])]<-1+Rating[i,(data1[i,j])]
}
}

for (i in 1:nr1)
#Loops through the Original matrix assigning the ranks to the stimuli and category boundaries
#It runs through the matrix up to the Number of Categories (CatBound +1)
#If statements tell the matrix how to assign the ranks to the dominance matrix
{
for (k in 1:(CatBound+1))
{
```

```
for (j in 1:nc1)

#Case when we have 0 in the rating category and we aren't in the last rating category
if (Rating[i,k]==0 &(k!=(CatBound+1)))
{
Rank<-F
Rank<-Rank+1/nc1
SuccData[i,k]<-Rank
F<-Rank
}
#Case when we have only assigned one stimulus to a particular rating category
if (Rating[i,k]==1)
{

#One Stimulus and not the Last Rating Category

if ((data1[i,j]==k) & (k!=(CatBound+1)))
{
Rank<-F
Rank<-((Rating[i,k]+1)/2)+Rank
SuccData[i,(j+CatBound)]<-Rank
Rank<-Rank+1
SuccData[i,k]<-Rank
F<-Rank
}

#One Stimulus and the Last Rating Category
if ((data1[i,j]==k) & (k==(CatBound+1)))
{
Rank<-F
Rank<-((Rating[i,k]+1)/2)+Rank
SuccData[i,(j+CatBound)]<-Rank
}
}

#Case when we assigned more than One Stimulus to a Rating Category
if (Rating[i,k]>1)
{
#More than One Stimulus First Category
if ((data1[i,j]==k) & (k==1))
{
Holder<-Rank
Rank<-((Rating[i,k]+1)/2)+Rank
SuccData[i,(j+CatBound)]<-Rank
Rank<-Rank-((Rating[i,k]+1)/2)
SuccData[i,k]<-(Holder+(Rating[i,k]+1))
F<-SuccData[i,k]
}
#More than One Stimulus and it isn't the First or Last Category

if ((data1[i,j]==k) & ((k>1)&(k<(CatBound+1))))
{
Rank<-SuccData[i,(k-1)]
Holder<-Rank
Rank<-((Rating[i,k]+1)/2)+Rank
SuccData[i,(j+CatBound)]<-Rank
```

```
Rank<<-Rank-((Rating[i,k]+1)/2)
SuccData[i,k]<-(Holder+(Rating[i,k]+1))
F<-SuccData[i,k]
}
#More than One Stimulus and it is the Last Category
if ((data1[i,j]==k) & (k==(CatBound+1)))
Rank<-SuccData[i,(k-1)]
Rank<-((Rating[i,k]+1)/2)+Rank
SuccData[i,(j+CatBound)]<-Rank
Rank<-Rank-((Rating[i,k]+1)/2)
}
}
}
}
#Resets the Rank and F(Final Rank)
Rank<-0
F<-0

}


#Gets the Dominance Matrix for this set of data

DMatrix<-2*SuccData-(NumStim+CatBound+1)
DMatrix



}
```

Dual Scaling Computations Program

```
DS_Command<-function(F,n=0,c=0,N=0,type=c("MC","S","R","PC","SC","CT",""),nsolutions)
#written by Adam E. Wyse
#F is the converted matrix (dominance or incidence format) that needs to be inputted
#n is the number of items
#c is the number of categories
#N is the number of subjects
#type specifies what data type we are putting in
#nsolutions is the number of solutions that the user wants to display
#This procedure was developed in accordance with Methods found in Nishisato's "Elements of Dual Scaling" (1994)

{
m<-c-1 #finds the number of category boundaries
fc<-matrix(colSums(F),ncol(F),1) #fc dual scaling matrix
fr<-matrix(rowSums(F),nrow(F),1) #fr dual scaling matrix
Dr<-matrix(0,nrow(F),nrow(F)) #Dr dual scaling matrix set to zeros
Dc<-matrix(0,ncol(F),ncol(F)) #Dc dual scaling matrix set to zeros
trf<-function(x) #computes the trace
{sum(diag(x))}


if ((type=="MC")|(type=="S")) #Multiple Choice or Sorting Format
{
diag(Dr)<-n
diag(Dc)<-(fc)
ft<-trf(Dr)
}

 else if ((type=="R")|(type=="PC")) #Ranking or Paired Comparison
{
diag(Dr)<-n*(n-1)
diag(Dc)<-N*(n-1)
ft<-N*n*(n-1)
}

 else if (type=="SC") #Successive Categories
{
diag(Dr)<-(n+m)*(n+m-1)
diag(Dc)<-N*(n+m-1)
ft<-N*(n+m)*(n+m-1)
}

else #Anything else, includes contingency tables
{
diag(Dr)<-(fr)
diag(Dc)<-(fc)
ft<-trf(Dr)
}

Power<-function(data1,power) #finds the power of a matrix
{
Y<-data1^(power)
Y[is.infinite(Y)]<-0 #assigns zero when it is infinite
Y
```

```
}

Dchalf<-Power(Dc,.5) #Power is used to raise Dc to the 1/2
Drneghalf<-Power(Dr,-.5) #Power is used to raise Dr to the -1/2
Dcneghalf<-Power(Dc,-.5) #Power is used to raise Dc to the -1/2
B<-Drneghalf%*%F%*%Dcneghalf #Computes the B matrix
tranB<-t(B) #Computes the transpose of the B matrix
BprimeB<-tranB%*%B #multiples the transpose of B and B
colvect<-matrix(1,ncol(F),1) #makes a column vector of 1's
rowvect<-matrix(1,1,ncol(F)) #makes a row vector of 1's
topC1<-Dchalf%*%colvect%*%rowvect%*%Dchalf #Computes denominator of the subtracted part

A<-topC1/ft #finds the subtracted part
Drnegone<-Power(Dr,-1) #Power is used to raise Dr to the -1
Dcnegone<-Power(Dc,-1) #Power is used to raise Dc to the -1
CMatrix<-BprimeB-A #Computes the C1 matrix eliminating trivial solution

if ((type=="R")|(type=="PC")|(type=="SC"))
#these types of data don't have the trivial solution II so we don't need to subtract off the correct
{
CMatrix<-BprimeB
}

s<-svd(CMatrix) #Singular-Value Decomposition of Matrix c1
newsquare<-s$d #equals the diagonal component from svd
xvector<-s$u #x vector from Singular-Value Decomposition of Matrix c1
xweights<-xvector*(sqrt(ft)) #assign initial x weights

# (Note: R normalizes vectors, so this is not needed in this program)

dualx<-Dcneghalf%*%xweights #obtain dual x weights for dual scaling
P<-(1/sqrt(newsquare))#new value
optimalY<-Drnegone%*%F%*%dualx #place y value in matrix off by a constant

percentexplain<-matrix(0,1,ncol(optimalY)) #holds the percent of information explained in a matrix

holdpercent<-0 #a place holder for the percent matrix that is set equal to 0

for (i in 1:ncol(optimalY)) #obtain correct optimal y weights for dual scaling
{
for (j in 1:nrow(optimalY))
{
optimalY[j,i]<-P[i]*optimalY[j,i]

}
percentexplain[1,i]<-holdpercent+(100*newsquare[i])/trf(CMatrix) #calculates percent explained
holdpercent<-percentexplain[1,i] #reset the holdpercent
}
optimalX<-dualx #we know that x should equal the dual x per the dual relations


#these commands create user specified output according to the number of solutions
if (nsolutions>ncol(F)) #user specified too many solutions
{
xdisplay<-optimalX
ydisplay<-optimalY
```

```
percentdisplay<-percentexplain
newdisplay<-matrix(newsquare,1,ncol(F))
}
else #user wants to extract a number solutions less than original dimension
{
xdisplay<-matrix(0,nrow(optimalX),nsolutions)
ydisplay<-matrix(0,nrow(optimalY),nsolutions)
percentdisplay<-matrix(0,1,nsolutions)
newdisplay<-matrix(0,1,nsolutions)

for (i in 1:nrow(optimalY)) #creates the correct display
{

for(j in 1:nsolutions)
{
ydisplay[i,j]<-optimalY[i,j]
percentdisplay[1,j]<-percentexplain[1,j]
newdisplay[1,j]<-newsquare[j]
}
}
for (i in 1:nrow(optimalX)) #creates the correct display
{
for(j in 1:nsolutions)
{

xdisplay[i,j]<-dualx[i,j]

                                                  }
}
# it is important to note that optimal x and optimal y usually have different dimensions


}

list(round(F,digits=10),round(xdisplay,digits=3),round(ydisplay,digits=3),round(percentdisplay,digits=3),signif(new
display, digits=3))
#outputs the original data, weights, eta, and percentexplained in proper format
#uses the round function to display the data in proper format
#round is used on F to erase the use of exponential notation on the conversion programs that R sometimes employs


}
```

Program to Create F* Matrix

```
SetClass<-function(data1,data2,n,NumCat)
#Written by Adam E. Wyse, 2005
#This function takes in as input the original data,
#the demographic variable, the number of stimulus,
#the number of categories
#**** This function is used to include ten new categories between a set of original data
#    and a demo variable. A centering matrix is also added to the original matrix.
{
NewCat<-10 #ten new categories added
m<-NumCat-1 #number of category boundaries in the original data
CatBound<-m+2+2*NewCat #catculates the new number of category boundaries between demo variables
demo1<-1 #sets Demo 1 to 1
demo2<-1+CatBound #sets Demo 2
Add<-demo1+NewCat
#This calculates the number that needs to be added to each element of
# the original data matrix. It is also the number of new category boundaries
# either above or below the centered data.
AddMatrix<-matrix(Add,nrow(data1),ncol(data1)) #Matrix that will be added to original data matrix
TransData<-data1+AddMatrix #Transforms Original Data Matrix
for (j in 1:nrow(data2)) #replaces all those classified as 2 with group2
{
if (data2[j,1]!=1)
{
data2[j,1]<-demo2
}
}
OutData<-cbind(TransData,data2) #Augments Transformed Data with demographic column
OutData
}
```

Appendix B

Explanation of How to Use Computer Programs

The user must first create a file that contains the successive-categories data that is to be analyzed and a second file that contains the demographic data. (These should be text files with a "txt" extension.) The user then reads the data into R using the *Matrix* and *Scan* functions selecting specific options of these functions to ensure that data are read in properly. (See below for an example.) The successive-categories data and the demographic data will be read in separately, and after the data are entered, the *SetClass* function constructs the $\mathbf{F}^*$ Matrix. The *SuccFormat* function is then run to create the dominance matrix $\mathbf{E}^*$ for the analysis to follow. Finally, the *DS_Command* function, with $\mathbf{E}^*$ as its input, does the actual dual scaling analysis. *DS_Command* displays the dominance matrix ([[1]]), the item weights and category boundaries (with category boundaries listed first) ([[2]]), the respondent scores ([[3]]), the cumulative percent of variation explained ([[4]]), and the squared correlation ratio ([[5]]). If the user is interested only in a standard analysis of the original data, then the *SetClass* function can be ignored. It is important to note that each function must be added to the workspace before it can be used in the analysis. This can be done by copying the program from a script and hitting the "F5" key. A sample session of how to use the programs follows.

Sample Session

```
> A1<-matrix(scan("thesis_example.txt"),ncol=6,byrow=TRUE) #Reads in the successive-categories data

Read 60 items

> A2<-matrix(scan("demo1.txt"),ncol=1,byrow=TRUE) #Reads in the demo variable

Read 10 items

> A1 #Displays the successive-categories data

      [,1] [,2] [,3] [,4] [,5] [,6]
 [1,]    4    1    3    4    2    3
 [2,]    3    3    2    1    2    3
 [3,]    3    2    3    2    4    2
 [4,]    4    1    4    3    2    4
 [5,]    3    3    3    4    3    1
 [6,]    4    4    3    4    2    2
 [7,]    2    2    2    4    3    1
 [8,]    4    3    2    3    2    4
 [9,]    1    2    3    2    4    3
[10,]    2    1    1    1    3    3
> A2 #Displays the demo variable

      [,1]
 [1,]    2
 [2,]    1
 [3,]    1
 [4,]    2
 [5,]    2
 [6,]    2
 [7,]    1
 [8,]    2
 [9,]    1
[10,]    2

> A3<-SuccFormat(A1,6,4) #Changes original data to the dominance matrix

> A4<-DS_Command(A3,6,4,10,"SC",5) #Finds the first 5 solutions for the dual scaling analysis of the dominance matrix A3
```

> A4

[[1]] #Dominance matrix for original data

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
 [1,]   -6   -2    4    7   -8    1    7   -4    1
 [2,]   -6    0    8    4    4   -3   -8   -3    4
 [3,]   -8    0    6    3   -4    3   -4    8   -4
 [4,]   -6   -2    2    6   -8    6    0   -4    6
 [5,]   -6   -4    6    1    1    1    8    1   -8
 [6,]   -8   -2    2    6    6    0    6   -5   -5
 [7,]   -6    2    6   -2   -2   -2    8    4   -8
 [8,]   -8   -2    4    7    1   -5    1   -5    7
 [9,]   -6    0    6   -8   -3    3   -3    8    3
[10,]   -2    2    8    0   -6   -6   -6    5    5
```

[[2]] # Displays item weights and category boundaries, where the first three rows are category boundaries and the last six rows are the items

```
       [,1]   [,2]   [,3]   [,4]   [,5]
 [1,] -2.095  0.111 -0.170 -0.391 -0.810
 [2,] -0.355 -0.359  0.254  0.158 -0.562
 [3,]  1.595 -0.810  0.589  0.758 -0.545
 [4,]  1.092  0.482 -1.443 -0.092  0.246
 [5,] -0.626  0.825 -0.270  2.324  0.919
 [6,] -0.015  0.224  0.297 -1.376  2.276
 [7,]  0.583  2.009  0.457 -0.825 -1.217
 [8,] -0.100 -0.990  1.884 -0.047  0.091
 [9,] -0.080 -1.492 -1.599 -0.509 -0.398
```

[[3]] #Respondent scores

```
       [,1]   [,2]   [,3]   [,4]   [,5]
 [1,] 1.372  0.461 -0.512 -1.438 -0.864
 [2,] 0.843 -0.965 -0.847  1.739  0.608
 [3,] 1.093 -0.754  1.061 -0.071  1.398
 [4,] 1.042 -0.398 -0.964 -1.775  0.971
 [5,] 1.092  1.090  0.979  0.302  0.080
 [6,] 1.041  1.360 -0.331  1.033  0.758
 [7,] 0.951  0.657  1.525  0.185 -1.233
 [8,] 1.176 -0.170 -1.378  0.681 -0.760
 [9,] 0.465 -1.319  1.278 -0.192  0.673
[10,] 0.586 -1.701  0.195  0.242 -1.690
```

[[4]] #Cumulative Percent Explained

```
      [,1]   [,2]   [,3]   [,4]   [,5]
[1,] 34.414 58.638 79.653 90.554 95.576
```

[[5]] #Squared Correlation Ratio

```
      [,1]   [,2]   [,3]   [,4]   [,5]
[1,] 0.138 0.0971 0.0843 0.0437 0.0201
```

> A5<-SetClass(A1,A2,6,4)

> A5 #Displays **F***

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
 [1,]   15   12   14   15   13   14   26
 [2,]   14   14   13   12   13   14    1
 [3,]   14   13   14   13   15   13    1
 [4,]   15   12   15   14   13   15   26
 [5,]   14   14   14   15   14   12   26
 [6,]   15   15   14   15   13   13   26
 [7,]   13   13   13   15   14   12    1
 [8,]   15   14   13   14   13   15   26
 [9,]   12   13   14   13   15   14    1
[10,]   13   12   12   12   14   14   26
```

> A6<-SuccFormat(A5,7,26) #Creates dominance matrix for A5

> A7<-DS_Command(A6,7,26,10,"SC",5)) #Finds the first 5 solutions for the dual scaling analysis of the dominance matrix A3

> A7[[2]] #Displays item weights and category boundaries, where the first 25 rows are category boundaries and the last seven rows are items

```
         [,1]   [,2]   [,3]   [,4]   [,5]
 [1,] -1.726 -0.056 -0.007 -0.014 -0.030
 [2,] -1.612 -0.064 -0.007 -0.015 -0.029
 [3,] -1.498 -0.072 -0.007 -0.015 -0.028
 [4,] -1.384 -0.080 -0.006 -0.015 -0.027
 [5,] -1.269 -0.088 -0.006 -0.015 -0.026
 [6,] -1.155 -0.096 -0.006 -0.016 -0.025
 [7,] -1.041 -0.104 -0.006 -0.016 -0.024
 [8,] -0.927 -0.112 -0.006 -0.016 -0.023
 [9,] -0.812 -0.120 -0.006 -0.016 -0.023
[10,] -0.698 -0.128 -0.005 -0.017 -0.022
[11,] -0.584 -0.136 -0.005 -0.017 -0.021
[12,] -0.367 -0.112  0.534  0.872  0.724
[13,] -0.060 -0.369  0.447  0.438  0.617
[14,]  0.283 -0.395  0.438  0.399  1.426
[15,]  0.559 -0.217 -0.003 -0.019 -0.012
[16,]  0.673 -0.225 -0.003 -0.020 -0.011
[17,]  0.787 -0.233 -0.003 -0.020 -0.010
[18,]  0.901 -0.241 -0.003 -0.020 -0.009
[19,]  1.016 -0.249 -0.003 -0.020 -0.008
[20,]  1.130 -0.257 -0.002 -0.021 -0.007
[21,]  1.244 -0.265 -0.002 -0.021 -0.006
[22,]  1.358 -0.273 -0.002 -0.021 -0.005
[23,]  1.473 -0.282 -0.002 -0.021 -0.005
[24,]  1.587 -0.290 -0.002 -0.022 -0.004
[25,]  1.701 -0.298 -0.002 -0.022 -0.003
[26,]  0.128  0.293  0.115 -2.671 -1.915
[27,] -0.122 -0.260 -1.170 -3.142  3.023
[28,] -0.025 -0.241 -0.618  0.902 -3.377
[29,]  0.042  0.220 -3.792  0.316 -1.108
[30,]  0.012 -0.751  0.116  3.476  1.506
[31,] -0.006  0.023  3.892 -0.753 -1.043
[32,]  0.393  5.480  0.133  0.562  0.503
```

> A7[[3]] #Respondent scores

```
       [,1]   [,2]   [,3]   [,4]   [,5]
 [1,] 1.020  0.820 -0.258  0.216 -1.357
 [2,] 0.972 -1.221  0.963 -1.920  0.388
 [3,] 0.975 -1.256 -0.032  0.395 -0.628
 [4,] 1.018  0.808  0.748  0.229 -1.946
 [5,] 1.018  0.775 -1.446  0.585  0.808
 [6,] 1.017  0.807 -1.158 -1.121  0.540
 [7,] 0.973 -1.234 -1.457  0.336  0.145
 [8,] 1.019  0.813  0.656 -1.160  0.375
 [9,] 0.971 -1.275  0.498  1.094  0.032
[10,] 1.016  0.741  1.488  1.343  1.649
```

> A7[[4]] #Cumulative Percent Explained

```
       [,1]   [,2]   [,3]   [,4]   [,5]

[1,] 89.889 98.815 99.313 99.624 99.816
```

Appendix C

QUESTIONNAIRE

Student-Life Stress Inventory

Please circle your current stress level:

Mild	Moderate	Severe
1	2	3

For all the questions below please respond by circling the correct response using the following scale:

1 = never 2 = seldom 3 = occasionally 4 = often 5 = most of the time

STRESSORS

A. As a student (frustrations):

1. I have experienced frustrations due to delays in reaching my goals.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

2. I have experienced daily hassles which affected me in reaching my goals.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

3. I have experienced lack of sources (money for auto, books, etc.).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

4. I have experienced failures in accomplishing the goals that I set.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

5. I have not been accepted socially (became a social outcast).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

6. I have experienced dating frustrations.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

7. I feel I was denied opportunities in spite of my qualifications.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

B. I have experienced conflicts which were:

8. Produced by two or more desirable alternatives.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

9. Produced by two or more undesirable alternatives.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

10. Produced when a goal had both positive and negative alternatives.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

C. I experienced pressures:

11. As a result of competition (on grades, work, relationships with spouse and/or friends).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

12. Due to deadlines (papers due, payments to be made, etc.).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

13. Due to an overload (attempting too many things at one time).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

14. Due to interpersonal relationships (family and/or friends, expectations, work responsibilities).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

D. I have experienced (changes):

15. Rapid unpleasant changes.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

16. Too many changes occurring at the same time.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

17. Change which disrupted my life and/or goals.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

E. As a person (self-imposed):

18. I like to compete and win.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

19. I like to be noticed and be loved by all.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

20. I worry a lot about everything and everybody.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

21. I have a tendency to procrastinate (put off things that have to be done).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

22. I feel I must find a perfect solution to the problems I undertake.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

23. I worry and get anxious about taking tests.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

II. REACTIONS TO STRESSORS:

F. During stressful situations, I have experienced the following (physiological):

24. Sweating (sweaty palms, etc.).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

25. Stuttering (not being able to speak clearly).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

26. Trembling (being nervous, biting fingernails, etc.).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

27. Rapid movements (moving quickly, from place to place).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

28. Exhaustion (worn out, burned out).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

29. Irritable bowels, peptic ulcers, etc.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

30. Asthma, bronchial spasm, hyperventilation.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

31. Backaches, muscle tightness (cramps), teeth-grinding.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

32. Hives, skin itching, allergies.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

33. Migraine headaches, hypertension, rapid heartbeat.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

34. Arthritis, over-all pains.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

35. Viruses, cold, flu.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

36. Weight loss (can't eat).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

37. Weight gain (eat a lot).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

G. When under stressful situations, I have experienced (emotional):

38. Fear, anxiety, worry.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

39. Anger.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

40. Guilt.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

41. Grief, depression.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

H. When under stressful situations. I have (behavioral):

42. Cried.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

43. Abused others (verbally and/or physically).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

44. Abused self (used drugs, etc.).

1 never 2 seldom 3 occasionally 4 often 5 most of the time

45. Smoked excessively.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

46. Was irritable towards others.

1 never 2 seldom 3 occasionally 4 often 5 most of the time

47. Attempted suicide.

1 never　　2 seldom　　3 occasionally　　4 often　　5 most of the time

48. Used defense mechanisms.

1 never　　2 seldom　　3 occasionally　　4 often　　5 most of the time

49. Separated myself from others.

1 never　　2 seldom　　3 occasionally　　4 often　　5 most of the time

I. With reference to stressful situations, I have (cognitive appraisal):

50. Thought about and analyzed how stressful the situations were.

1 never　　2 seldom　　3 occasionally　　4 often　　5 most of the time

51. Thought and analyzed whether the strategies I used were most effective.

1 never　　2 seldom　　3 occasionally　　4 often　　5 most of the time

Appendix D

Student Demographics Page

Please answer the following questions about yourself so that we can better interpret the information that we are collecting for this study. Every answer given is confidential.

Please circle or fill in the blank with the correct response:

1. Gender: Male (1) Female (2)

2. Academic Standing: Freshman (1) Sophomore (2) Junior (3) Senior (4) Grad (5)

3. GPA: 1.5 1.9 (1) 2.0 2.4 (2) 2.5 2.9 (3) 3.0 - 3.4 (4) 3.5 - 4.0 (5)

4. Age: 17 (1) 18 (2) 19 (3) 20 (4) 21 (5) 22 (6) 23 (7) 24 (8) 25 (9) Other: __________ (10)

5. College: CAST (1) CCIS (2) CIAS (3) COB (4) COE (5) COLA (6) COS (7) NTID (8)

6. Major(s):__________________________________

7. Ethnicity: African American (1) Asian (2) Caucasian (3) Latino (4) Native American (5) Other (6)

8. Employment status: employed (1) unemployed (2)

9. Marital status: single (1) married (2) divorced (3)

10. Have you had an appointment at the Student Health Center in the last 6 months?

Yes (1) No (2) Don't Know (3)