

GENERALIZED PROCRUSTES ANALYSIS: A TOOL FOR EXPLORING AGGREGATES AND PERSONS

James W. Grice
Oklahoma State University

Kimberly K. Assad
Michigan State University

ABSTRACT

Gower (1975) introduced *Generalized Procrustes Analysis* (GPA) as a multivariate statistical technique for analyzing three-dimensional data matrices. The current paper presents a non-technical introduction to the logic underlying GPA and then presents a completely worked example using genuine data. Specifically, self and peer ratings obtained from students attending a Summer Science Academy are analyzed and discussed. It is shown that GPA offers a powerful set of tools for exploring data at both the aggregate and individual level. A number of issues regarding the current analysis methods are also discussed.

INTRODUCTION

Imagine a female student who rates herself and nine other people whom she knows personally on 30 adjective trait terms (e.g., friendly, outgoing, agreeable) using a 5-point Likert-type scale. A personality psychologist subsequently records her responses in a 30 x 10 2-dimensional matrix:

Person #1

1	1	5		4
2	5	2		1
3	4	1		5
.				
.				
30	4	3		2
	1	2	...	10

Rated individuals

The psychologist next recruits more people to participate in the study, obtaining 49 additional 30 x 10 matrices:

GENERALIZED PROCRUSTES ANALYSIS

Unlike the personality psychologist, the market researcher is confronted with data that are matched on one dimension (cereals) but not matched on the other (the personal descriptors that differ across participants). Nonetheless, the goal of the researcher is to determine which breakfast cereals are alike and most similar to the ideal breakfast cereal. How is the researcher to reach this goal?

Gower's (1975) approach to the type of data gathered by the personality and market researchers above involved an inventive combination of Procrustes rotation and Analysis of Variance (ANOVA), which he referred to as *Generalized Procrustes Analysis*. The essence of this analysis can best be described using a simple data set:

		Cereal A	Cereal B	Cereal C	Ideal Cereal
Joe	<i>Sweet</i>	-3	-2	2	3
	<i>Crunchy</i>	1	2	2	2
Mary	<i>Expensive</i>	-2	-3	-2	-1
	<i>Healthy</i>	-2	-1	2	2

Joe and Mary elicited their own terms to describe breakfast cereals, and then they rated three cereals and the “ideal cereal” on 7-point Likert-type scales (scored -3 to +3) constructed from their unique descriptors. The ratings for Joe and Mary can be plotted separately, as shown in Figure 1. For these data the four cereals are matched and the two dimensions are not matched. The goal of Generalized Procrustes Analysis (GPA) is to maximally align the multivariate configurations of the matched figures in the two spaces. Examination of Figure 1 clearly reveals that both Joe and Mary viewed cereals A and B as highly similar to each other and distinct from cereal C and the ideal cereal. These latter cereals, C and the ideal, are themselves rated as highly similar. The key to GPA is to treat the unmatched dimensions as essentially arbitrary, and then rotate the two spaces to maximum similarity using Procrustes rotation (Cliff, 1966; Schönemann, 1966). In this example, rotating the axes in Joe's graph approximately -90° would align the four cereals very closely with the cereals in Mary's 2-dimensional space.

The rotated alignment, however, between the two sets of ratings would not be perfect. The next major feature of GPA thus turns toward quantifying the degree of similarity between the two rotated multivariate configurations of cereals. Gower (1975) accomplished this goal by computing a *consensus matrix*, which is simply an average of the rotated matrices of ratings. With this matrix of means in hand, the techniques underlying ANOVA can then be used to partition the total amount of variability in the rotated ratings into two unique portions: *consensus* and *residual*. If Joe's and Mary's rotated ratings match the consensus matrix closely, the lion's share of the variance will be allocated to the consensus term. If the rotated ratings are quite discrepant, then the residual variance term will be relatively large. In order to quantify the magnitude of effect—that is, agreement among raters regarding the multivariate configuration of matched figures—a *consensus proportion* can furthermore be computed as the ratio of the consensus

APPLIED MULTIVARIATE RESEARCH

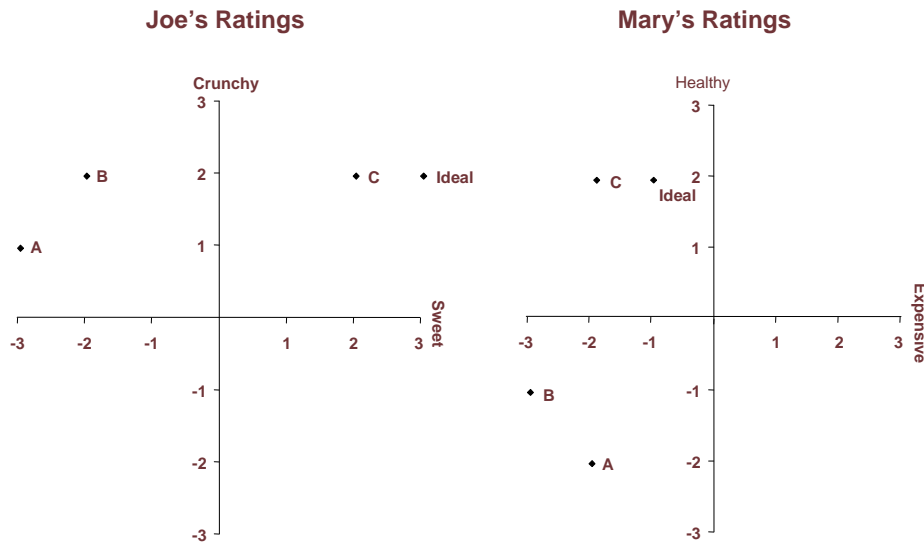


Figure 1. Bi-plots of two individuals' ratings of four cereals.

sum of squares to the total sum of squares in the ANOVA. The consensus proportion is thus kin to the traditional multiple R^2 statistic in regression, which ranges in value from 0 to 1 and indicates the proportion of overlap among the rotated ratings. It can be tested for statistical significance most efficiently using a randomization test (Wakeling, Raats, & Halliday, 1992). The result for Joe and Mary was equal to .99 ($p < .001$), indicating near-perfect similarity in their multivariate configurations of the cereals, as can be clearly seen in Figure 1. In addition to this useful summary statistic, a number of plotting routines can also be used to visualize the configuration of the matched figures in the consensus (average) matrix (see Gabriel, 1971; Gower & Dijksterhuis, 1994; Slater, 1977). These techniques typically revolve round Principal Components Analysis or similar multivariate statistical techniques.

In addition to these analyses and statistics that are based on aggregated data, what is perhaps most impressive about GPA is the ability to focus on the persons in the analysis. This is typically done in two different ways. First, each person's complete matrix of rating values can be related to the consensus matrix, and the degree of similarity between the two quantified. Particular individuals who deviate greatly from the consensus can thus easily be identified. Moreover, the ANOVA procedures discussed above can be refined so that specific points of discrepancies among the raters can be identified. For instance, if Joe and Mary disagreed primarily in their views of 'cereal A', the refined ANOVA results would help the investigator identify this difference in the data. Second, the general and specific points of departure from the consensus matrix can be examined using powerful graphing routines. The multivariate configuration of a particular person's matrix of ratings can be plotted and compared to the consensus configuration to visually examine points of disparity. Additionally, all of the persons' particular dimensions can be plotted simultaneously with the consensus configuration, even if the rating scales are not matched as with Joe ("sweet", "crunchy") and Mary ("expensive", "healthy") above. These plotting tech-

GENERALIZED PROCRUSTES ANALYSIS

niques again revolve around Principal Components Analysis and another multivariate technique referred to as extension analysis (see Grice, 2007).

In order to demonstrate the aggregate and person-centered features of GPA, we present data below collected from high school students attending a 1-week Summer Science Academy at Oklahoma State University. These students shared a common living space and participated in numerous activities over the course of the academy. At the end of the week we asked the students to rate themselves and their peers on scales constructed from their own personal constructs and scales constructed from adjectives for the Big Five personality traits. The data were therefore similar to the hypothetical studies described at the beginning of this manuscript. Did the high school students reach a consensus with regard to themselves and their peers on the Big Five traits? Did the average, multivariate configuration of the students' views of each other derived from the personal constructs match the configuration derived from the Big Five traits? Were particular students discrepant in their views of themselves and their peers compared to everyone else's views? In the analysis below, these questions will be addressed, and a number of the specific details regarding GPA will be introduced and discussed.

Method

Participants

Twenty-five high school students participated in the Summer Science Academy at Oklahoma State University, a program designed for Oklahoma high school students with the intention of developing interest in psychological research. Of these students, twenty (5 males, 15 females) voluntarily participated in this study after having their parents or guardians sign a consent form. Participants ranged in age from 13 to 16 years ($M = 14.55$, $Mdn = 14.00$, $SD = 1.10$) and were obtained from across Oklahoma. Six participants were Caucasian, 4 African American, 7 Asian, 2 Native American, and 1 was Hispanic. All participants were tested individually on computers, and did not take more than 2 hours to complete the experiment. One participant, however, failed to complete all of the procedures appropriately, and her data were consequently removed from the analyses.

Procedures

Personal Construct grid. Each participant was initially handed a sheet containing both labeled pictures of 21 students in the Summer Science Academy, and 4 spaces to induce cues of the 4 students who did not have their pictures taken, to insure the participants remembered the other students' names. Each participant was taken individually into a private workspace to complete the experiment. The "ideal self" and the names of the 25 students were entered into Idiogrid—software for managing and analyzing repertory grids and other types of self-report, personality data (Grice, 2002). The participants were then administered a sentence completion task developed by Grice, et al. (2004) to elicit their bipolar personal constructs. For instance, the participants were required to complete the following sentence with a single word or short phrase: "Generally speaking, I am the type of person who is _____." After entering a response, the participants were immediately asked to enter the opposite word or phrase, thus yielding bipo-

APPLIED MULTIVARIATE RESEARCH

lar personal constructs (e.g., happy-sad, shy-outgoing). After eliciting the 10 personal constructs, participants rated (in an individually determined random order) each of the 26 figures (viz., ideal self, and 25 individuals, including themselves) on each construct using a 5-point Likert-type scale ranging from “Very Inaccurate” to “Very Accurate.” Each item stem was displayed as “____ is the type of person who is ____” (as opposed to ____). The first blank contained the figure being rated, and the other 2 blanks contained the poles of a personal construct; for example, “Robert is the type of person who is shy (as opposed to outgoing).” The 260 ratings for each participant were recorded in a 10 x 26 (constructs x figures) matrix, or grid, for further analysis in Idiogrid, and the names of the 25 students were changed to pseudonyms to protect confidentiality in the analyses and results reported below.

Big Five grid. For each participant the 26 figures from the personal construct grid were loaded into Idiogrid. The participants then rated each figure on 20 Big Five trait descriptors. The descriptors were selected from Goldberg’s list of 100 lexical marker items (Goldberg, 1999) and altered slightly to fit the item stems. Four items were selected for each of the Big Five Personality traits, with two of the four items for each trait being reverse-keyed. The figures were then rated (in an individually determined random order) by each participant using the same 5-point Likert scale. Each item stem was presented as “____ is the type of person who ____.” The first blank contained the figure being rated, and the second blank contained the Big Five descriptor being considered; for example, “Robert is the type of person who thinks of others first.” The 520 ratings were recorded in a 20 x 26 (Big Five descriptor traits x figures) grid for subsequent analysis, and the names of the 25 students were again changed to the same pseudonyms to protect confidentiality.

Results

In the personal construct grids the constructs differed from participant to participant, while 25 of the 26 figures were matched across grids. Recall that each participant rated the other 24 students in the Summer Science Academy as well as himself or herself and the ideal self. Since the ideal self was not strictly matched across grids, it was removed from both the personal construct and Big Five grids. Given the remaining figures were matched across grids the participants views of themselves and their peers could then be examined.

Grid Pre-Scaling

The first decision that a researcher must make when utilizing GPA regards the scaling of the original grid ratings. Three pre-scaling options can be considered that remove systematic sources of variation between raters (for a review see Dijksterhuis & Gower, 1991). The first pre-scaling option is referred to as *dimensional scaling*, and it can be applied when the grids or matrices of ratings differ with respect to their dimensions. If this scaling is applied each grid value will be divided by the square root of the number of non-matched figures in the grid, effectively reducing the variability of the ratings in large grids. In this way larger grids will not have a statistical ‘advantage’ over smaller grids in the ANOVA portion of the results that is due simply to their size. In this study each participant

GENERALIZED PROCRUSTES ANALYSIS

rated his or her peers on 10 personal constructs. The non-matched figures (viz., the personal constructs) were therefore equal in number across all of the grids, and dimensional scaling was not necessary.

The second pre-scaling option, referred to as *centering scaling*, essentially removes differences in scale means from the ratings. This scaling was applied to the current grids in Idiogrid; consequently, the grid values for each participant were subtracted from their respective construct means prior to being submitted to the GPA algorithm. The mean of each centered construct was therefore set equal to zero, and potential individual differences in construct means was thus removed.

The third pre-scaling option, *isotropic scaling*, was also applied to the current grids. This scaling technique essentially “shrinks” or “stretches” the original grid ratings to remove individual differences in scale usage. For example, if one participant endorses many different points on the entire scale (5 scale points) while another participant endorses primarily the mid-point of the scale, this difference may impact the ANOVA portion of the results. Applying isotropic scaling to these two participants’ grids would decrease the overall variability of the first grid and increase the overall variability of the second grid. The computation of isotropic scaling values is slightly involved, and descriptions of the procedures can be found in ten Berge (1977) and Gower (1975). The computed isotropic scaling values will be centered around unity, with values greater than unity indicating that the original ratings were stretched (that is, their variability was increased). Values less than unity indicate that the original ratings were shrunk (that is, their variability was decreased). As can be seen in Table 1, the isotropic scaling values for the current 19 participants’ personal construct grids show the ratings for Eric were stretched the most compared to the other 18 participants. Examination of Eric’s original grid ratings in fact revealed that he never used the end-points of the rating scale. Rhonda’s ratings were also stretched because she never endorsed the two lowest points on the rating scale. These individual differences may be of interest in another context, but the current goal is to examine the multivariate configuration of the rated people in the personal construct grids. Consequently, the isotropic scaling option was applied to essentially remove this systematic source of between-participant variability.

The Consensus Matrix

The central feature of GPA is undoubtedly the consensus grid, or matrix, which records the means computed from the rescaled and rotated original grids. The grids are rescaled according to the investigator’s chosen options (see above), and then rotated to maximal agreement using the well-known Procrustes rotation technique. The GPA algorithm, as originally outlined by Gower (1975), works in an iterative fashion, passing through the complete set of grids and producing a new consensus matrix at the end of each pass. The most recent consensus matrix is continually compared to the preceding consensus matrix (from the previous pass through the grids) until the two match at a level specified by the investigator; that is, until the quantified difference between the two consensus matrices fails to exceed some *tolerance* criterion (e.g., .001). Once the algorithm converges the final consensus matrix is saved, and the rescaled and rotated grids are then compared to the consensus matrix using standard ANOVA formulas and procedures. In this way individual grids that deviate from the consensus matrix (i.e.,

APPLIED MULTIVARIATE RESEARCH

the average grid) can easily be identified, and particular points of variability in the ratings can also be identified. The ANOVA results also yield an overall index of variability among the rescaled and rotated grids referred to as the *consensus proportion*. When converted to a percentage it will range from 0 to 100, with higher values indicating greater similarity in the organization of the matched figures across raters. In the extremes, a consensus proportion of 0 would indicate ratings that are completely dissimilar (e.g., as one might expect with random numbers) while a consensus proportion of 1.0 (100%) would indicate perfect agreement among raters in the multivariate configuration of the matched figures. Lastly, consensus proportions between pairs of individual grids can also be computed and then submitted to further analysis to identify clusters of raters who are similar to one another.

Table 1. Isotropic scaling factors for personal construct and Big Five grids

Grid	Personal Construct	Big Five
Tasha	0.98	0.90
Jeff	0.91	0.80
Kayla	0.98	0.89
Tom	1.02	0.88
Jenny	0.97	1.09
Claire	1.06	1.09
Eric	1.51	1.70
Rhonda	1.42	1.16
Debbie	1.05	0.89
Sally	0.88	0.95
Andrea	0.89	0.99
Zack	0.84	0.95
Paige	1.24	1.28
Matt	0.90	0.84
Samantha	0.89	0.92
Sue	1.06	1.19
Jessica	1.01	1.42
Erin	1.02	1.42
Lindsey	0.97	0.83

The results of the GPA for the personal construct grids yielded a consensus proportion of .45, which indicated only moderate average agreement among the 19 students when considering themselves and their peers. This consensus proportion was tested for statistical significance using a randomization test developed by

GENERALIZED PROCRUSTES ANALYSIS

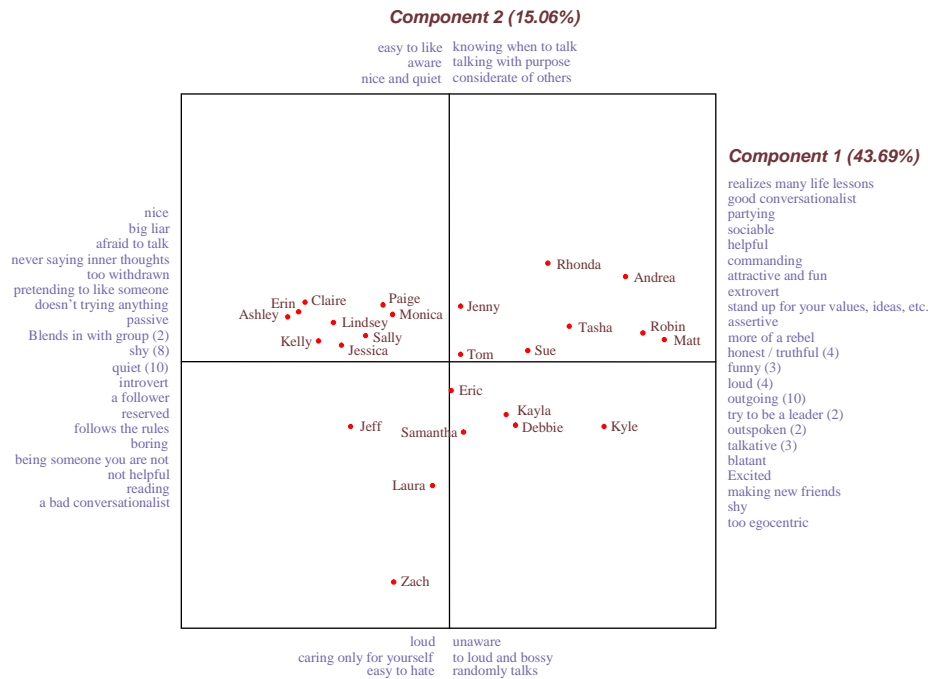


Figure 2. Bi-plot of consensus matrix from personal construct grids. Highly salient personal constructs are also indicated on the two axes.

Wakeling, Raats, and MadFie (1992). The test works by randomizing the matched figures in each of the 19 grids and then conducting the GPA. This process is repeated numerous times and the consensus proportion is computed and recorded for each trial. The observed consensus proportion can then be compared to the distribution of proportions generated from randomized grids. In the current data 500 trials were generated, and not a single value equaled or exceeded the observed consensus proportion. The observed proportion of .45 was therefore judged to be statistically significant at the .05 level (observed $p < .002$; that is, less than 1/500).

A common strategy is to submit the consensus matrix to a Principal Components Analysis (PCA), the purpose of which is to reduce the dimensionality of the consensus matrix so that the organization of the rated targets (in this case, rated students) can be visually examined in bi-plots. The relationships between the targets and the individual rating dimensions can also be examined in these bi-plots using extension analysis. Grice and his colleagues (Grice, 2007; Grice, Jackson, McDaniel, 2006; Grice, 2004) have shown how extension analysis can be conducted with personality data. For the current data, the first two principal components explained 58.75% of the variance in the consensus matrix values and were used to create the axes in Figure 2. An extension analysis was then conducted in which the personal constructs for the 19 students were mapped into the two-dimensional space formed by the principal components. A total of 190 constructs could potentially be mapped into the space, but the structure coefficients from the PCA were examined to select the most salient constructs shown in Figure 2. As can be seen, the first component differentiates people who are outgoing and talka-

APPLIED MULTIVARIATE RESEARCH

tive versus shy and introverted. Matt, Robin, and Andrea are clearly outgoing whereas Ashley, Erin, and Claire are shy. The second component appears to differentiate people who are nice and considerate (e.g., Rhonda and Andrea) versus those who are selfish and disagreeable (e.g., Zach and Laura). The multivariate configuration in Figure 2 thus represents the aggregate view computed from the 19 personal construct grids.

Analysis of Variance

As mentioned above, GPA is a combination of Procrustes rotation and analysis of variance (ANOVA). The ANOVA results for the personal construct grids are reported in Tables 2 and 3, and they can be examined to identify points of agreement and disagreement among the participants' views of themselves and their peers. The residuals reported in the tables are most useful for this process and can be examined with respect to their relative magnitudes. As can be seen in Table 2, the personal construct residual values are relatively homogeneous, ranging in value from 2.59 to 3.16 for the personal construct grids. Rhonda's grid produced the largest residual, indicating that her ratings deviated most from the consensus matrix; whereas Claire's grid showed the greatest degree of agreement with the consensus matrix. With smaller sets of grids and for different purposes these residuals can be used to identify 'outliers'; that is, people who deviate from the average view of the rated figures. In other contexts, for instance, the discrepant views of a person in a focus group could be identified and explored, or the novel views of a person on a management team could be extracted and investigated.

The ANOVA residuals for the rated peers can also be computed and examined for specific points of departure from the consensus matrix. As shown in Table 3, these residuals ranged in value from 1.56 to 3.37. It appears the 19 participants showed the greatest discrepancies with respect to their ratings of Zach (residual = 3.37) and Debbie (residual = 3.23). As described by Dijksterhuis and Punter (1991), the ANOVA residuals can be broken down even further to identify exactly which participants contributed most to the varied views of Zach and Debbie. Although these specific residuals are not reported here, they indicated that Eric's view of Debbie was most discrepant from the consensus. A bi-plot of Eric's first two principal components (explaining 72.20% of the rating variance) extracted from his grid are shown in Figure 3. It can be seen that Eric viewed Debbie as highly similar to Zach, both of whom were construed as "immature", "dependent", "uptight", and "serious." In the consensus matrix in Figure 2, Debbie was viewed as most like Kayla, and neither were located on the extreme ends of the dimensions in the bi-plot or close to Zach. Comparing these visual plots demonstrates clearly how the views of specific individuals can be related to the aggregate results.

Procrustes statistics

The consensus matrix and consensus proportion are computed from all of the grids in the analysis, and the latter statistic indicates the general degree of similarity among the grids with respect to the configurations of the rated people. The same principles used to compute the consensus proportion can also be applied to pairs of grids. Specifically, the *procrustes statistic* can be computed for

GENERALIZED PROCRUSTES ANALYSIS

Table 2. Analysis of variance results for the nineteen students' personal construct and Big Five grids

Grid	Personal Construct		Big Five	
	Residual	Total	Residual	Total
Tasha	2.94	6.44	2.23	5.93
Jeff	2.61	3.89	2.30	4.66
Kayla	2.95	4.91	2.19	5.74
Tom	3.12	6.17	2.38	5.21
Jenny	2.91	5.60	2.30	5.71
Claire	2.59	3.07	2.43	5.02
Eric	2.88	4.67	2.32	4.95
Rhonda	3.16	6.35	2.35	5.18
Debbie	2.71	4.53	2.33	4.89
Sally	2.80	5.05	2.35	5.58
Andrea	3.06	6.40	2.24	6.28
Zack	2.78	4.36	2.46	4.27
Paige	2.90	5.41	2.31	5.89
Matt	3.07	7.29	2.30	6.43
Samantha	2.70	3.78	2.40	5.38
Sue	3.03	6.71	2.33	5.73
Jessica	2.97	4.93	2.40	4.65
Erin	2.84	3.79	2.41	4.57
Lindsey	3.11	6.65	2.48	3.92
Total SS	55.11	100.00	44.51	100.00
Consensus Proportion		.45		.55

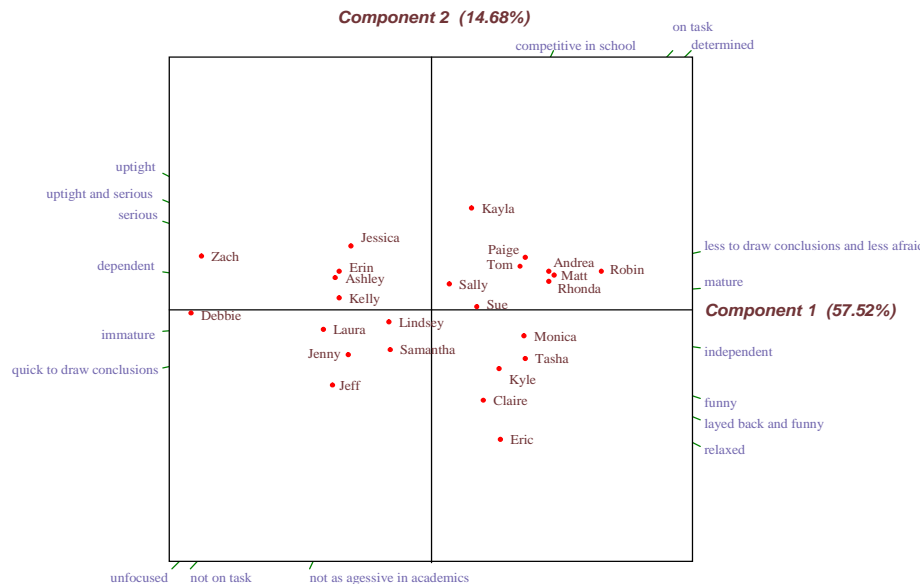


Figure 3. Bi-plot of a Eric's personal construct grid that was discrepant from the consensus matrix.

APPLIED MULTIVARIATE RESEARCH

Table 3. Analysis of variance results for the twenty-five rated people in the personal construct and Big Five grids

Rated Person	Personal Construct		Big Five	
	Residual	Total	Residual	Total
Tasha	2.09	3.64	1.64	3.47
Jeff	2.02	3.46	1.41	3.69
Kayla	2.38	4.56	2.28	4.56
Tom	1.90	2.68	1.64	3.15
Kelly	1.59	2.99	1.38	2.69
Ashley	1.85	4.00	1.46	2.89
Jenny	2.30	2.94	1.75	2.84
Claire	2.58	5.05	1.87	4.22
Eric	1.61	2.72	1.66	3.09
Rhonda	2.02	3.59	1.90	4.49
Debbie	3.23	5.37	2.44	5.15
Kyle	1.99	4.72	1.43	5.27
Sally	1.96	2.85	1.49	2.29
Andrea	2.07	5.46	1.84	5.45
Zach	3.37	6.59	1.99	6.98
Paige	2.49	3.56	1.93	4.04
Matt	2.65	6.10	1.88	6.57
Robin	2.43	5.86	1.73	4.67
Laura	2.89	4.45	1.79	4.04
Monica	1.87	2.31	1.66	2.23
Samantha	2.10	3.40	1.90	3.33
Sue	2.29	3.59	1.96	3.91
Jessica	1.56	2.61	1.71	3.10
Erin	1.56	3.57	1.76	4.31
Lindsey	2.32	3.94	2.02	3.58

each pair of rescaled and rotated grids to indicate the extent to which any two participants differ with respect to the configurations of the rated people. Converse to the consensus proportion, high values for the procrustes statistic indicate greater dissimilarity between the grids, and while there is no limit on its magnitude a value of zero indicates perfect agreement between two grids.

Given that the statistic can be computed for all pairs of grids it is common practice to report the procrustes statistics in matrix form. The matrix appears much like a correlation matrix, with the grids comprising the rows and columns and the procrustes statistics reported in the off-diagonal cells of the table. The main diagonal of the table consists of zeros, indicating that each grid is identical to itself. This matrix can then be submitted to a Singular-Value-Decomposition or

GENERALIZED PROCRUSTES ANALYSIS

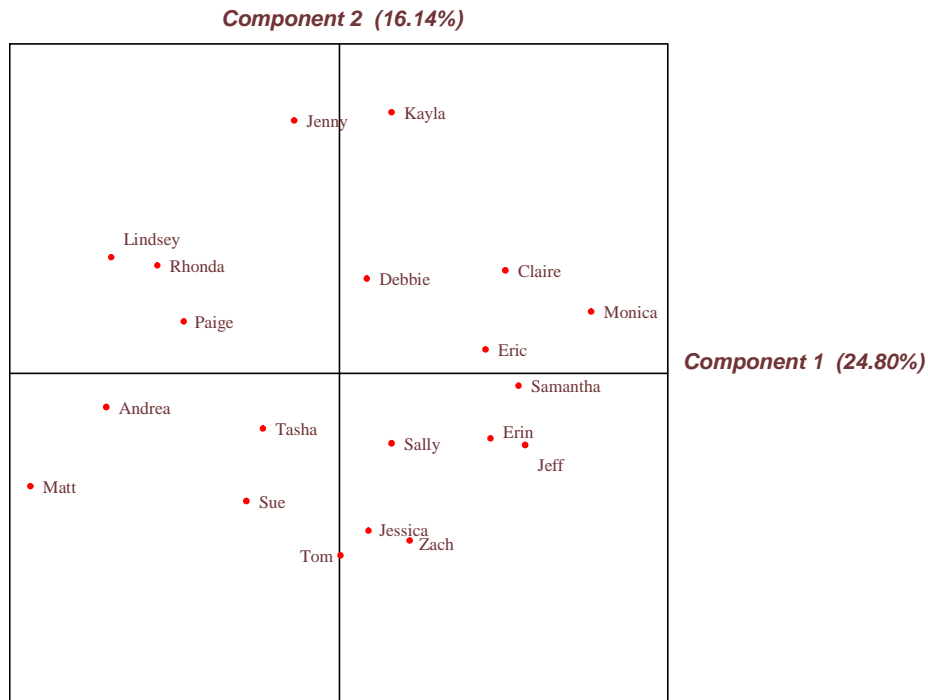


Figure 4. Bi-plot of procrustes statistics computed from personal construct grids.

a Principal Components Analysis and the results plotted in a bi-plot. The latter analysis was chosen for the procrustes statistics computed for the personal construct grids. The first two components, explaining 40.94% of the variance, are plotted in Figure 4. It can be seen that Jenny and Kayla were more similar to one another in their views than they were to the other participants. Similarly, Zach, Tom, and Jessica were relatively similar to one another in their ratings of themselves and their peers. Examination of such a plot can therefore aid in the identification of clusters of individuals who show relative agreement and disagreement in their ratings. For the current analyses, however, it should be noted that the procrustes statistics (not reported) were relatively homogeneous, ranging in value from .21 to .41, which is consistent with the homogeneous residuals reported in Table 2. Consequently, the clusters of people in Figure 4 may not represent substantive differences in their views. For example, the procrustes statistic for Jenny and Kayla (0.22), who appear close to one another in Figure 4, is not much lower than the procrustes statistic for Jenny and Tom (0.37), who appear quite discrepant in the figure. In order to investigate the practical relevance of these differences bi-plots created from Principal Components Analyses conducted on Jenny's, Kayla's, and Tom's individual grids could be constructed and examined.

APPLIED MULTIVARIATE RESEARCH

Big Five grids

The Big Five grids in which the participants rated themselves and their peers on 20 marker items for the Big Five personality traits were also analyzed with GPA (sans the ideal self). The same pre-scaling options used for the personal construct grids were chosen, and the isotropic scaling factors are reported in Table 1. Eric’s ratings were again stretched more than the other 18 students as he again chose not to use the end-points of the rating scale. As stated above, however, these individual differences were not of primary importance in this study.

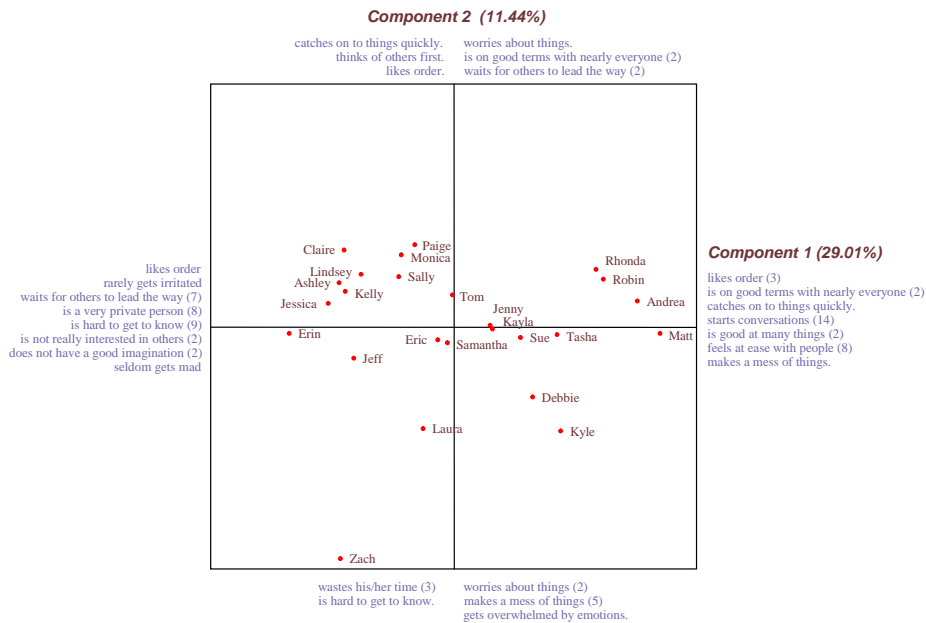


Figure 5. Bi-plot of consensus matrix from Big Five grids. Highly salient traits terms are also indicated on the two axes.

The consensus proportion for the Big Five grids was equal to .55 ($p < .002$, 500 randomized trials) and was slightly higher than the result for the personal construct grids. A Principal Components Analysis was conducted on the consensus matrix, and the first two components that explained 40.45% of the variance in the consensus matrix were plotted in Figure 5. Also shown in the figure are the 25 rated people and the Big Five marker items with salient structure coefficients. The first component clearly differentiates people who “start conversations” and “feel at ease with people” from those who are “hard to get to know” and “very private.” The second component differentiates people who think “...of others first” and who are “...on good terms with nearly everyone.” Matt, Andrea, Robin, and Rhonda are viewed as similar to one another and opposite to Erin, Jessica, and others in Figure 5. The former individuals are, on average, viewed as being at ease with others and able to start conversations, while the latter are viewed as disinterested in others and hard to get to know. Zach is again relatively disparate from the other rated people, and is viewed, on average,

GENERALIZED PROCRUSTES ANALYSIS

as someone who wastes time and makes a mess of things. The constructs located on the opposite end of the second component also apply to Zach in a negative way, as he is not on good terms with others and does not think of others first. With respect to the ANOVA results, Table 2 again revealed a great deal of homogeneity among the residuals for the participants' Big Five ratings. In other words, no single student's grid deviated distinctly from the consensus matrix compared to the other students in the study. With respect to the views of particular students, Table 3 indicated that Debbie was again rated in a manner that was relatively inconsistent with the consensus view. Examination of the specific residuals (not shown) revealed that Eric viewed Debbie more negatively than he did Zach, whereas in the consensus matrix (see Figure 5), Zach alone was viewed as highly disparate from most of the other students.

When comparing the personal construct (Figure 2) and Big Five (Figure 5) bi-plots, it becomes obvious that the 25 people are arranged in a similar pattern in both graphs. While the descriptive terms associated with the poles of the components may not match perfectly across the two figures, the organization of the rated people is highly similar. In fact, a GPA conducted on the two consensus matrices produced an extremely high consensus proportion of .97 ($p < .002$, 500 randomized trials). In terms of the aggregated data, then, the participants grouped themselves and their peers into highly similar clusters using either their own personal constructs or the Big Five Marker items.

Discussion

Gower's *Generalized Procrustes Analysis* (GPA; 1975) is an inventive combination of Procrustes rotation and analysis of variance. In the results above it was shown that GPA is a flexible and powerful technique for modeling three-dimensional matrices of data. Specifically, a number of questions were addressed regarding the intra- and inter-personal perceptions of students attending a Summer Science Academy. First, did the high school students reach a consensus with regard to their views of themselves and their peers? For the personal construct ratings the *consensus proportion* from the GPA was equal to .45, and for the Big Five ratings the value was equal to .55. The commonality found among the multivariate configurations of the students' ratings of themselves and their peers, although statistically significant, was therefore modest in both the personal construct and Big Five grid grids. Second, did the average, multivariate configuration of the students derived from the personal construct grids match the configuration derived from the Big Five trait grids? To address this question, we computed the consensus matrix from each rating task. As described above, a consensus matrix is an average grid that is computed from all of the participants' rescaled and rotated original grids. These two consensus matrices were themselves then submitted to a GPA, and the resulting consensus proportion was extremely high, .97, indicating very high similarity between the two configurations. Examination of the principal component bi-plots of the consensus matrices also indicated a high degree of similarity in the multivariate configurations of the rated students in the personal construct and Big Five consensus matrices. Third, did particular students emerge from the analysis as highly discrepant in their ratings of themselves and the other students when

APPLIED MULTIVARIATE RESEARCH

compared to the consensus? This question was answered on a general level with respect to each student's overall grid responses, and on a more specific level with respect to the ratings of particular students in each participant's grid. Several students were identified, for instance, to deviate generally from the consensus, although the extent of the differences were not noteworthy when compared to all of the participants in the study. On a more specific level, points of disagreement with regard to particular students (Zach and Debbie) in the group were clearly identified in both the Personal Construct and Big Five grids. These specific points of disagreement were furthermore attributed primarily to one student rater, Eric, as the differences between his views of Zach and Debbie and those expressed in the consensus matrix could be clearly seen in the comparison of the aggregate and individual bi-plots.

The analyses addressing this third question in fact reveal what is perhaps the most intriguing feature of GPA; namely, the capability of deftly moving from the aggregate to the individual in the analysis. Despite analyzing 4750 total ratings in the personal construct grids and 9500 total ratings in the Big Five grids, the individuals in the study were never lost from view nor were they swept away completely into the aggregate. Through the specific analysis of variance results and the visual examination of the bi-plots particular points of similarity and dissimilarity among the persons were discovered and examined. Recent years have seen renewed interest in sophisticated analyses that place considerably more focus on the individual relative to the aggregate (e.g., Molenaar, 2004, 2005), and Grice and his colleagues (Grice, 2004, 2007; Grice et al., 2004; Grice et al., 2006) have explored techniques that connect person-centered and aggregate-based results. In a similar vein, the recent emphasis on reporting and interpreting effect sizes will necessitate connecting aggregate patterns of results to the lives of individuals who are impacted by published research. While the aggregate may reveal important features of the data that can be meaningfully tied to variables and to hypotheses, the testing ground of most psychological theories will be the person, whether considered in isolation or considered as embedded in a social situation. With GPA a researcher can place his or her focus on the person and the aggregate in a single set of analyses. Consequently, GPA could prove particularly fruitful in the study of groups, such as families, classrooms, or organizations. It could also prove fruitful for studying how an individual changes over time in his or her views of self and others, or how a person's perceptions change across contexts. In the analyses above, rating matrices were obtained from 20 different students, but matrices obtained from the same person on different occasions or from the same person rating different thoughts, emotions, or actions in different situations could also be analyzed. The flexibility of GPA can in fact be seen in the variety of domains for which it has been used; for instance, in the study of individuals' perceptions of genetic engineering (Frewer, Howard, & Shepherd, 1997), meat (Russell & Cox, 2003), commercial ports (Williams & Langron, 1984), and their views of the 'personalities' of pigs (Wemelsfelder, Hunter, Mendl, & Lawrence, 2000).

Technical Issues

Beginning with Gower's (1975) original article, a number of technical issues have consistently been raised with respect to GPA. Perhaps foremost is the

GENERALIZED PROCRUSTES ANALYSIS

long-standing criticism of the sensitivity of Procrustes rotation. It has been shown, for instance, that even when two matrices of data are generated randomly, the Procrustes transformation can rotate the matrices to a level of agreement that can be surprisingly high (e.g., see Horn, 1967). In the context of GPA, this property of Procrustes rotation would suggest that the consensus proportion may be inflated for some data sets. For instance, a consensus proportion of .70 may seem high given the possible range of values from 0 to 1; however, this result may simply be an artifact of the rotation's capitalization on chance patterns in the data. The randomization test developed by Wakeling, et al. (1992) is therefore an essential part of any GPA, since it allows the researcher to compare the observed consensus proportion to a distribution of proportions created from randomized versions of the original data matrices. Traditional cut-points for statistical significance (viz., .05 and .01) can then be used to determine if the observed consensus proportion is statistically significant.

A second issue regarding GPA is the application of the centering, dimensional, and isotropic pre-scaling transformations. Specifically, when are these transformations to be applied to the original data? Unfortunately, no simple answer can be offered to this question, but we generally recommend running a series of analyses with any given data set. When conducting a GPA, the goal of most researchers is to match the multivariate configurations of the rated targets (e.g., people, things, places, etc.) across raters. Removing potential, systematic differences in scale usage or dimensionality of the rating matrices appears to be consistent with this goal. The primary analysis should therefore be based on ratings that have been transformed with each of the three scaling options. Of course, if the rating matrices are all of the same dimensionality, as with the repertory grids in this study, then dimensionality scaling is superfluous. After the primary results are interpreted, the researcher should conduct the GPA again without one of the scaling options. The results can then be compared to the completely scaled results and the impact of the particular scaling option can be thoroughly assessed.

A third issue involves analyses based on grids of different dimensionality. Consider, for instance, the following two sets of ratings:

		Cereal A	Cereal B	Cereal C	Ideal Cereal
Joe	<i>Sweet</i>	-3	-2	2	3
	<i>Crunchy</i>	1	2	2	2
	<i>Pricey</i>	2	-1	-3	2
Mary	<i>Expensive</i>	-2	-3	-2	-1
	<i>Healthy</i>	-2	-1	2	2

Joe has now rated the four cereals on three dimensions whereas Mary has rated the cereals on only two dimensions. As described above GPA can still be used to analyze such data by treating the dimensions as arbitrary and 'padding' Mary's ratings with a third row of zeros. As described above, the overarching goal of

APPLIED MULTIVARIATE RESEARCH

GPA is to align the multivariate configurations of the cereals. Joe's ratings comprise a 3-dimensional space whereas Mary's comprise only a 2-dimensional space. Mary's ratings, however, can be considered to comprise a 3-dimensional space in which the cereals do not extend into the third dimension. In other words, the ratings for the cereals are all zero on the third dimension. In computer programs such as Idiogrid (Grice, 2002), small matrices will automatically be padded with zeros prior to the analysis, and the consensus matrix will be labeled in such a way as to indicate the arbitrary nature of the dimensions. The consensus matrix for Joe's and Mary's ratings above, for instance, appears as follows:

	Cereal A	Cereal B	Cereal C	Ideal Cereal
<i>Dim_1</i>	-0.25	-1.14	-0.10	1.49
<i>Dim_2</i>	-1.87	-0.76	1.60	1.03
<i>Dim_3</i>	0.00	0.00	0.00	0.00

The labels 'Dim_1', 'Dim_2', and 'Dim_3' refer to the arbitrary dimensions, and the zeros in the third row indicate that the two matrices could not be matched in the third dimension. If other raters who used three dimensions were included in the analysis, the third row in the consensus matrix above would report values other than zero. It is thus by padding smaller matrices with zeros that matrices of different dimensionality can be analyzed in GPA. Of course, the matched figures must be the same and in the same order across matrices. In other words, both Joe and Mary needed to rate the same cereals that are recorded in identical columns in their two matrices.

Finally, the graphical methods associated with GPA play a critical role in interpreting and reporting the results. In many cases these methods involve bi-plots created from the results of Principal Components Analyses. The figures created for the results above show how useful they can be for summarizing the aggregate results as well as relating individual data to the aggregate. Examples of similar plots can be found in Monteleone, Raats, and Mela (1997), Oreskovich, Klein and Sutherland (1991), and Russell and Cox (2004). Furthermore, Gower and Dijksterhuis (1994) have discussed creating bi-plots from matrices with both quantitative and qualitative variables. Similar developments on both the applied and analytical fronts should contribute to the continued growth and popularity of Generalized Procrustes Analysis.

REFERENCES

- Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, 31, 33-42.
- Dijksterhuis, G. B., & Gower, J. C. (1991/2). The interpretation of generalized procrustes analysis and allied methods. *Food Quality and Preference*, 3, 67-87.

GENERALIZED PROCRUSTES ANALYSIS

- Dijksterhuis, G., & Punter, P. (1991). Interpreting generalized procrustes analysis 'analysis of variance' tables. *Food Quality and Preference*, 2, 255-265.
- Frewer, L. J., Howard, C., & Shepherd, R. (1997). Public concerns in the United Kingdom about general and specific applications of genetic engineering: Risk, benefit, and ethics. *Science, Technology, & Human Values*, 22, 98-124.
- Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58, 453-467.
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mer-vielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe: Vol. 7* (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40, 33-51.
- Gower, J. C., & Dijksterhuis, G. B. (1994). Multivariate analysis of coffee images: A study in the simultaneous display of multivariate quantitative and qualitative variables for several assessors. *Quality & Quantity*, 165-184.
- Grice, J. W. (2007). Person-centered structural analyses. In R. Robins, C. Fraley, and R. Krueger (Eds.) *Handbook of research methods in personality psychology* (pp. 557-572). New York: Guilford Press.
- Grice, J. W. (2004). Bridging the idiographic-nomothetic divide in ratings of self and others on the Big Five. *Journal of Personality*, 72, 203-242.
- Grice, J. W. (2002). Idiogrid: Software for the management and analysis of repertory grids. *Behavior Research Methods, Instruments, & Computers*, 34, 338-341.
- Grice, J., Burkley, E., Burkley, M., Wright, S., & Slaby, J. (2004). A sentence completion task for eliciting personal constructs in specific domains. *Personal Construct Theory & Practice*, 1, 60-75.
- Grice, J. W., Jackson, B., & McDaniel, B. (2006). Bridging the idiographic-nomothetic divide: A follow-up study. *Journal of Personality*, 74, 1191-1218.
- Horn, J. L. (1967). On subjectivity in factor analysis. *Educational and Psychological Measurement*, 27, 811-820.
- Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, 2, 201-218.
- Molenaar, P. C. M. (2005). How generalization works through the single case: A simple idiographic process analysis of an individual psychotherapy. *International Journal of Idiographic Science*, 1, 1-20.
- Monteleone, E., Raats, M., & Mela, D. (1997). Perceptions of starchy food dishes: Application of the repertory grid technique. *Appetite*, 28, 255-265.
- Oreskovich, D. C., Klein, B. P., & Sutherland, J. W. (1991). Procrustes analysis and its applications to free-choice and other sensory profiling. In Lawless and Klein (Eds.) *Sensory Science Theory and Applications in Foods* (pp. 353-393). New York: Marcel Dekker.
- Russell, C. G., & Cox, D. N. (2004). Understanding middle-aged consumers' perceptions of meat using repertory grid methodology. *Food Quality and Preference*, 15, 317-329.

APPLIED MULTIVARIATE RESEARCH

- Schönemann, P. H. (1966). A generalized solution of the orthogonal Procrustes problem. *Psychometrika*, *31*, 1-10.
- Slater, P. (1977). *The measurement of intrapersonal space by grid technique: Vol. 2. Dimensions of intrapersonal space*. London: Wiley.
- Ten Berge, J. M. F. (1977). Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, *42*, 267-276.
- Wakeling, I. N., Raats, M. M., Halliday, J. H. (1992). A new significance test for consensus in generalized procrustes analysis. *Journal of Sensory Studies*, *7*, 91-96.
- Wemelsfelder, F., Hunter, E. A., Mendl, M. T., & Lawrence, A. B. (2000). The spontaneous qualitative assessment of behavioural expressions in pigs: First explorations of a novel methodology for integrative animal welfare measurement. *Applied Animal Behaviour Science*, *67*, 193-215.
- Williams, A. A., & Langron, S. P. (1984). The use of free-choice profiling for the evaluation of commercial ports. *Journal of the Science of Food and Agriculture*, *35*, 558-568.

Notes

Correspondence concerning this article should be addressed to James W. Grice, Department of Psychology, 215 North Murray, Oklahoma State University, Stillwater, OK 74078. Electronic mail may be sent to james.grice@okstate.edu.