

# SPSS and SAS programs for generalizability theory analyses

CHRISTOPHER MUSHQUASH and BRIAN P. O'CONNOR  
*Lakehead University, Thunder Bay, Ontario, Canada*

The identification and reduction of measurement errors is a major challenge in psychological testing. Most investigators rely solely on classical test theory for assessing reliability, whereas most experts have long recommended using generalizability theory instead. One reason for the common neglect of generalizability theory is the absence of analytic facilities for this purpose in popular statistical software packages. This article provides a brief introduction to generalizability theory, describes easy to use SPSS, SAS, and MATLAB programs for conducting the recommended analyses, and provides an illustrative example, using data ( $N = 329$ ) for the Rosenberg Self-Esteem Scale. Program output includes variance components, relative and absolute errors and generalizability coefficients, coefficients for D studies, and graphs of D study results.

Consistently accurate measurement is a necessary goal in almost all empirical psychological investigations. Classical test theory, item response theory, and generalizability theory (G theory) are the three broad, overlapping approaches to assessing measurement fidelity. Although the latter two approaches have long been widely recommended by experts, classical test theory continues to be, by far, the most commonly used method in the empirical literature. A primary reason why item response theory and G theory are rarely used is the lack of facilities for conducting such analyses in popular statistical software packages. The present article describes simple SPSS, SAS, and MATLAB programs for conducting G theory analyses, with the hope of facilitating and encouraging the use of this analytic technique.

Classical test theory provides an assortment of useful individual methods for assessing reliability (e.g., test-retest, alternate forms, internal consistency, and interobserver agreement), but it considers only one source of measurement error at a time (i.e., for occasions, forms, items, or observers). There is also no way of combining the numerous possible reliability indices to obtain an overall estimate of reliability, no way of determining the relative importance of the various sources of error, and no way of assessing possible interactions between the sources of error. Researchers are also not provided with explicit information on how many more items, occasions, forms, or observers are required to obtain dependable scores from their respondents (Webb, Rowley, & Shavelson, 1988).

G theory is a more encompassing, informative, and useful alternative. The errors that exist in observed scores can stem from different sources. G theory provides an all-

at-once way of revealing and comparing the sources of error in a common metric. G theory provides estimates of the variance contributed by each source, and it provides estimates of the variance associated with interactions between the various sources. For example, if an instrument is administered on two occasions, G theory provides estimates of the variance contributed by persons, items, and occasions of measurement and of each of the four possible interactions between these facets (persons  $\times$  items, persons  $\times$  occasions, items  $\times$  occasions, and persons  $\times$  items  $\times$  occasions). It also provides helpful forecasts of the improvements in measurement reliability that can be obtained by altering the numbers of persons, items, occasions, and so forth.

The modern, most influential presentation of G theory was that by Cronbach, Gleser, Nanda, and Rajaratnam (1972). Excellent recent introductions and reviews of G theory have been provided by Brennan (2000a, 2000b, 2001), Hoyt and Melby (1999), Kieffer (1999), Marcoulides (1999, 2000), Strube (2000), Shavelson, Webb, and Rowley (1989), and Webb et al. (1988). Fortran programs for G theory analyses have been provided by Crick and Brennan (1983) and are available from Robert Brennan at [www.education.uiowa.edu/casma/GenovaPrograms.htm](http://www.education.uiowa.edu/casma/GenovaPrograms.htm). However, a simple program interface for popular software packages has not yet been available. We therefore wrote flexible, easy-to-use SPSS, SAS, and MATLAB G theory programs that conduct the analyses that are likely to be of interest to most users. Before providing a description of the programs and illustrative results, the primary concepts from G theory must be introduced.

## G Theory Concepts

An individual observation or measurement is merely a sample estimate from an infinite and varied universe of possible measurements. Measurements could be taken on a multitude of occasions, using various items and forms of a measuring instrument. Individual measurements are

---

This work was supported by a grant from the Social Sciences and Humanities Research Council of Canada. Address correspondence to B. P. O'Connor, Department of Psychology, Lakehead University, 955 Oliver Road, Thunder Bay, ON, P7B 5E1 Canada (e-mail: [brian.oconnor@lakeheadu.ca](mailto:brian.oconnor@lakeheadu.ca)).

estimates of an individual's true scores and are parts of a *universe of admissible observations*. Forms, items, occasions, and raters are called *facets*, which can include any characteristic of a measurement procedure that is a potential source of error. The levels within each facet (e.g., the different items or the different occasions of measurement) are *conditions* that can be infinitely large (Webb et al., 1988). The *object* of measurement is usually persons, which is typically not considered a source of error, because people vary and their true score differences are real, systematic, and of great interest to investigators (Eason, 1991; Kieffer, 1999). The objects of measurement do not create error variance and are, therefore, not considered a facet. Anything that generates systematic variance can be the object of measurement. Other possibilities include schools, businesses, work groups, or occasions.

For each person (which, for simplicity, will be the objects of measurement in this article), the means of the scores from the various conditions and facets provide the best estimates of the person's true, or *universe*, score. This grand mean is always a flawed estimate of the universe scores, because of the errors contributed by the measurement facets. G theory decomposes, estimates, and reveals these measurement errors, which are termed *variance components*. A *G coefficient* is produced for each data set, representing the universe score variance divided by the observed score variance. G coefficients range between zero and one. When the G coefficient for a data set is high, investigators can generalize the obtained scores across the study facets—hence, the term *generalizability theory*.

The data for G theory studies can be crossed or nested. In a fully *crossed* data set, the objects of measurement have scores on all of the same facets and conditions. The data are *nested* when this is not the case—that is, when each condition of a facet does not appear in combination with each condition of other facets. For example, the data are nested when there are different items or occasions of measurement for different respondents. It is possible to have a G theory data set in which there are crossed data for some facets and conditions and nested data for other facets and conditions. Fully crossed designs are preferable because they provide estimates of the most possible sources of error of measurement, which are otherwise confounded in nested designs.

Facets can be *fixed* or *random*. A facet is random if its conditions can be exchanged with any other of the conditions from the same facet (Kieffer, 1999). For example, items from a course examination item pool can typically be exchanged with any other items from the same pool. Randomly selecting the conditions permits researchers to generalize across the universe of conditions for the facet. Facets are considered fixed when their levels are not exchangeable and are of specific interest to the investigator, as in many experiments. Generalization, in this case, is restricted to the conditions of the research.

A distinction is made between *relative* and *absolute* G coefficients, both of which are computed by the present programs. A relative G coefficient reflects the degree to which the objects of measurement maintain their rank

order across facets, regardless of possible changes in raw score elevations. It is analogous to the reliability coefficients of classical test theory. Absolute G coefficients are more stringent and reflect both the degree of consistency in the rank ordering of measurement objects and the consistency in the elevations of the raw scores. Absolute G coefficients are most useful when the actual values of the obtained scores are important or meaningful to the investigator. These typically involve performance measurements where there is a cutoff value that is deemed particularly meaningful (e.g., for driving tests or for categorical psychiatric diagnoses). For technical reasons, the terms *phi coefficients* or *index of dependability coefficients* are used instead of *absolute G coefficients* (Shavelson & Webb, 1991, p. 93). Only those variance components that reflect the rank ordering of persons are involved in the G estimates for relative decisions, whereas all of the variance components (except that for the universe scores) are involved in the G estimates for absolute decisions.

In single-facet fully crossed designs, G coefficients are identical to the intraclass correlations that are often used to estimate interobserver reliability (Hoyt & Melby, 1999, p. 339). Relative G coefficients are identical to the intraclass correlations produced by Shrout and Fleiss's (1979) ICC-3 formula, and absolute G (phi) coefficients are identical to the intraclass correlations produced by Shrout and Fleiss's ICC-2 formula. *G study* refers to the analyses that are conducted on the actual data—that is, on the objects and facets of measurement in a given data set. A *D study* (or decision study) is performed when the results of G studies are used to forecast the G coefficients that would be obtained for alternative, hypothetical levels of the facets. For example, a D study could be used to forecast the changes in G coefficients that would result from doubling or tripling the numbers of items or occasions. (There is a direct analogy with the Spearman–Brown prophecy formula.) The present programs have a D study utility.

### SPSS, SAS, and MATLAB Programs for G Theory

Two kinds of G theory programs were created in both SPSS and SAS formats. The G1 version is easier to use but is more restricted in the kinds of permissible research designs. The G2 version requires only slightly more involvement on the part of users, and it is more flexible with regard to possible research designs, analytic options, and program output. A MATLAB version of the G1 program is also available.

**G1.sps and G1.sas.** The computations for the G1 programs are performed within the Matrix–End Matrix environment of SPSS and within the PROC-IML environment of SAS. Users simply read in their data, enter their specifications, and then click Run or Submit to obtain their results.

The data can be entered into the programs directly, in matrix form, or SPSS and SAS data files can be located and read by the programs. Each row of the data file contains the data for one object of measurement (e.g., one person). The columns of the data file contain the facet condition scores. Missing values are not permitted.

Users specify the nature of their data set by entering values, in accordance with a brief provided guide, into statements near the top of the programs. One statement asks the user to enter the number of levels or conditions for Facet 1; a second statement asks the user to enter the number of levels or conditions for Facet 2; and a third statement asks the user to enter the type of design. On the third statement, a "1" indicates a single-facet design; a "2," a single-facet nested design; a "3," a two-facet fully crossed design (e.g.,  $p \times i \times h$ ); a "4," a two-facet nested design [as in  $p \times (i:h)$ ]; a "5," a two-facet nested design [as in  $(i:p) \times h$ ]; a "6," a two-facet nested design [as in  $i:(p \times h)$ ]; a "7," a two-facet nested design [as in  $(i \times h):p$ ]; and an "8," a two-facet nested design [as in  $i:h:p$ ]. Users then enter their preferred D Study values for the two facets, as well as the kind of coefficients they would like to view in graphs of the D study results.

Once the nature of the data set has been specified, the entire file is run to obtain the results. Users need not bother viewing all of the subsequent program computation statements, which are nevertheless accompanied by comments for interested users. The notation and variable names used in the programs were the same as those used by Brennan (2001). The G1 programs are appropriate for random facet, balanced designs. Program output includes sums of squares, mean squares, degrees of freedom, variance components, proportions for the variance components, relative and absolute error terms, G and phi coefficients, the results for user-specified D studies, and plots of the D study results. Variance components are computed using the ANOVA matrix algorithm described by Brennan (2001, pp. 79 and 439). This algorithm is also used in Crick and Brennan's (1983) GENOVA program.

**G2.sps and G2.sas.** The G2 programs involve use of a combination of syntax written by the authors and of analytic facilities for variance components that are available in more advanced versions of the SPSS and SAS packages. Researchers wishing to conduct G theory analyses encounter three awkward features of the built-in SPSS and SAS variance components routines: (1) The routines run very slowly, even for modest-sized data sets; (2) the data that are read into the built-in routines must be arranged or structured very differently from the arrangement most researchers will have in their data; and (3) the built-in routines provide variance component estimates but do not provide important, additional G theory statistics. Nothing can be done for the slow speed of the built-in routines, but the other two shortcomings can be remedied.

The G2 programs consist of syntax illustrating how data sets can be readily converted for entry into the built-in routines, syntax for running the built-in routines and (for SPSS only) for automatically saving the variance components output to a data file, and syntax that then reads the variance components data and provides the additional G theory coefficients for random designs, including graphs of D study results. A number of different procedures for estimating variance components can be selected when the built-in SPSS and SAS routines are used, including ANOVA, MINQUE, maximum likelihood, and restricted

maximum likelihood (for reviews, see Brennan, 2001, chap. 7; Searle, Casella, & McCulloch, 1992). These options are particularly useful for unbalanced designs.

**Program availability.** All of the programs can be downloaded from [flash.lakeheadu.ca/~boconno2/gtheory/gtheory.html](http://flash.lakeheadu.ca/~boconno2/gtheory/gtheory.html). They can also be obtained by e-mailing the author at [brian.oconnor@lakeheadu.ca](mailto:brian.oconnor@lakeheadu.ca) or by sending a stamped, self-addressed disk mailer to Brian O'Connor, Department of Psychology, Lakehead University, 955 Oliver Road, Thunder Bay, ON, P7B 5E1 Canada.

### Illustrative Example

The G1.sps program was used to analyze data for the popular Rosenberg Self-Esteem Scale (Rosenberg, 1979; see Blascovich & Tomaka, 1991, for a review). G theory analyses for this measure have not been previously reported in the literature. The scale consists of 10 items, and responses were provided on 8-point Likert scales. The participants were 329 undergraduates, all of whom completed the measures on two occasions, 5 months apart.

The user specifications for the G1.sps program for these data are presented in Appendix A. The program output is presented in Appendix B and in Figure 1. The largest estimated variance component ( $P = 1.418$ ; see Appendix B) was for persons. The persons variance component is the estimated variation in the respondents' scores when the score for each respondent represents his or her mean score across all items and both occasions of measurement. In other words, averaging over items and occasions, respondents differed substantially in their self-esteem scores. This variability for the objects of measurement is not error. The variability is desirable because it represents systematic individual differences in self-esteem. The remaining variance components represent error variance, and their magnitudes reveal the importance of the various sources of error.

The moderate-sized variance component for items ( $F1 = .564$ ) indicates variation in the degrees to which the items assess self-esteem. The zero variance component for occasions ( $F2 = .000$ ) indicates that self-esteem scores, when averaged across persons and items, remained consistent across the two occasions of measurement. The substantial interaction between persons and items ( $P * F1 = .816$ ) indicates that the rank ordering of persons differed across the 10 items. The interaction between persons and occasions ( $P * F2 = .416$ ) indicates that the rank ordering of persons differed slightly across occasions. Somewhat different correlates of self-esteem may thus emerge, depending on which occasion of measurement is used in one's analyses. The large three-way interaction ( $P * F1 * F2 = 1.81$ ) is a confounded and ambiguous term. It could reflect, in whole or in part, a three-way interaction between persons, items, and occasions. But this variance component is also the residual and may be caused by facets that were not included in the research design. The two possibilities cannot be disentangled.

The G coefficients for relative (.789) and absolute (.765) decisions are the reliabilities of the scores across the 10 items and two occasions. Although these values are

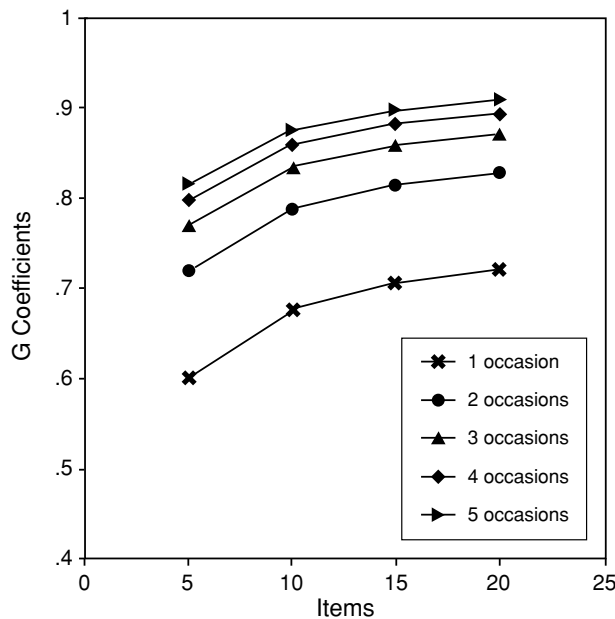


Figure 1. SPSS plot of D study G coefficients for the Rosenberg Self-Esteem Scale.

close to the conventional .80 criterion for reliability, they nevertheless provide reason for concern, because these reliabilities are based on two occasions of measurement, whereas most uses of the self-esteem scale involve just one occasion of measurement. The somewhat modest G coefficients and the interaction between persons and occasions suggest that just one occasion of measurement may not be sufficient. This possibility can be evaluated by perusing the D study results.

The relative error variance (.380) coefficient is similar to error variance in classical test theory. The square root of this coefficient is equivalent to the standard error of measurement. The absolute error variance (.437) represents differences between the respondents' observed scores and their universe scores. Absolute error variance is based on all of the error variances in the research design, whereas relative error variance is based on all variance components that involve both persons and one other facet. Both error variances are, of course, inversely related to the G coefficients. The D study results in Appendix B reveal the error variances and G and phi coefficients for various numbers of items and occasions. A plot of the D study G coefficients appears in Figure 1. Halving the number of items (to five) and the number of occasions of measurement (to one) results in notable decreases in reliability. For example, using one half of the scale items on just one occasion of measurement would involve a G coefficient of only .601. Most disconcerting is the fact that use of the full scale with just one occasion of measurement, which is the most common use of the scale in the literature, involves a G coefficient of only .676. At first glance, additional occasions of measurement may not seem necessary, because of the low variance component for occasions. But the interaction terms involving occasions of measurement were nonnegli-

gible. Increasing the number of occasions of measurement consequently helps suppress measurement error, although the gains are minimal beyond three occasions.

The G theory results provide a more comprehensive and sobering account of the reliability of the Rosenberg Self-Esteem Scale when compared with classical test theory psychometrics. The Cronbach alpha values for the scale in the present sample were .88 and .86 for the two occasions of measurement, and the test-retest correlation was .72. These impressive values are similar to those commonly reported for this scale (see Blascovich & Tomaka, 1991). But the interactions between persons, items, and occasions are neglected in classical test theory. The G theory analyses exposed these important, previously hidden sources of error in this popular scale. The D study results indicate how they can be reduced.

## REFERENCES

- BLASCOVICH, J., & TOMAKA, J. (1991). Measures of self-esteem. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 115-160). New York: Academic Press.
- BRENNAN, R. L. (2000a). (Mis)conceptions about generalizability theory. *Educational Measurement: Issues & Practice*, *19*, 5-10.
- BRENNAN, R. L. (2000b). Performance assessments from the perspective of generalizability theory. *Applied Psychological Measurement*, *24*, 339-353.
- BRENNAN, R. L. (2001). *Generalizability theory*. New York: Springer.
- CRICK, J. E., & BRENNAN, R. L. (1983). *Manual for GENOVA: A generalized analysis of variance system* (American College Testing Technical Bulletin 43). Iowa City: ACT, Inc.
- CRONBACH, L. J., GLESER, G. C., NANDA, H., & RAJARATNAM, N. (1972). *The dependability of behavioral measurements: Theory of generalizability of scores and profiles*. New York: Wiley.
- EASON, S. (1991). Why generalizability theory yields better results than classical test theory: A primer with concrete examples. In B. Thompson (Ed.), *Advances in educational research: Substantive findings, methodological developments* (Vol. 1, pp. 83-98). Greenwich, CT: JAI.
- HOYT, W. T., & MELBY, J. N. (1999). Dependability of measurement in counseling psychology: An introduction to generalizability theory. *Counseling Psychologist*, *27*, 325-352.
- KIEFFER, K. M. (1999). Why generalizability theory is essential and classical test theory is often inadequate. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 149-170). Stamford, CT: JAI.
- MARCOULIDES, G. A. (1999). Generalizability theory: Picking up where the Rasch IRT model leaves off? In S. E. Embretson & S. L. Hershberger (Eds.), *The new rules of measurement: What every psychologist and educator should know* (pp. 129-152). Mahwah, NJ: Erlbaum.
- MARCOULIDES, G. A. (2000). Generalizability theory. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 527-551). San Diego: Academic Press.
- ROSENBERG, M. (1979). *Conceiving the self*. New York: Basic Books.
- SEARLE, S. R., CASELLA, G., & McCULLOCH, C. E. (1992). *Variance components*. New York: Wiley.
- SHAVELSON, R. J., & WEBB, N. M. (1991). *MSSS generalizability theory: A primer*. Newbury Park, CA: Sage.
- SHAVELSON, R. J., WEBB, N. M., & ROWLEY, G. L. (1989). Generalizability theory. *American Psychologist*, *44*, 922-932.
- SHROUT, P. E., & FLEISS, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420-428.
- STRUBE, M. J. (2000). Reliability and generalizability theory. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding more multivariate statistics* (pp. 23-66). Washington, DC: American Psychological Association.
- WEBB, N. M., ROWLEY, G. L., & SHAVELSON, R. J. (1988). Using generalizability theory in counseling and development. *Measurement & Evaluation in Counseling & Development*, *21*, 81-90.

**APPENDIX A**  
**User Specifications Section of the G1 SPSS Program**

---

- \* G1.sps for Generalizability Theory analyses
  - \* There are two main ways of entering data into this program.
  - \* Method 1: You can specify the data to be analyzed on a GET statement, as in the example below. On the GET statement, the data matrix must be named SCORES (as in the example); "FILE = \*" will use the currently active SPSS data set; "FILE = C:\filename" will use the specified SPSS data file on your computer. The \* beside the GET command below converts the command into a comment. To use the GET command, remove the \* that appears before GET; specify the data for analysis on FILE =; and specify the variables for analysis on VARIABLES =.
  - \* GET scores / file = \* /variables = all / missing = omit.
  - \* Method 2: A data matrix can be entered directly, within the MATRIX--END MATRIX program, using a COMPUTE statement, as in the trial-run data below. The data matrix must be named "scores"; the data must appear inside curly brackets; place a comma between the data values for each case; and place a semicolon at the end of the data values for each case.
  - \* A third method for entering data is to use a READ statement, which is a MATRIX--END MATRIX command. READ can be used to read a raw data file on your computer. See the online documentation for MATRIX--END MATRIX.
  - \* Enter the number of levels/conditions of Facet 1 (e.g., # of items).  
compute nfacet1 = 10.
  - \* Enter the number of levels/conditions of Facet 2 (e.g., # of occasions); You can ignore this step for single-facet designs.  
compute nfacet2 = 2.
  - \* For two-facet designs, Facet 1 is the facet with the fastest-changing conditions in the columns of your data matrix. For example, if the first 10 columns/variables contained the data for 10 different items measured on occasion 1, and if the next 10 columns/variables contained the data for the same 10 items measured on occasion 2, then items would be the fastest-changing facet. As you slide from one column to the next across the data matrix, it is the item levels that change most quickly. You would therefore enter a value of "10" for NFACTET1 and a value of "2" for NFACTET2 on the above statements.
  - \* Enter the design of your data on the "COMPUTE TYPE =" statement below:
    - enter "1" for a single-facet fully-crossed design, as in P \* F1
    - enter "2" for a single-facet nested nested design, as in F1 : P
    - enter "3" for a two-facet fully-crossed design, as in P \* F1 \* F2
    - enter "4" for a two-facet nested design, as in P \* (F1 : F2)
    - enter "5" for a two-facet nested design, as in (F1 : P) \* F2
    - enter "6" for a two-facet nested design, as in F1 : (P \* F2)
    - enter "7" for a two-facet nested design, as in (F1 \* F2) : P
    - enter "8" for a two-facet nested design, as in F1 : F2 : P.
  - compute type = 3.
  - \* Enter D-study values for Facet 1; enter the values inside curly brackets, and place a comma between the values.  
compute dfacet1 = {5,10,15,20}.
  - \* Enter D-study values for Facet 2; enter the values inside curly brackets, and place a comma between the values. You can ignore this step for single-facet designs.  
compute dfacet2 = {1,2,3,4,5}.
  - \* At the very bottom of this file, after the END MATRIX statement, is a GRAPH command that can be used to plot the results for the D-study values that you specified above. Specify the data that you would like to plot by entering the appropriate number on the COMPUTE GRAPHDAT statement:
    - enter "1" for relative error variances;
    - enter "2" for absolute error variances;
    - enter "3" for G-coefficients;
    - enter "4" for phi coefficients.
  - compute graphdat = 3.
  - \* End of user specifications. Now just run this whole file.
-

**APPENDIX B**  
**Output From the G1 SPSS Program for the Rosenberg Self-Esteem Scale**

GENERALIZABILITY THEORY ANALYSES:

Design Type 3: two-facet fully-crossed design, as in P \* F1 \* F2

Number of persons/objects ('P'):

329

Number of levels for Facet 1 ('F1'):

10

Number of levels for Facet 2 ('F2'):

2

ANOVA Table:

	df	SS	MS	Variance	Proport.
P	328.000	11795.298	35.961	1.418	.282
F1	9.000	3383.220	375.913	.564	.112
F2	1.000	7.423	7.423	.000	.000
P*F1	2952.000	10158.830	3.441	.816	.162
P*F2	328.000	1958.227	5.970	.416	.083
F1*F2	9.000	28.156	3.128	.004	.001
P*F1*F2	2952.000	5343.694	1.810	1.810	.360

Error Variances:

Relative	Absolute
.380	.437

G-coefficients:

G	Phi
.789	.765

D-Study:

Entered D-Study values for Facet 1:

5 10 15 20

Entered D-Study values for Facet 2:

1 2 3 4 5

In the D-study results below, the levels of Facet 1 appear in the first column, and the levels of Facet 2 appear in the first row.

D-Study Absolute Error Variances

.000	1.000	2.000	3.000	4.000	5.000
5.000	1.055	.665	.536	.471	.432
10.000	.735	.437	.337	.287	.257
15.000	.629	.360	.271	.226	.199
20.000	.576	.322	.238	.196	.170

D-Study Relative Error Variances

.000	1.000	2.000	3.000	4.000	5.000
5.000	.941	.552	.422	.358	.319
10.000	.679	.380	.281	.231	.201
15.000	.591	.323	.233	.189	.162
20.000	.547	.294	.210	.167	.142

D-Study G Coefficients

.000	1.000	2.000	3.000	4.000	5.000
5.000	.601	.720	.770	.799	.816
10.000	.676	.789	.835	.860	.876
15.000	.706	.815	.859	.883	.898
20.000	.722	.828	.871	.894	.909

D-Study Phi Coefficients

.000	1.000	2.000	3.000	4.000	5.000
5.000	.573	.681	.726	.751	.767
10.000	.658	.765	.808	.832	.846
15.000	.693	.797	.840	.862	.877
20.000	.711	.815	.856	.879	.893