G String IV

Version 6.1.1

User Manual

Revised February 19, 2015

Ralph Bloch & Geoff Norman

Hamilton, Ontario, Canada

Contents

G String IV and urGENOVA

G String IV is a program to code data entry and compute generalizability coefficients based on variance component estimates from urGENOVA. It was designed and coded by Ralph Bloch at papaworx© as part of a project commissioned by The Medical Council of Canada and subsequently developed further. G String IV is written in C# for the "DOT.NET®" framework on the Windows® platform.

To contact the author, write: Ralph Bloch: <ralph.bloch@cogeco.ca>

urGENOVA was written by R.L. Brennan at the University of Iowa and can be downloaded from their website: http://www.education.uiowa.edu/casma/GenovaPrograms.htm. However, urGenova.exe is included in the installation package for G String IV.

Because G_String relies heavily on Brennan's formulation, the manual will reference appropriate pages from his book: Brennan, R.L. Generalizability Theory. New York, Springer, 2001.

To execute, G_String_III requires DOT.NET. If you do not have DOT.NET framework 3.0 or later on your computer, it can be downloaded free from the Web at:
http://www.microsoft.com/downloads/details.aspx?FamilyID=10cc340b-f857-4a14-83f5-25634c3bf043&displaylang=en

urGENOVA is a traditional command line program written in ANSI C; users must specify their parameters by means of a somewhat cryptic control file since urGENOVA does not have a graphical user interface. Also, urGENOVA has difficulties with current long directory and file names. G_String takes care of that. While urGenova provides the variance components for the individual effects, it does not calculate variance coefficients under different conditions; G_String does that as well.

Loading from disc

To load G_String and urGENOVA, load the Installation disc in your CD drive. Click on "ReadMe.txt" to see the conditions of use and run Setup.exe to install G_String and urGENOVA.

Loading from the Web

G_String can also be downloaded from the Web. It is available at the website:
http://fhsperd.mcmaster.ca/g_string/download.html

Complete instructions for downloading are on the website.

Introduction

G String IV is a program designed to complement urGENOVA, the standard software for estimating the variance components for unbalanced, random effects G-study designs. G_String queries the user through a total of 12 steps, and uses the responses to generate the control language text required by urGENOVA. When the variance components have been computed by urGENOVA, G_String calculates absolute and relative variance coefficients for the G-Study and D-Studies using specifications provided by the user.

Terminology

To clarify some of the instructions to follow, a brief introduction to the terminology of G theory is necessary. In particular, different authors use different conventions; so to avoid confusion, we have described the approach used here.

G theory is structured around different sources of variation or variance, called ***facets.*** In any measurement situation, we can consider an object of measurement **p.** In G theory, the facet associated with the object of measurement is called the ***facet of differentiation*** or ***D facet.*** In any study, there is only one facet of differentiation.

Every observation of this object of measurement is subject to error, derived from various sources depending on the "universe of observation" defined by the researcher. These sources are called ***facets of generalization*** or ***G*** facets, and address the question: "To what extent can I generalize from a measurement taken under one situation to another with a different level of the facet of generalization?" These may be of two kinds: ***fixed facets,*** which are held constant, and ***random facets***, which are the focus of the generalization. Random facets contribute to the relevant error; fixed facets do not.

> EXAMPLE: If we have a situation in which multiple raters are rating different essays by students, if we wish to look at inter-rater reliability on the same essay, ***rater*** is a random facet and ***essay*** is a fixed facet.

Every generalizability coeffficient has the same form, called an ***Intraclass Correlation,*** which is a ratio of the variance associated with the facet of differentiation, which Brennan calls $\boldsymbol{\tau}$ (tau), to the sum of this variance and the error variance, called $\boldsymbol{\delta}$ (delta) and $\boldsymbol{\Delta}$ (Delta) by Brennan. That is, the coefficient is:

$$\boldsymbol{G} = \frac{\sigma(\tau)^2}{\sigma(\tau)^2 + \sigma(\delta)^2}$$

$$or\ \frac{\sigma(\tau)^2}{\sigma(\tau)^2 + \sigma(\Delta)^2}$$

The facet of differentiation always contains $\boldsymbol{\tau}$**.** All other variance components associated with facets of generalization contribute to $\boldsymbol{\tau}$**,** $\boldsymbol{\delta}$ or $\boldsymbol{\Delta}$ (or both $\boldsymbol{\delta}$ and $\boldsymbol{\Delta}$) depending on the specific choice of G coefficient. Whether one uses $\boldsymbol{\delta}$ or$\boldsymbol{\Delta}$ depends on whether a score is to be interpreted *relative to* other scores, or in an *absolute* sense. For the relative interpretation, one uses $\boldsymbol{\delta}$***,*** and produces a G coefficient called the ‘Relative Error’ coefficient or Eρ^2. For the absolute interpretation, one uses Δ and produces a coefficient called the ‘Absolute Error’ coefficient or Φ.

The idea of relative versus absolute error deserves further explanation. It is perhaps easiest understood by example. Consider, the internal consistency, which reflects the correlation between items on a scale. Usually, there is only one version of the questionnaire and everyone always fills out the same questionnaire. Under these circumstances, the main effect of item (the variance corresponding to overall differences between items) is irrelevant, since its net effect would be to just add or subtract a constant from everyone. Under these circumstances, the appropriate error term is the relative error term, since this excludes the main effect. (If, on the other hand, there were multiple tests, such as different examinations at different sites, then the main effect might affect different people differently, and one should use the absolute error term). On the other hand, a study of inter-rater reliability would typically assume that whatever raters involved in the study were a sample of all possible raters. So any main effect of Rater would contribute error to an individual determination. In this case, one would use the absolute error term. This terminology is described in detail in Brennan.

One limitation of the Brennan formulation is that one has a limited choice between selecting the absolute error coefficient (where main effects of all facets of generalization are included) or the relative error term (where all main effects are excluded). In general the differences between the two coefficients are small and of little consequence. However, if one wishes to compute a mixed coefficient, where some facets are treated using absolute error and some with relative error, it is simply a case of computing the G coefficient by hand, including the main effects for the absolute error facets and excluding them for the relative error facets. Any interaction involving facets which are using absolute error would also be contained in the error term. The assignment of terms to $\boldsymbol{\delta}$ and $\boldsymbol{\Delta}$ in the ANOVA table produced by G String can be used as a guideline.

There is also a specific terminology associated with the design specification. A facet A may be ***crossed*** with another facet B where each level of A occurs at all levels of B, written as AB. Or A may be ***nested*** in B, where each level of A occurs at only one level of B, written as A:B.

> In the example above, if each rater rated only one essay question, then rater is nested in essay. If each rater rated all essays, rater is crossed with essay.

Finally, the facet of differentiation may be nested in various other facets, which are called ***facets of stratification*** or ***S facets.*** These will be described later.

Exclusions and Conditions

There are several exclusions and limitations in G_String. Some result from the specific approach used by urGENOVA; others are a consequence of logistical concerns.

1. Levels of nesting

G_String is restricted to only one level of nesting **of facets of generalization.** That is, there can be multiple nested facets of the form G1:G2, G1:G2:D or G1:D, but there cannot be any nested facets of the form G1:G2:G3. The reason is that for unbalanced designs and three or more levels of nesting the form of the averaged sample size is not defined.

2. Facet of Differentiation

G String is presently limited to a maximum of 1500 levels of the facet of differentiation (1500 subjects). If your design exceeds this, write to Ralph Bloch and we can furnish a modified version.

3. Stratification facets

For practical reasons, G_String will not handle more than 4 S facets. If the user has more, it is suggested that you collapse over facets that are unlikely to contribute error variance.

4. Stratification facets and facets of generalization

G_String cannot analyze any design where a facet of generalization G is nested in a facet of stratification S. This may occur quite often, for example when a user is running multiple OSCE circuits, *rater*, a G facet, is inevitably nested in *circuit*, an S facet. urGENOVA cannot analyze this. We are developing G_String V, which will deal with this situation using bootstrapping methods, but the current version cannot handle this situation.

5. Missing data and unintentional nesting

Usually, nesting is a deliberate choice of the researcher, as in the above example. However, there is one situation where nesting can arise from unintentional factors. When data are missing—for example, individual subjects completing a questionnaire skip particular questions—urGENOVA can handle only by turning ***question*** into a nested facet (so that one respondent may have responses to 6 questions, another has responses to a different 8 questions, and so on) (Brennan, urGenova Manual, Appendix F.). This approach is quite primitive, as it loses any information about systematic differences among questions. For accidentally missing data, G String IV replaces missing values implicitly by the grand mean. This solution may be adequate, if only a small fraction of scores are missing. G String IV warns the user when this occurs. Users are strongly advised to deal with missing data **before** using G_String using standard statistical interpolation methods.

6. Nested Crossed Facets

In some experiments, two or more facets may be nested in another parent facet. At this point in time, G String IV can handle a maximum of two nested facets, e.g. raters and test items nested in a station. Should the need arise to increase the maximum number of nested facets within a given parent facet, the program can be extended accordingly.

Data Structure

urGENOVA requires the data to be in ASCII text files (.dat or .txt). G_String is set up to handle tab-delimited or fixed format text files. ASCII files can be easily generated from a spreadsheet such as Excel or Quattro Pro. Simply click on "Save as" and save as a "Text - tab delimited (*.txt)" file.

Many databases have a series of index variables at the beginning of each field. G_String has the capability to set the starting point of the data array and skip over these fields, as long as they are in ASCII numeric format. Using tab-delimited text, G_String will automatically adjust the field size to accommodate the data, and will read data with varying numbers of columns, varying decimal points, etc. For fixed format, you can set the field width manually. Data must occur in sequential fields of identical width, with no other variables in between. This can usually be done with a spreadsheet program.

Like all previous versions, G String IV requires that the data be ordered, so that all records related to a particular level of a facet are together. Thus, to take a simple case: 10 students wrote an examination with 4 questions. Six students were marked by Teacher 1; 4 by Teacher 2. The data must be laid out as follows:

Teacher	Student	Question			
1	1	X	X	X	X
1	2	X	X	X	X
1	3	X	X	X	X
1	4	X	X	X	X
1	5	X	X	X	x
1	6	X	X	X	X
2	1	X	X	X	X
2	2	X	X	X	x
2	3	X	X	X	X
2	4	X	X	X	X

Interspersing records from Teacher 1 and Teacher 2 is not permitted. Records must group all observations at a single level together. This can be easily achieved using the Sort function in a spreadsheet program like Excel, before producing the .txt file.

G String IV contains a significant improvement over prior versions in its ability to automatically create counts of the number of levels of each facet. This feature is accessed at startup by selecting "Auto-Index" as described on page 10. It is particularly valuable for nested designs, where the number of levels may vary for each nest. (In one recent example, we had ratings on 1100 general practitioners by anywhere from 1 to 35 colleagues. In previous versions this would require entry of 1100 2-digit numbers into the program). To achieve this, it is necessary to define subscripts (which was not necessary for prior versions). That is, in the above example, one column must define the levels of Teacher (1 – 2); another column must define the levels of Student (1 → 10). **There is no constraint on the form of these indices except than that each level of each facet (Teacher, Student) must be unique!** That is, Teacher or Student could be defined by their names (J Smith) or their birthdates (e.g. 10 Nov 1944) as long as all records for a particular level occur together and have the same identifier.

For the repeated observations in each record (in this case, question), G String IV has two ways to handle it. If the user enters a blank in the Column box, it signals to G String that the user will enter the number of levels when prompted. This is identical to previous versions of G String. The design is then specified by the combination of nested and crossed facets and the number of levels of each on the record. For example, in an OSCE with 6 stations and the same 4 items per station, both Station and Item would have blanks entered in Column, and a total of 12 observations would appear on each record.

If the number of observations per record may vary (for example, a varying number of raters rating each teacher with a single score), then G String IV can be used to automatically determine the number of ratings for each teacher (how many observations per record. In this case, the user enters "-1" in the Column field. However, the number of levels varying from record to record only applies to one facet. So, for a second example, if students were rating teachers with a single global rating, the repeated observations on each record would be Student nested in Teacher (S:T), which may vary in number from teacher to teacher. However, if the rating had multiple items, the data base should be reconfigured so that the repeated observations on each record is Item , (typically crossed with Student and Teacher and the same number of items) and each record would correspond to a single Student's ratings of a single Teacher.

It is not absolutely necessary to add an extra index column for the facet corresponding to individual records. When no specific index exists that changes from record to record, the corresponding column field can be left blank. G_String_IV will then determine the number of records automatically.

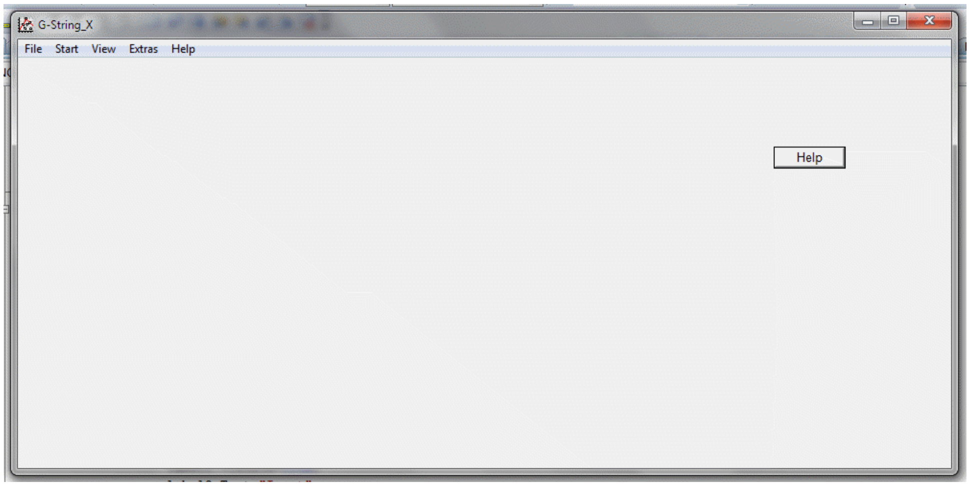
G String IV calculates the actual 'Grand Mean' and displays it at the end of the 'COMMENTS' section in the control and output file. The adjusted scores delivered to urGENOVA have been normalized by subtracting the Grand Mean from each score in order to minimize numerical errors arising from calculating small differences of large numbers . Consequently, the 'Grand Mean' displayed by urGENOVA is essentially zero. If you want to reconstitute the actual mean for each level, simply add the Grand Mean as calculated by G String IV.

Getting Started

G_String guides the user through all the steps of setting up a control file for urGENOVA, feeds the control file to urGENOVA, and allows the user to inspect and modify the control file and view the result file via a familiar Windows® user interface. G_String has built-in help screens. After urGENOVA has executed, G_String can then compute G coefficients under user control.

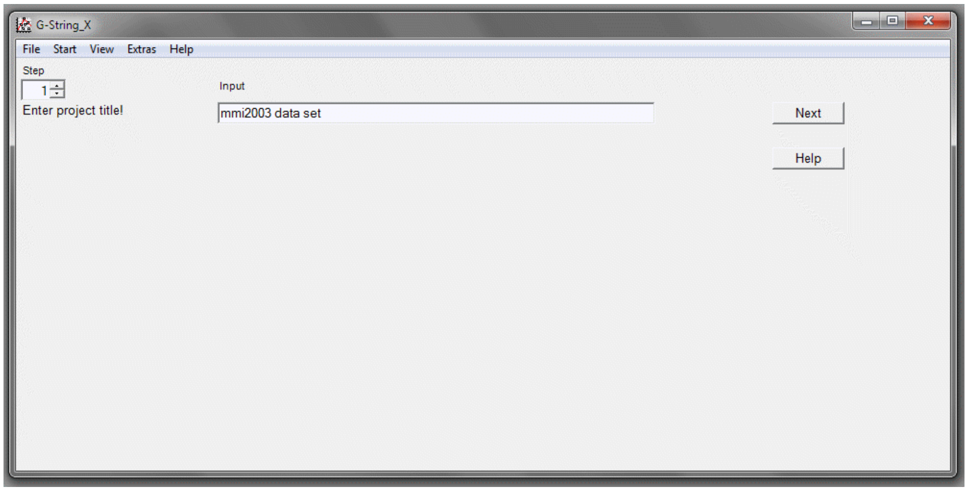
To start G_String, click on G_String.exe or a shortcut. Then, in G_String click on "Start."

At this point, a sub-menu with three options is displayed. *"Start fresh"* is the usual approach, where you are creating a new G String run and all fact and all levels of each facet will be user-specified. *"Start over"* enables you to do multiple runs of the same database, in order to perform or refine D-Studies that were not done during the initial analysis . Selecting *"Auto index"* tells G String to automatically count the number of levels of each facet. As described in detail on page 9, this is very useful for unbalanced nested designs with large numbers of subjects and/or raters.



STEP 1: Selecting a Title

"Title" can be any combination of letters and numbers up to 80 characters. It is not actually used in the calculations, so can be omitted, but it appears in the output.



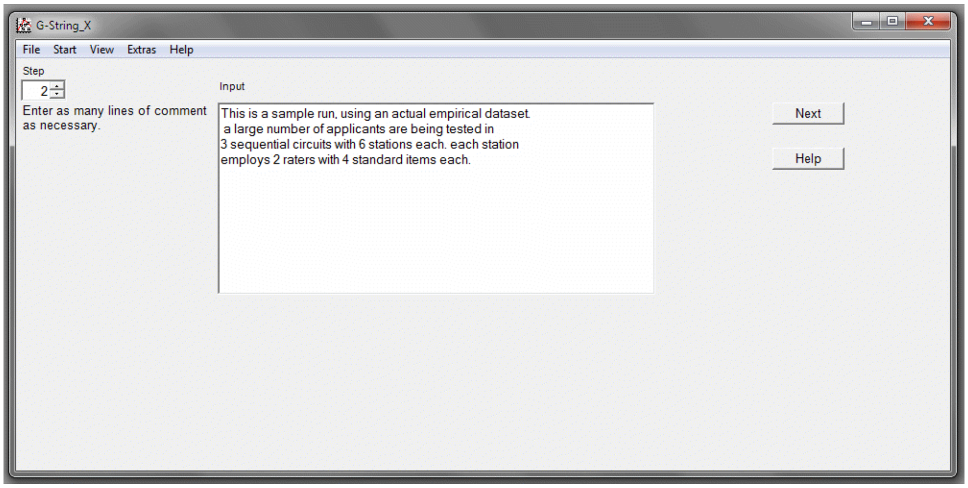
Step 1

Enter project title!

Input: mmi2003 data set

STEP 2: Entering Comments

Comment fields are optional and are not used in the calculations but copied into the result (output) file. G_String adds some comment lines automatically.



Step 2

Enter as many lines of comment as necessary.

Input:

This is a sample run, using an actual empirical dataset.
a large number of applicants are being tested in
3 sequential circuits with 6 stations each. each station
employs 2 raters with 4 standard items each.

STEP 3: Defining “Subjects”

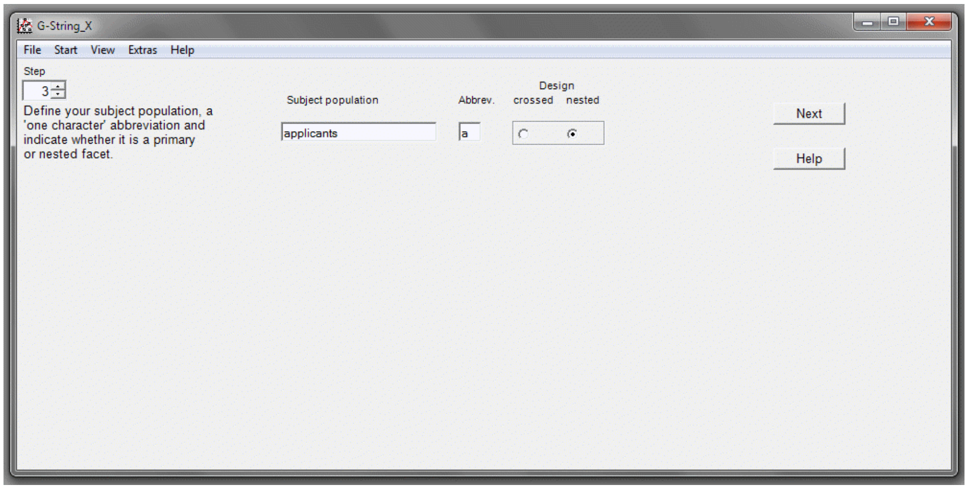
“Subject” is the variable describing the people or things that were measured in the study—the “object of measurement.” This is also the “facet of differentiation.” In Brennan’s terminology, “Subject”’ is always labelled **p**. While in G theory the designation “Subjects” is to some extent arbitrary, usually reliability or G coefficients are referenced to subjects. Usually, but not always, the data records are arranged subject by subject.

“Subject” is usually crossed with other factors, such as item or rater (e.g. a series of students being rated by 3 raters on a 10-item test), which would be the repeated measures in a simple analysis. However, “Subject” may also be nested.

> EXAMPLE: Student may be nested in Year (freshman, sophomore, senior); Patient may be nested in Gender or Physician Practice, and can be both crossed with some variables and nested with others. G_String easily deals with this situation. Facets such as Year, Gender, Physician as above are labelled "Stratification facets" and are handled somewhat differently, as will be described (Brennan, p. 153).

While, in principle, "Subject" may be nested in many stratification facets, in practice G_String is restricted to 4 stratification facets.

If *"Auto-Index"* is selected, a Column box will also be displayed. You must specify in which column of the data base the index for the "Subject" facet is located. This is described earlier on page 9.



Step 3

Define your subject population, a 'one character' abbreviation and indicate whether it is a primary or nested facet.

Subject population: applicants

Abbrev.: a

Design: crossed ○ nested ●

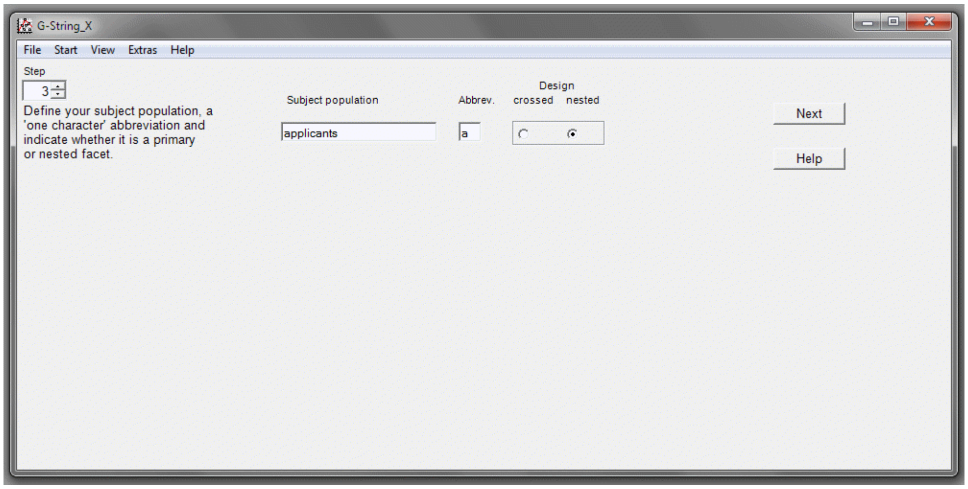
STEP 4: Defining the Other Facets

A "facet" in a design is any factor (in ANOVA jargon) or variable used to categorize the data for analysis. In G theory, "Subject" is always a factor, and is not counted explicitly at this step. Some variables are crossed with others, some are nested.

> EXAMPLE: The present example is a 6-station OSCE. There were 3 circuits (C), with 6 Applicants (A) each. Applicant is nested in Circuit. Station is crossed with Applicant (all applicants do all stations). All stations have 2 raters, with the same 4 items in each station Therefore, Item is crossed with Station but Rater is nested in Station, since each Station has its own raters but Items were constant across stations.

In Step 4, you simply specify the number of facets in addition to subjects. For the OSCE, this would be 4 (Circuit, Station, Item, Rater).

As described earlier, any number of facets with fixed levels occurring on the actual record line can be specified. For this purpose you leave the column fields empty for these facets. You will then be prompted to manually enter the actual fixed levels. If, however, the number of levels per record line has to be determined automatically, the record line may contain only one facet. In this case enter '-1' in the corresponding column field.



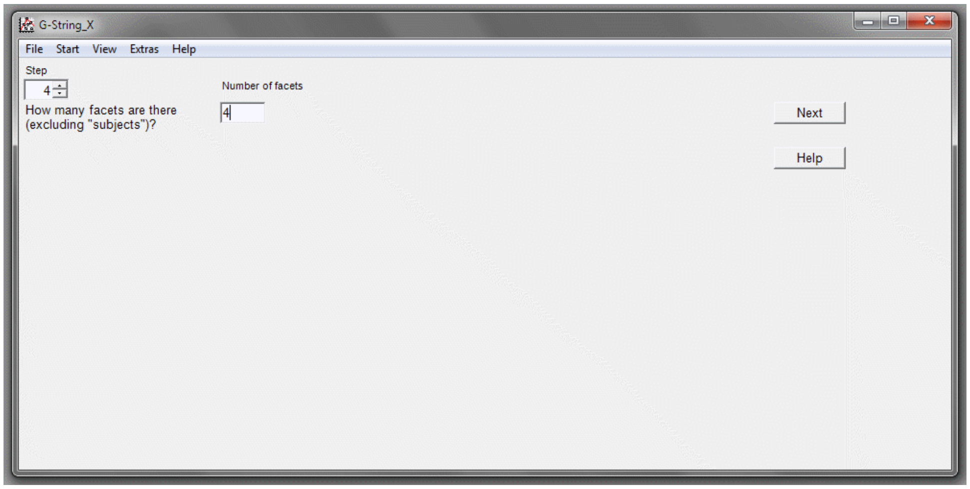
Step 3

Define your subject population, a 'one character' abbreviation and indicate whether it is a primary or nested facet.

Subject population: applicants

Abbrev.: a

Design: ○ crossed ● nested



Step 4

How many facets are there (excluding "subjects")?

Number of facets: 4

STEP 5: Naming and Specifying the Facets

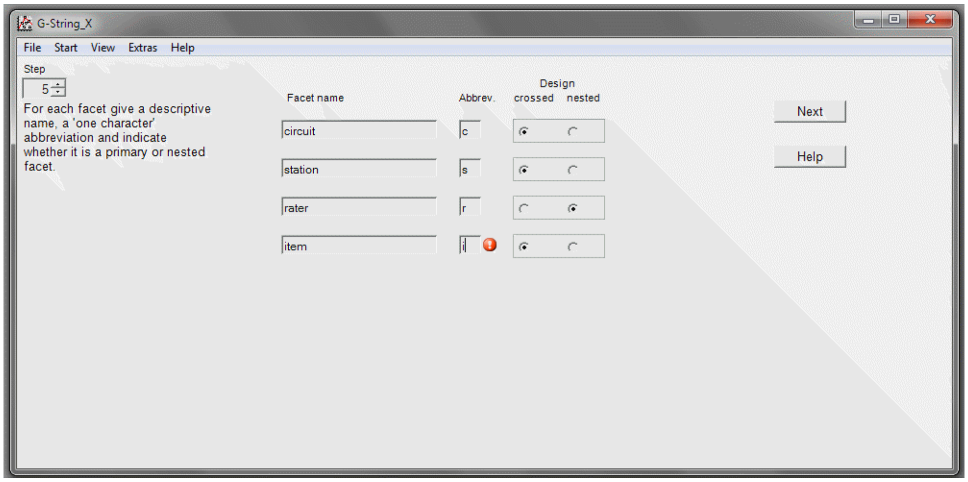
In this step, you name the facets and indicate which are nested in other facets.

- Give each facet a descriptive name and a corresponding one-character, unique, lowercase abbreviation.
- If a variable is nested in one or more other variables (see Step 4), then you change the default "crossed" to "Nested."

  In the OSCE example, Applicant is nested in Circuit, (Screen 3) and Rater is nested in Station.

- Variables must be listed in the order they are encountered in the data file, from slowest-moving to fastest.

In the OSCE example: if the data have one record per student, with all data for each station, then the data for each rater, then the responses on each item, the order of additional variables would be: Circuit, Station, Rater, Item.



Step 5

For each facet give a descriptive name, a 'one character' abbreviation and indicate whether it is a primary or nested facet.

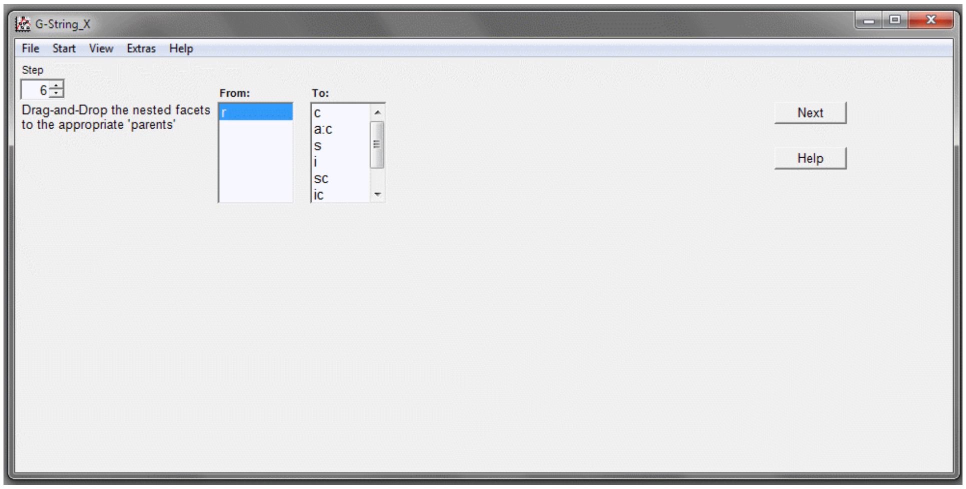
Facet name	Abbrev.	Design crossed	Design nested
circuit	c	●	○
station	s	●	○
rater	r	○	●
item	i	●	○

STEP 6: Facet Nesting

In this step, nested factors are "drag-and-dropped" to the right side so that they are located under the factor in which they are nested. Every possible combination of crossed facets is shown in this box, and a facet can be nested in more than one other facet, e.g. a:ic.

Pick a nested facet up with the mouse cursor from the list on the left and drop it on the desired combination in the list on the right.

In the example, Applicant has already been dragged under Circuit. Rater will be dragged to Station (s)



Step 6

Drag-and-Drop the nested facets to the appropriate 'parents'

From:
r

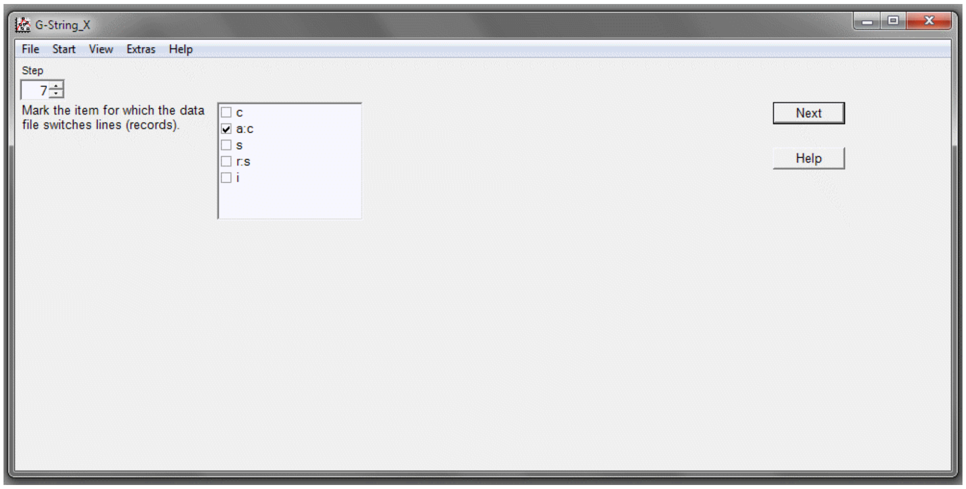
To:
c
a:c
s
i
sc
ic

STEP 7 Identifying the Data Structure

Based on the specification of nested and crossed factors in STEP 6, G_String creates a list corresponding to the order in which the data are expected to occur.

> In the OSCE example, the list would be: Subject then Station then Rater then Item, listed as:
>
> - Circuit
> - Applicant: Circuit
> - Station
> - Rater: Station
> - Item

You now specify which variable corresponds to the physical record (in Excel, each row). For example, if all data for one student was on one line, the check is put beside "Applicant: Circuit" (a:c). If each station is listed on one line (with all raters and items), the check is beside Station.



Step 7

Mark the item for which the data file switches lines (records).

□ c
☑ a:c
□ s
□ r:s
□ i

STEP 8: Specifying Sample Sizes

(If "Auto Index" is selected, the number of levels of each facet will be computed automatically and the corresponding fields will contain the appropriate number of levels. When the number of detected levels is more than 30, their value will not be displayed.)

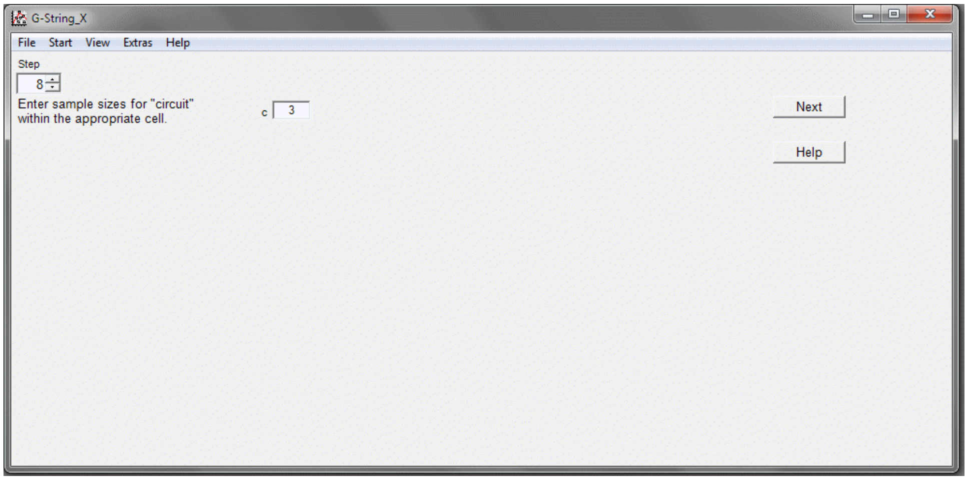
At this step, G_String cycles through all the variables you specified, and asks for "sample size." The "sample size" is the number of levels of each facet and must be > 1.

> In the OSCE example, "sample size" for Station is just the number of Stations

For *nested* variables, you must specify the number of levels at each level of the nesting variable.

> For Subject, this is the number of Applicants in each Circuit 6, 6, 6. For Rater this will be the number of raters per station, 2, 2, 2, 2, 2, 2.

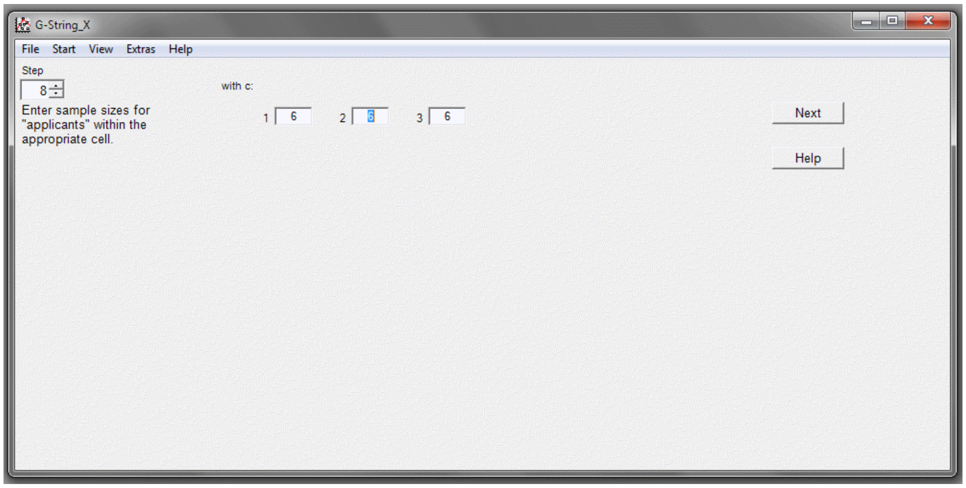
As a default, once you enter the levels for the first box and press the 'tab' key, G_String will automatically assign the same number of levels for all boxes. If the numbers differ, simply overwrite the pre-assigned numbers. The sequence below illustrates how all the levels are being entered.



Step 8

Enter sample sizes for "circuit" within the appropriate cell.

c 3

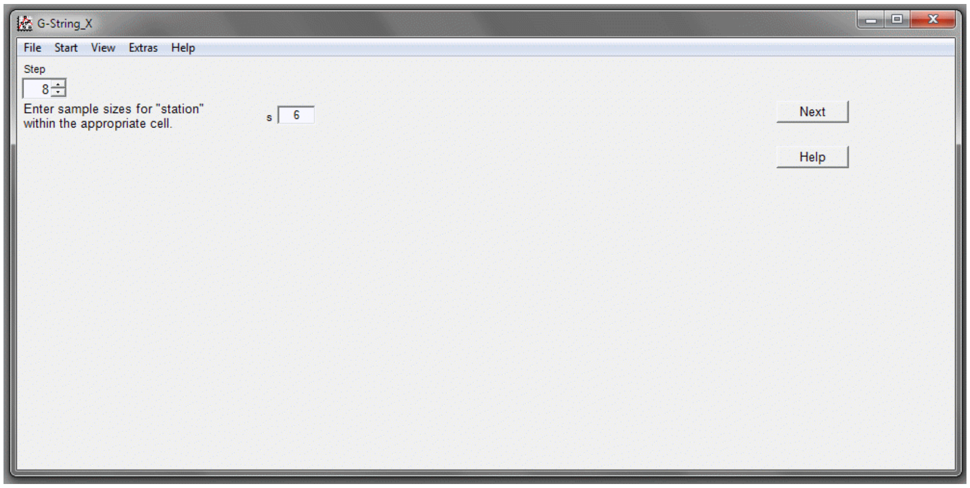


Step 8

Enter sample sizes for "applicants" within the appropriate cell.

with c:

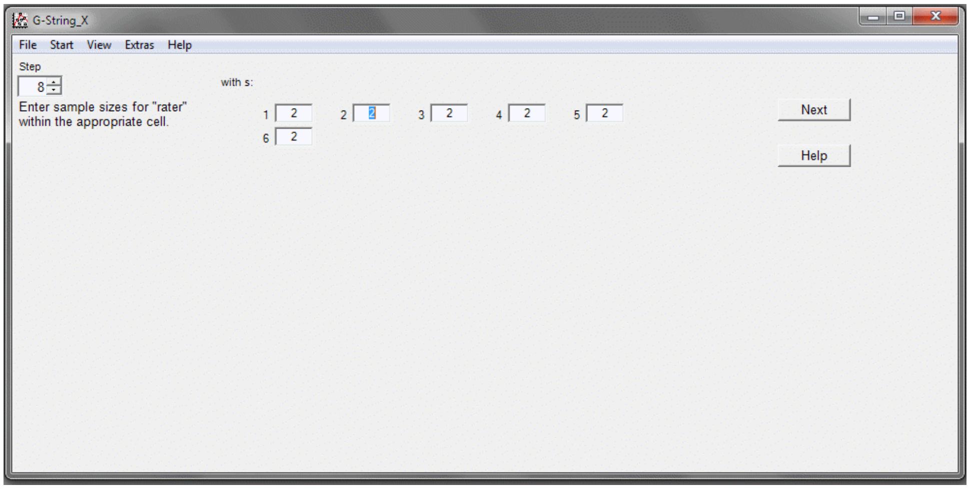
1 6 2 6 3 6



Step 8

Enter sample sizes for "station" within the appropriate cell.

s 6

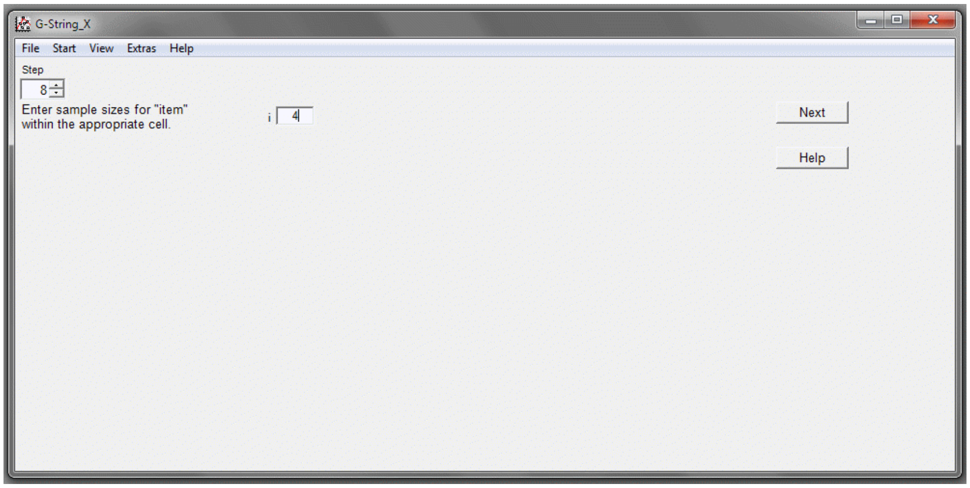


Step 8

Enter sample sizes for "rater" within the appropriate cell.

with s:

1 2   2 2   3 2   4 2   5 2
6 2



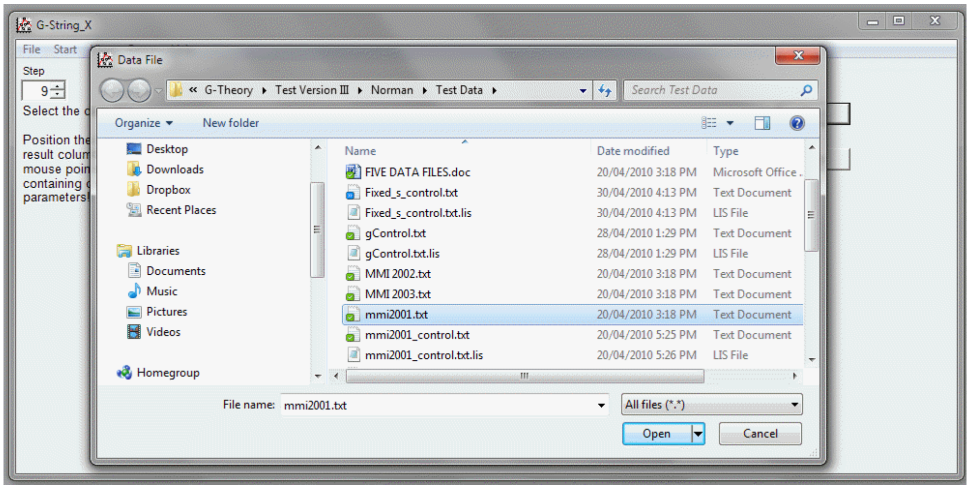
Step 8

Enter sample sizes for "item" within the appropriate cell.

i 4

STEP 9: Locating and Specifying Data File

At this step, you first tell G_String where the data file is located using the usual Browse function. G_String then reads the first few records from this file. It assumes that the actual data are listed sequentially beginning at a specific column of each data line in the data file. Recall that data must be in an ASCII text file.



Select the c

Position the result colum mouse poin containing parameters!

Name	Date modified	Type
FIVE DATA FILES.doc	20/04/2010 3:18 PM	Microsoft Office .
Fixed_s_control.txt	30/04/2010 4:13 PM	Text Document
Fixed_s_control.txt.lis	30/04/2010 4:13 PM	LIS File
gControl.txt	28/04/2010 1:29 PM	Text Document
gControl.txt.lis	28/04/2010 1:29 PM	LIS File
MMI 2002.txt	20/04/2010 3:18 PM	Text Document
MMI 2003.txt	20/04/2010 3:18 PM	Text Document
mmi2001.txt	20/04/2010 3:18 PM	Text Document
mmi2001_control.txt	20/04/2010 5:25 PM	Text Document
mmi2001_control.txt.lis	20/04/2010 5:26 PM	LIS File

File name: mmi2001.txt

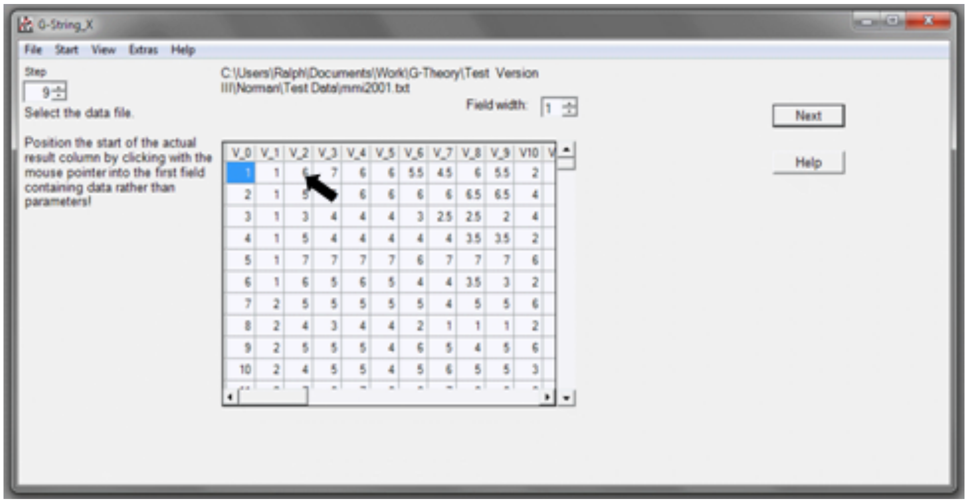
All files (*.*)

You then select the column where the actual data start by mouse-clicking directly on the first cell containing data (in any row). urGENOVA will ignore anything to the left of this.

For tab-delimited files, G_String will create the correct field width. For fixed field data (no delimiters), first set the start column as above, then with the "Field Width" selector, indicate the width in columns of each individual data field (including blanks).

In the example, the first 2 columns are identifiers, so the cursor is placed in the 3rd column.

The cursor arrow must be located in the first actual data field, *not* on the headers.



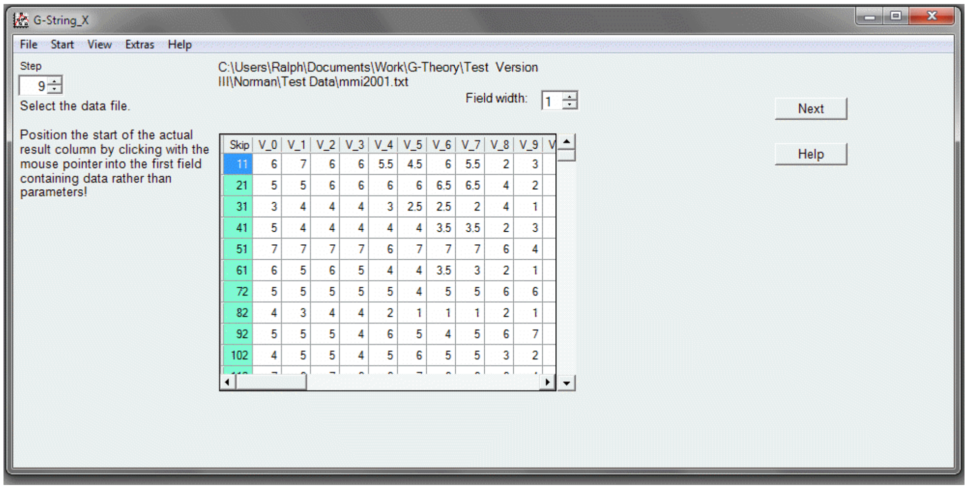
C:\Users\Ralph\Documents\Work\G-Theory\Test Version III\Norman\Test Data\mmi2001.txt

Field width: 1

Select the data file.

Position the start of the actual result column by clicking with the mouse pointer into the first field containing data rather than parameters!

V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V10
1	1	6	7	6	6	5.5	4.5	6	5.5	2
2	1	5	[illegible]	6	6	6	6	6.5	6.5	4
3	1	3	4	4	4	3	2.5	2.5	2	4
4	1	5	4	4	4	4	4	3.5	3.5	2
5	1	7	7	7	7	6	7	7	7	6
6	1	6	5	6	5	4	4	3.5	3	2
7	2	5	5	5	5	5	4	5	5	6
8	2	4	3	4	4	2	1	1	1	2
9	2	5	5	5	4	6	5	4	5	6
10	2	4	5	5	4	5	6	5	5	3



Step 9

Select the data file.

Position the start of the actual result column by clicking with the mouse pointer into the first field containing data rather than parameters!

C:\Users\Ralph\Documents\Work\G-Theory\Test Version III\Norman\Test Data\mmi2001.txt

Field width: 1

Skip	V_0	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9
11	6	7	6	6	5.5	4.5	6	5.5	2	3
21	5	5	6	6	6	6	6.5	6.5	4	2
31	3	4	4	4	3	2.5	2.5	2	4	1
41	5	4	4	4	4	4	3.5	3.5	2	3
51	7	7	7	7	6	7	7	7	6	4
61	6	5	6	5	4	4	3.5	3	2	1
72	5	5	5	5	5	4	5	5	6	6
82	4	3	4	4	2	1	1	1	2	1
92	5	5	5	4	6	5	4	5	6	7
102	4	5	5	4	5	6	5	5	3	2

STEP 10: Options

urGENOVA allows you to specify a number of options. G_String assumes some default values that you don't have to change, unless you know what you are doing.

NREC: the number of data records that will be printed in the output file. Useful to check that the data are being read as expected.

Outname: the name of the output file. This will be assigned a name and stored in the same directory as the data file, unless you choose a new name and directory.

ET prints the expected T term equations.

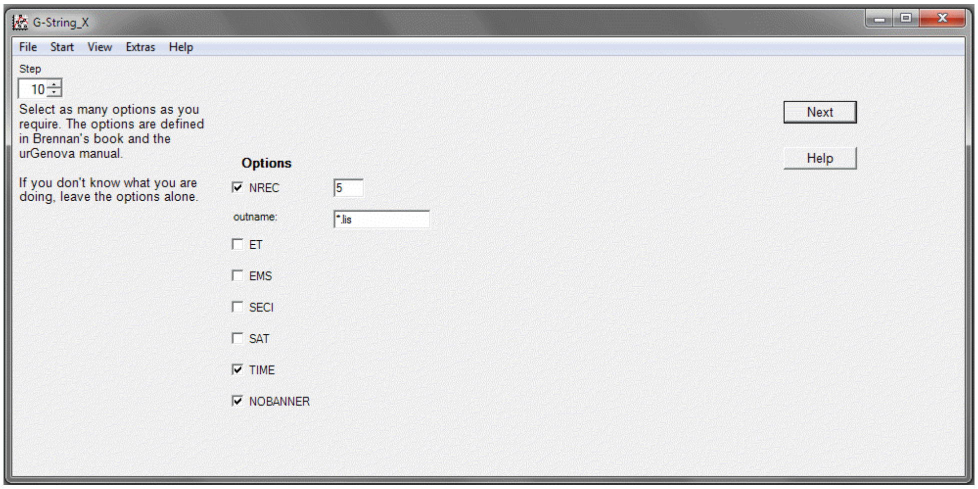
EMS prints the equations for the expected mean squares as sums of variances.

SECI .nn is the standard error and ."nn" confidence interval for the estimated variance component (.nn is a fraction between .00 and 1.0, usually .95).

SAT is a second confidence interval estimate, due to Satterthwaite (see the GENOVA manual).

TIME Time and date of processing will be printed (default is ON).

NOBANNER Banner will not be printed (default is ON).



Step 10

Select as many options as you require. The options are defined in Brennan's book and the urGenova manual.

If you don't know what you are doing, leave the options alone.

Options

☑ NREC 5

outname: *.lis

☐ ET

☐ EMS

☐ SECI

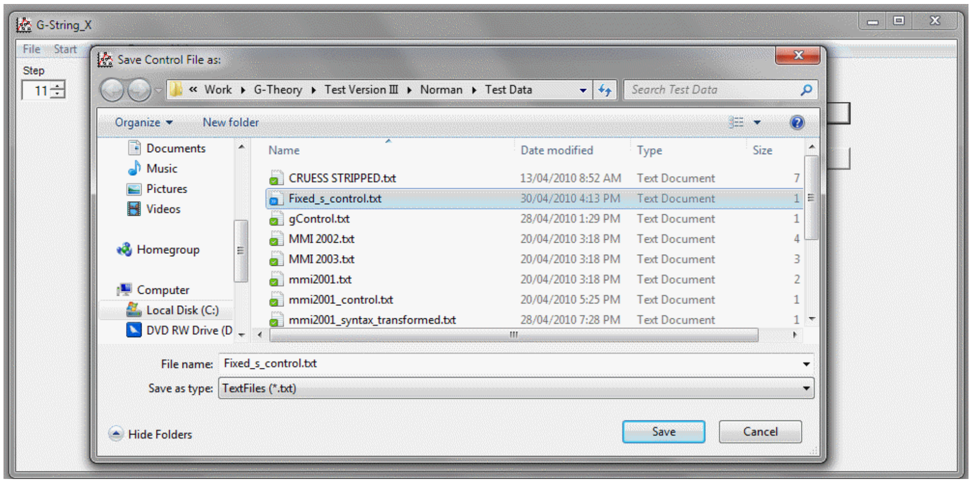
☐ SAT

☑ TIME

☑ NOBANNER

STEP 11: Save Control File

You have now completed the specification and generated a control language file. By default, it is called "gControl.txt" and stored in the same directory as the data file; however, at this step you can give it a more meaningful name and place it in any directory of your choice.



Step 11

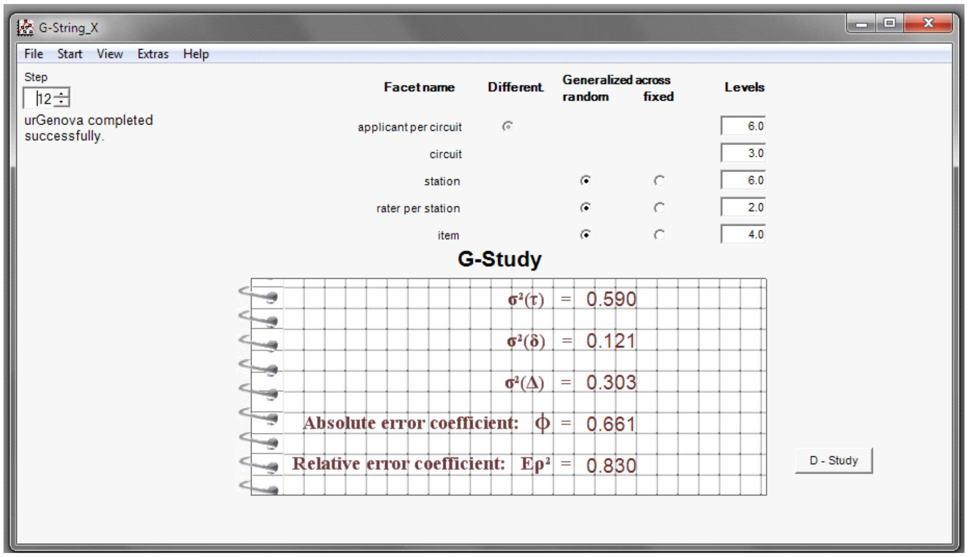
Name	Date modified	Type	Size
CRUESS STRIPPED.txt	13/04/2010 8:52 AM	Text Document	7
Fixed_s_control.txt	30/04/2010 4:13 PM	Text Document	1
gControl.txt	28/04/2010 1:29 PM	Text Document	1
MMI 2002.txt	20/04/2010 3:18 PM	Text Document	4
MMI 2003.txt	20/04/2010 3:18 PM	Text Document	3
mmi2001.txt	20/04/2010 3:18 PM	Text Document	2
mmi2001_control.txt	20/04/2010 5:25 PM	Text Document	1
mmi2001_syntax_transformed.txt	28/04/2010 7:28 PM	Text Document	1

File name: Fixed_s_control.txt

Save as type: TextFiles (*.txt)

STEP 12: Calculating Variance Components

Once you saved the proper control file path, urGENOVA is executed automatically to calculate the variance components and the coefficients of variance for the G study are generated.



Step 12

urGenova completed successfully.

Facet name	Different	Generalized across random	fixed	Levels
applicant per circuit	●			6.0
circuit				3.0
station		●	○	6.0
rater per station		●	○	2.0
item		●	○	4.0

G-Study

$\sigma^2(\tau) = 0.590$

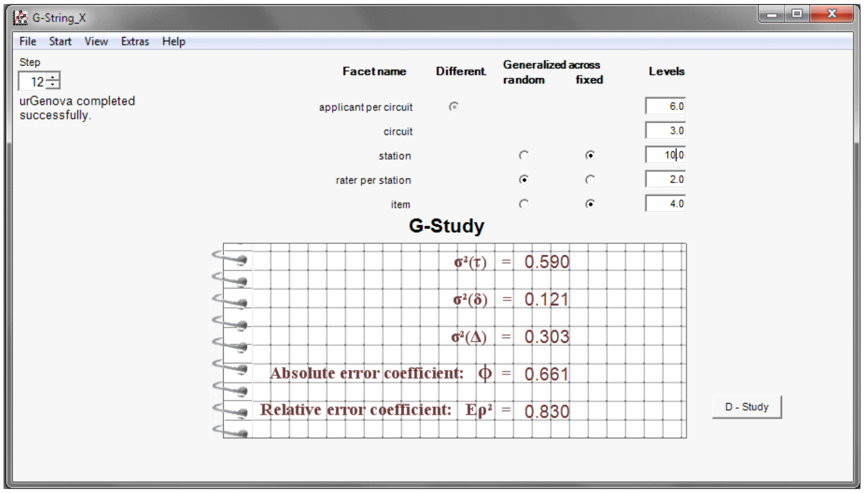
$\sigma^2(\delta) = 0.121$

$\sigma^2(\Delta) = 0.303$

Absolute error coefficient: $\Phi = 0.661$

Relative error coefficient: $E\rho^2 = 0.830$

STEP 13: Coefficients of Variance in G-Study and D Studies



Step 12

urGenova completed successfully.

Facet name	Different	Generalized across random	fixed	Levels
applicant per circuit	●			6.0
circuit				3.0
station		○	●	10.0
rater per station		●	○	2.0
item		○	●	4.0

G-Study

$\sigma^2(\tau)$ = 0.590

$\sigma^2(\delta)$ = 0.121

$\sigma^2(\Delta)$ = 0.303

Absolute error coefficient: Φ = 0.661

Relative error coefficient: $E\rho^2$ = 0.830

This screen displays the output from the calculation of the G coefficient, and then permits the user to conduct repeated D studies. The output follows the convention of Brennan, in particular the rules for calculation of G coefficients (4.1.6, p. 109) and the section on Mixed Models (4.3, p. 120). A brief explanation is required.

Generalizability theory is an extension of classical test theory (CTT). In CTT, every observation is comprised of a True Score or Signal, and Error. The reliability coefficient is the ratio of the True Score VAR(τ) to the total variance (VAR(τ) + VAR (δ)). G theory extends this formulation by considering that error may have multiple sources, which we have called "facets of generalization." Depending on the measurement situation, you may wish to generalize over some facets (called "random" facets by Brennan), and keep others constant (called "fixed" facets by Brennan).

> In the OSCE, if we set Rater as random and Item and Station as fixed, we will compute the equivalent of the Inter-Rater Reliability. If we set Item as random and fix Rater and Station, we are computing the equivalent of internal consistency.

The calculation amounts to moving variance components between the error term δ and the Signal term τ. Screen 11 displays δ and τ as well as Δ, described next.

There is a further refinement in G theory. Sometimes, we wish to interpret a person's score *relative* to those of other people. In this situation, the fact that some raters may be more strict or lenient than others, or some items harder or easier, is irrelevant. This amounts to ignoring the main effects of the facets of generalization, and only interactions with Subject are included. This is the error term δ. However, if we wish to put an absolute interpretation on scores, we must include main effects, which is the term Δ on Screen 12. In turn, the Absolute Error coefficient or Φ contains Δ whereas the Relative Error coefficient or $E\rho^2$ contains δ. See page 7 earlier for further explanation.

Some additional comments about stratification facets are necessary. You may stratify for two reasons:

a) To test an experimental validity hypothesis (for example, senior students will do better than junior students), which we will call an “experimental stratification facet”, or

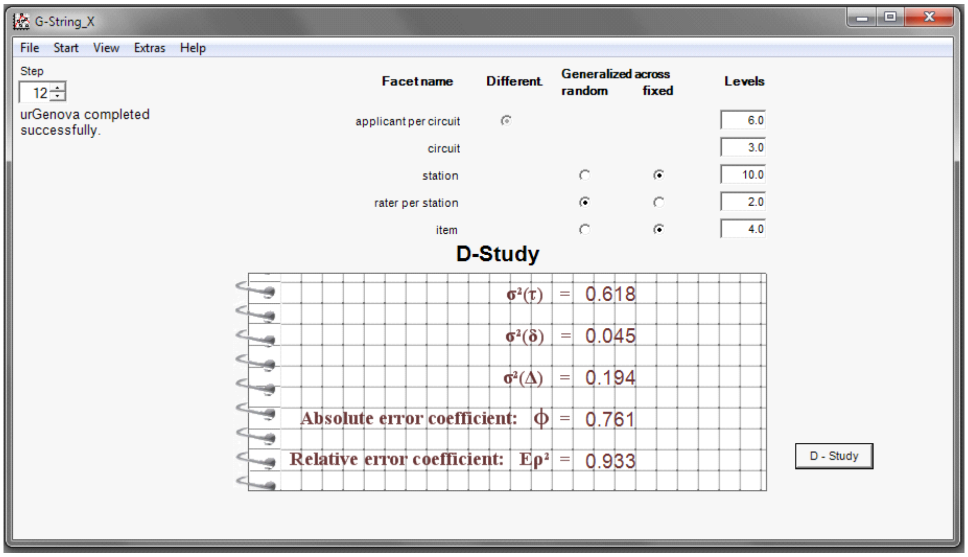
b) The stratification may be a result of the logistics of the test administration, which we will call a “logistical stratification facet.”

In the first case, you are only interested in person variance within each stratum, so the Relative Error term should be used. In the second, variation due to stratification contributes error in interpretation of an individual score, so the Absolute Error term should be used, since variance due to Stratification facets is included in Δ. (Stratification is introduced in Brennan p. 153, but this is a more complete treatment).

The first automatic output on this screen considers all facets as facets of generalization. Further, it computes averages over each facet, based on the sample sizes in the original study. So the calculated coefficient Eρ^2 is the G coefficient for the original test.

However, on Screen 11, G_String will calculate G coefficients with any combination of fixed facets and facets of generalization, and any sample sizes—so-called D studies—in order to examine the effect of each facet on the overall generalizability. You can also calculate the equivalent of classical coefficients by ‘treating’ one facet at a time as ‘random’ and fixing the remaining facets.

> In the OSCE example, if you want to compute the equivalent of inter-rater reliability in the OSCE, you would: a) Set Item and Station as Fixed facets, and b) Set the sample size for Rater=1. (If you keep sample sizes for Item and Station, you are calculating inter-rater for the average of N_s = 6 stations and N_1 = 4 items). More likely you would also fix sample sizes for Item and Station at 1 to determine inter-rater reliability for a single rater in one station with one item.



Step 12

urGenova completed successfully.

Facet name	Different	Generalized across random	fixed	Levels
applicant per circuit	●			6.0
circuit				3.0
station		○	●	10.0
rater per station		●	○	2.0
item		○	●	4.0

D-Study

$\sigma^2(\tau)$ = 0.618

$\sigma^2(\delta)$ = 0.045

$\sigma^2(\Delta)$ = 0.194

Absolute error coefficient: ϕ = 0.761

Relative error coefficient: Eρ^2 = 0.933

If you wish to calculate different coefficients (D studies), simply re-enter the new combination of facets, identifying fixed facets and facets of generalization, and the new sample sizes and click on “compute.” The new coefficient will be calculated and displayed in the screen and in the printout. Refer to page 22 and examples in Appendix 2. for detailed explanations.

Note that, in the case of nested variables, the number of levels is within each nest. For D studies you must keep this constant across nests so it is a “balanced” design.

In the OSCE study, there are 6 applicants per circuit and 2 raters per station.

Interpreting the Output

The computer output contains many more details of the above calculations and will be described next. This output is generated when the process of study calculation is finished, and is created as a ‘.txt.lis’ file in the target directory. Below is a sample output from the example. Annotations are in this font. On some computer operating systems you may have to delete the secondary extension ‘.lis’ to be able to read the file.

```
                              CONTROL CARDS FOR RUN 1
                       Control Cards File Name:  ~Temp.txt
                                 mmi2003 data set

GSTUDY   mmi2003 data set
COMMENT
COMMENT  Processing date: 06/06/2010 2:49:22 PM
COMMENT
COMMENT  This is a sample run, using an actual empirical dataset.
COMMENT  a large number of applicants are being tested in
COMMENT  3 sequential circuits with 6 stations each. each station
COMMENT  employs 2 raters with 4 standard items each.
COMMENT
COMMENT
COMMENT%  applicant  (a)
COMMENT%  circuit  (c)
COMMENT%  station  (s)
COMMENT%  rater  (r)
COMMENT%  item  (i)
COMMENT
COMMENT  The calculated 'Grand Mean' = 4.4010.
COMMENT  G_String III normalizes scores by subtracting the Grand Mean from each score.
COMMENT
OPTIONS  NREC 5 "*.lis" TIME NOBANNER
EFFECT     c         3
EFFECT   * a:c       6 6 6
EFFECT     s         6
EFFECT     r:s       2 2 2 2 2 2
EFFECT     i         4
FORMAT   30  0
PROCESS  "~Temp.dat"
```

This is an image of the control card input for urGENOVA created by G_String in response to user input. Note how the “EFFECT” lines completely describe the design, with circuits, applicants nested in circuits (6 / circuit), stations, rater nested in station (2/station) and item. The calculated Grand Mean over all the scores is 4.4010.

```
                              INPUT RECORDS FOR RUN 1
                                 mmi2003 data set

RECORD NUMBER 1:
  1.599   2.599   1.599   1.599   1.099   0.099   1.599   1.099  -2.401  -1.401
```

urGENOVA images the data on the first 5 records. Some are omitted from this example. The Grand Mean has been subtracted from the actual scores.

```
                         MEANS FOR MAIN EFFECTS FOR RUN 1
                                mmi2003 data set

Means for c
-0.045   0.528  -0.483

Means for a:c
  0.089  -0.078  -0.839  -0.214   0.932  -0.161   0.849  -1.339   1.297  -0.318
  0.995   1.682  -0.943  -0.672   0.745  -0.016  -1.130  -0.880

Means for s
  0.238  -0.283  -0.102   0.207   0.203  -0.262

Means for r:s
  0.418   0.057  -0.887   0.321   0.030  -0.234   0.314   0.099  -0.679   1.085
 -0.873   0.349

Means for i
  0.205  -0.119  -0.047  -0.040
```

urGENOVA outputs the means for each variable.

Below is the ANOVA table created by urGENOVA. The format is conventional, except that the right column is “variance component” and is used in the calculation of G coefficients. (Negative variance components are set to zero when computing G coefficients)

```
                          ANOVA TABLE FOR RUN 1
                            mmi2003 data set
-----------------------------------------------------------------------------
Effect              df             T            SS            MS            VC
-----------------------------------------------------------------------------
c                    2   16882.85677     147.89583      73.94792       0.10650
a:c                 15   17393.52604     510.66927      34.04462       0.58973
s                    5   16778.13368      43.17274       8.63455      -0.27757
r:s                  6   17005.35069     227.21701      37.86950       0.46282
i                    3   16747.93634      12.97541       4.32514       0.01591
cs                  10   17075.20312     149.17361      14.91736       0.17605
cr:s                12   17351.61458      49.19444       4.09954       0.02040
ci                   6   16898.97569       3.14352       0.52392       0.00179
as:c                75   18021.96875     436.09635       5.81462       0.27813
ar:cs               90   18619.43750     321.05729       3.56730       0.82002
ai:c                45   17420.10417      10.45920       0.23243      -0.00642
si                  15   16803.64583      12.53675       0.83578      -0.00231
ri:s                18   17044.84722      13.98438       0.77691       0.02484
csi                 30   17118.02083      14.16204       0.47207       0.01000
cri:s               36   17420.29167      11.87500       0.32986       0.00711
asi:c              225   18144.87500      69.62934       0.30946       0.01113
ari:cs             270   18845.75000      77.54688       0.28721       0.28721
-----------------------------------------------------------------------------
Mean                     16734.96094
-----------------------------------------------------------------------------
Total              863                  2110.78906
-----------------------------------------------------------------------------
        Grand Mean:         0.
```

Below is the first output from G_String. It is a calculation of the overall test generalizability, so a) there are no fixed facets, and b) the number of levels of each facet corresponds to the original study.

The allocation of individual terms is based on the specification of random or fixed facets. This is according to the rules in Appendix 1, abstracted from Brennan. Basically, in this case, the facet of differentiation, a, is in Tau; all interactions with Tau are in delta, and all interactions and main effects are in Δ.

```
Date and time at beginning of Run 1:  Sun Jun  6 14:49:22 2010
Processor time for run: 0 seconds

Computation sequence for G-Study
'a'    Differentiation          6.00
'c'     Stratification          3.00
's'             Random          6.00
'r'             Random          2.00
'i'             Random          4.00
------------------------------------------------------------------------------------
----

Pattern   Var. Comp.    Levels      Signature      Rule
------------------------------------------------------------------------------------
----
c              0.1065    1                   s     Delta only
a:c            0.5897    1                   ds    tau only
s              0.0000    (6.0)               r     Delta only
r:s            0.0386    (12.0)              r     Delta only
i              0.0040    (4.0)               r     Delta only
cs             0.0293    (6.0)               r     Delta only
cr:s           0.0017    (12.0)              r     Delta only
ci             0.0004    (4.0)               r     Delta only
as:c           0.0464    (6.0)               dr    Delta and delta
ar:cs          0.0683    (12.0)              dr    Delta and delta
ai:c           0.0000    (4.0)               dr    Delta and delta
si             0.0000    (6.0*4.0)           r     Delta only
ri:s           0.0005    (12.0*4.0)          r     Delta only
csi            0.0004    (6.0*4.0)           r     Delta only
cri:s          0.0001    (12.0*4.0)          r     Delta only
asi:c          0.0005    (6.0*4.0)           dr    Delta and delta
ari:cs         0.0060    (12.0*4.0)          dr    Delta and delta

RESULTS:

s2(T)        = 0.590
s2(D)        = 0.303
s2(d)        = 0.121
Er2          = 0.830
Phi          = 0.661
```

The first 5 outputs are shown in Screen 11:

Below is an example of D studies. The user can control two aspects of the computation: a) which facets are random and which are fixed, and b) how many levels of each. These are used for different purposes:

Random vs. Fixed Facets.

In G theory, one can compute the equivalent of classical coefficients such as inter-rater reliability, internal consistency, and so on, by restricting the analysis, setting one facet at a time as random, and setting the "n" for this facet equal to 1.

> In the example above, to compute inter-rater reliability for a single rating and a single station, one would declare ***rater*** as random, ***station*** and ***item*** as fixed, and set all the levels equal to 1. If one wanted the inter-rater reliability of the total score over all 4 items, no. of levels of ***item*** would remain 4. To look at internal consistency (across items) ***item*** becomes the random facet, ***rater*** and ***station*** fixed, and levels remains at 4 (since internal consistency is for the total score, so amounts to averaging by number of terms).

The number of levels is a matter of judgment, and is based on whether the reliability is for a single (item, rater, station) or for the mean across all items, raters, and stations. To understand how this works, we have taken the above example and created a number of D-study scenarios:

Random Facet(s)	Fixed Facet(s)	N_{rater}	N_{item}	$N_{station}$	Interpretation
S	R,I	2	4	1	Inter-station reliability of total score from 2 raters and 4 items
S	R,I	1	1	1	Inter-station reliability for any single item from any rater
S	R,I	2	4	6	Inter-station reliability for total score from 2 raters and 4 items
R	S,I	1	4	6	Inter-rater reliability for total score from 4 items, 6 stations
R	S,I	1	4	1	Inter-rater reliability for total score on any station
R	S,I	1	1	1	Inter-rater reliability for any item, any station
I	R,S	1	4	1	Internal consistency (across items) for 1 rater, 1 station
I	R,S	1	1	1	Average inter-item correlation
I	R,S	2	1	1	Average inter-item correlation for mean of 2 raters

Changing levels – D Studies.

To this point, we have set the number of levels as either the original design number or 1, depending on whether we wish to compute reliability for the single item or the number of levels of the facet in the original study. We can also vary the number of items at will, to determine the optimal combination of levels of each facet in the design. In this case, the interest is in the overall test reliability, so there are no fixed facets, but we might vary number of levels at will.

Note that when we proceed with D studies, the design is balanced by definition, since we input the number of levels of each facet as a single number. Thus unbalanced designs only arise in the initial calculation of the G coefficient from the original data.

> For example, are we better to have 6 stations with 2 raters (Nr=2, Ns=6), or 12 stations with 1 rater (Nr=1, Ns=12)? What do we gain in going from 12 stations to 18?

```
Computation sequence for D-Study
'a'     Differentiation             6.00
'c'      Stratification             3.00
's'              Random            12.00
'r'              Random             1.00
'i'              Random             4.00
-------------------------------------------------------------------------------------------
```

```
Pattern   Var. Comp.    Levels      Signature      Rule
------------------------------------------------------------------------------------------
c              0.1065    1                   s     Delta only
a:c            0.5897    1                  ds     tau only
s              0.0000    (12.0)              r     Delta only
r:s            0.0386    (12.0)              r     Delta only
i              0.0040    (4.0)               r     Delta only
cs             0.0147    (12.0)              r     Delta only
cr:s           0.0017    (12.0)              r     Delta only
ci             0.0004    (4.0)               r     Delta only
as:c           0.0232    (12.0)             dr     Delta and delta
ar:cs          0.0683    (12.0)             dr     Delta and delta
ai:c           0.0000    (4.0)              dr     Delta and delta
si             0.0000    (12.0*4.0)          r     Delta only
ri:s           0.0005    (12.0*4.0)          r     Delta only
csi            0.0002    (12.0*4.0)          r     Delta only
cri:s          0.0001    (12.0*4.0)          r     Delta only
asi:c          0.0002    (12.0*4.0)         dr     Delta and delta
ari:cs         0.0060    (12.0*4.0)         dr     Delta and delta

RESULTS:

s2(T)       = 0.590
s2(D)       = 0.264
s2(d)       = 0.098
Er2         = 0.858
Phi         = 0.690


Computation sequence for D-Study
'a'     Differentiation             6.00
'c'      Stratification             3.00
's'              Random            18.00
'r'              Random             1.00
'i'              Random             4.00
------------------------------------------------------------------------------------------

Pattern   Var. Comp.    Levels      Signature      Rule
------------------------------------------------------------------------------------------
c              0.1065    1                   s     Delta only
a:c            0.5897    1                  ds     tau only
s              0.0000    (18.0)              r     Delta only
r:s            0.0257    (18.0)              r     Delta only
i              0.0040    (4.0)               r     Delta only
cs             0.0098    (18.0)              r     Delta only
cr:s           0.0011    (18.0)              r     Delta only
ci             0.0004    (4.0)               r     Delta only
as:c           0.0155    (18.0)             dr     Delta and delta
ar:cs          0.0456    (18.0)             dr     Delta and delta
ai:c           0.0000    (4.0)              dr     Delta and delta
si             0.0000    (18.0*4.0)          r     Delta only
ri:s           0.0003    (18.0*4.0)          r     Delta only
csi            0.0001    (18.0*4.0)          r     Delta only
cri:s          0.0001    (18.0*4.0)          r     Delta only
asi:c          0.0002    (18.0*4.0)         dr     Delta and delta
ari:cs         0.0040    (18.0*4.0)         dr     Delta and delta

RESULTS:

s2(T)       = 0.590
s2(D)       = 0.213
s2(d)       = 0.065
Er2         = 0.901
Phi         = 0.734
```

Appendix 1: Rules used to compute G coefficients

In this section we describe the rules used to generate the G coefficients. The notation follows the conventions used in Brennan. The rules are derived primarily from Brennan, in particular, the rules specified in Chapter 4, p.122, rules 4.3.1—4.3.3 , However, the formalism of Stratification Facets and the rules related to the contribution of these facets to the G coefficients are new. These rules have been incorporated in the algorithms used by G String to generate the coefficients. They are intended for reference only.

DEFINITIONS

Rule 0:

Facets are of 3 types:

1) Facet of differentiation, p, defined in Screen 3. There is only 1 p.

2) Stratification facets, S1, S2…. These are of the form p:S1,S2, defined in Screen 4,5.

Note: The term "Stratification" is consistent with the terminology of Brennan, Section 5:2.

3) Facets of Generalization: G1,G2,G3… (defined in Screen 4,5).

Facets are of three types: (i) Differentiation; (ii) Stratification; (iii) Generalization.

Facets of Generalization are subdivided further into random and fixed facets of generalization.

Rule 1:

Facets of stratification (S_i) appear in ANOVA (and eventually in G String IV), but cannot be facets of generalization or differentiation

Note 1: One implication of Rule 1 is that NO S_i will appear in the formulae to calculate the coefficients in Screen 12.

Facets of Stratification can be recognized by the fact that they provide containers for a nested facet of Differentiation.

Rule 2:

Nesting of variables may arise in several different ways.

a) P:S_i —by definition, P can only nest in S_i. These are handled in Rule 1.

b) G_i:P

c) G_i:G_j

These are handled slightly differently in the rules to follow.

Note: Nesting of facets results in elimination of certain interactions in the ANOVA, but these are handled automatically by urGENOVA. There are also implications for the division by "n" in the D studies.

Rule 3:

Facets of Generalization are specified as of two types in the calculation of G coefficients: Random facets, Rj, and Fixed facets Fk. These are specified in Screen 12 (and can be changed by the user on successive calculations).

RULES FOR CREATING τ, δ, Δ

Rule 4: (Brennan Rule 4.3.1, p. 122)

τ = {p (including p:S) +all p x F_k interactions not containing any R_j + all Main effects of form F_i:p not containing any R_j}

> *NOTE 1: The reason behind this rule with respect to nested variables is that with fully crossed design, τ contains all interactions between p and F but not the main effect of F. With nested design, the variance due to nesting (e.g. VAR(F_i:p)) actually contains the pxF$_i$ interaction so is in τ term. See Brennan p. 123, where he says explicitly that, in the design pxI:H, Delta = VAR(pi:h)/$n_i'n_h'$ = VAR(pi)+ VAR(pih)/ $n_i'n_h'$. See also note in δ* below.
>
> *Note 2: When a facet is a facet of generalization, its main effect will be in Δ. However, when it is a fixed facet, (if it is not nested in p as below), the main effect does not move to τ. See Brennan section 4.4.1. He states that "fixing a facet affects which variance components contribute to τ and δ but it does not change their sum." However, in the example it DOES change sum of τ and Δ since, when facet is random its main effect is in Δ but when it is fixed, main effect is not in τ*

All effects that contain the facet of differentiation but no random facet of Generalization contribute to $\sigma^2(\tau)$.

Rule 5 (Brennan Rule 4.3.3, p. 122)

δ= {all terms containing p and R_j}, including specifically all terms of form p$x R_i x R_j$. R_j:p, pxR_j:F_i, pxF_i:R_j, pxF_ixR_j.

> *NOTE: The reason behind this rule with respect to nested variables is that, with fully crossed design, δ contains all interactions between p and R_j. With nested design, the variance due to nesting (e.g. VAR(R_j:p), VAR(pxF_i:R_j)) actually contains the pxR_j interaction(R_j + R_j x p in the first case, PxF_i + pxF_ixR_j in the second case) so is in error term. See Brennan p. 123, where he says explicitly that, in the design pxI:H, Delta = VAR(pi:h)/$n_i'n_h'$ = [VAR(pi)+ VAR(pih)]/ $n_i'n_h'$. Since it contains interaction between **p** and the random variable, R_j (implicitly) , it goes into δ from Brennan's Rule 4.3.3.*

All effects that contain the facet of Differentiation and at least one random facet of Generalization contribute to $\sigma^2(\delta)$.

Rule 6 (Brennan Rule 4.3.2, p.122)

Δ = {all terms containing R_j OR all terms containing S_j }, specifically including all main effects of R_j, all interactions of form p x R_j , all interactions between R_j and other facets, e.g. R_jF_k and R_jR_k. Also, all terms containing S_j to left of colon including main effect of S, and all interactions between S_j and G facets; but excluding terms where S_j is to the right of the colon.

All effects that contain at least one random facet of Generalization and all effects that contain a Stratification facet (unless the S facet is to the right of the colon) contribute to $\sigma^2(\Delta)$

Note: Stratification facets are of two types—those that might be termed "experimental," where there is an anticipation that there will be a large main effect of the stratification facet (e.g. Educational Level) and those that are part of the design, but the expectation is that there would be no effect of S (e.g. Day, Circuit, in an OSCE). For experimental strata: (a) the remaining facets are crossed (every stratum gets the same measures (for example, all persons get the same test items)) since this is the only way that one can test hypotheses about differences, (b) the RELATIVE ERROR term is appropriate, as generalization is within facet. For design facets, the strata may contribute error in interpretation of particular scores, so the appropriate term is the ABSOLUTE ERROR term.

RULES FOR CREATING THE DIVISOR OF EACH FACET IN THE G COEFFICIENT

Rule 7

For balanced designs, the divisor of each facet in a term in τ, δ, or Δ (except for terms involving P or S) is the number of levels of the facet in the term. For terms in p or s the divisor is always 1. For nested facets of the form g_1: g_2 the divisor is n_{g1} x n_{g2}

Note: This is formalized throughout Brennan in the use of the h, H notation, where VAR(H) = VAR(h)/n_h

For balanced designs, the divisor for each facet of generalizabiity is the number of levels of the facet. For p and s facets, the number of levels is always 1. For nested facets of form g_1: g_2 the divisor is n_{g1} x n_{g2}

Unbalanced designs are of three types, and each requires different treatment.

Type 1: ***p***:S (Person nested in a stratification facet)

Since each person can only be situated in a single stratum, the divisor for any term involving a stratification facet is always 1.

Type 2: G:***p*** (Facet of generalization nested in person)

In this design, at least one facet is nested within the facet of differentiation. A typical example is where students in a class rate their teacher, or employees in a company rate their supervisor, so each teacher is rated by different students and each supervisor by different employees. In this case, the G coefficient is divided by the number of levels of person (which is set at 1) x the *average* number of levels of rater, which according to Brennan is computed as the 'Harmonic mean"—basically the average of the (1/n) terms, 1 for each ***p.***

$$\tilde{n}_g = \frac{n_p}{\sum_p \frac{1}{n_{g:p}}}$$

For unbalanced designs of form G:***p***, the divisor is the harmonic mean of the number of observations at each level of G.

Type 3: $G_1 : G_2$ (One facet of generalization nested in another)

This situation, where one facet of generalization is nested in another, is encountered quite frequently. Some examples:

- a case-based test, where each case has different questions (and different numbers of questions)
- a questionnaire with different questions in each subscale
- an OSCE, with different checklists with different numbers of items, in each station

In this situation, as in balanced designs, there is an "n" associated with the appearance of G_1:G_2 and another associated with G_2. For G_1, the "sample size" is simply the total number of observations of $G_{2,}$ which Brennan calls "n_+" (p. 219, p. 232). For a balanced design, this is just $n_{G1:G2}$ x n_{G2}.

$$n_+ = \sum_{g2} n_{g1:g2}$$

For example, if 5 subscales had 2, 4, 5, 7, and 11 items each, n+ = 29. If 5 subscales had 3 items each, n+ would be 5 x 3 = 15.

However, the imbalance in G_1 also affects the denominator of $G_{2.}$ According to Brennan (p. 232), whenever G_2 appears, the variance will be divided by $\breve{n}_{g2}$ which equals:

$$\breve{n}_{g2} = \frac{{n_+}^2}{\sum_{g2} {n_{g1:g2}}^2}$$

In the above example, this equals $29^2 / (2^2 + 4^2 + 5^2 + 7^2 + 11^2) = 841/215 = 3.9$.

The basic formulation can be extended to more than one nested facet of generalization by computing the product of the harmonic means.

Type 4: G:***p***:S, G_1 G_2 :***p***:S   (Facet of generalization nested in person nested in facet of stratification)

This design is an extension and combination of Type 2 and Type 1. Person is nested in one or more stratification facets and at least one facet is nested within the facet of differentiation. As in Type 2, this may arise from a situation where students in a class rate their teacher, or employees in a company rate their supervisor. However now the fact if differentiation is also nested (for example, teacher within school or supervisor within company). As before, the divisor for terms involving S is always 1, since each person can only occur at one level of S. Similarly, for the facet of differentiation, ***p,*** the G coefficient is divided by 1. And as before for the facets of generalization nested in the stratification and differentiation facets, the divosor is the harmonic mean

Thus, in the G:***p***:S design, the specific terms are S, ***p***:S , and G:***p***:S. Variance due to S and G:***p***:S would be in Δ term; only G***p***:S would be in the δ term. The divisors for these terms in the G coefficient are S/1, ***p***:S / 1, and G***p***:S/ $\tilde{n}_{g.}$

where as before $\tilde{n}_g$ is the 'Harmonic mean"— the average of the (1/n) terms***.***

In the G_1G_2:***p***:S design, the specific terms are S, ***p***:S , G_1:***p***:S, G_2:***p***:S, and G_1G_2:***p***:S. As in the single G facet case, divisors for ***p*** and for S are 1; any term involving a facet of generalization (G) will be divided by the harmonic mean of the n's contributing to the variance. And the product term G_1G_2:***p***:S is divided by the product of the two harmonic means.

RULES FOR DIVISION IN INTERACTION TERMS

Rule 8

For every interaction term in τ, δ, or Δ, the divisor is the product of the divisors of the facets making up the interaction. Thus a term of the form G x H will be divided by the product of the divisors of the individual terms as defined in Rule 7.

The divisor of interaction terms is the product of the divisors of the individual facets

APPENDIX 2: EXAMPLES OF G - STUDY RESEARCH DESIGNS

In this section, we describe a number of common designs ranging from simple, classical one factor reliability designs reformulated in G theory nomenclature, to complex multi-facet nested designs. We have obviously not exhausted the possibilities, but rather have attempted to identify and provide examples of some of the more common designs.

The intent is to demonstrate how each design is formulated in the notation of G String. We describe each design, then reformulate the design in G theory language. We describe any specific requirements for the format of the input data, and the sequence of inputs to the screens required to specify the design. Finally, we show how to iterate values on Screen 12 to conduct a variety of D studies.

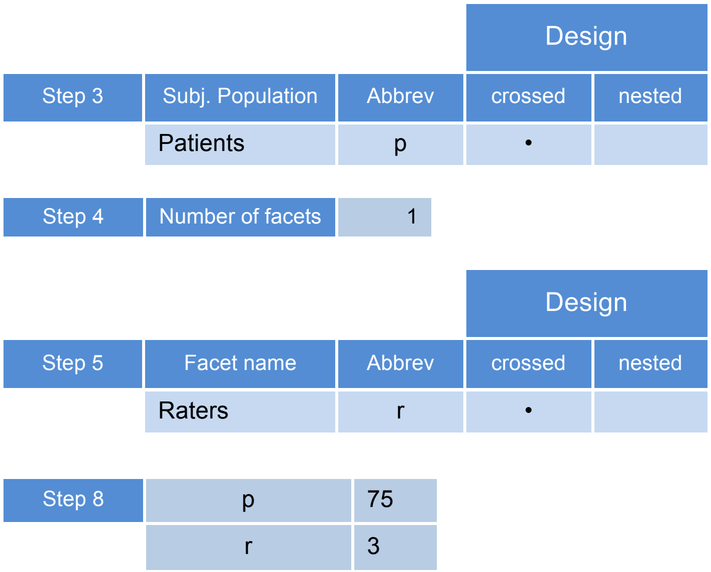
1) One Facet Designs

DESIGN 1.1. Inter-Rater Reliability

A clinical researcher examines clinician judgment of severity of illness for patients with Congestive Heart Failure. She locates complete records of 75 patients, and distributes these to 3 respirologists, who rate each case on a 0-100 scale, where 100 is "Perfect Health.

This example is a typical design for classical test theory. However, for illustrative purposes, we will recast it as a G theory study. The facet of differentiation is Patient; the single facet of generalization is Rater. The design is crossed.

The input screens would resemble:



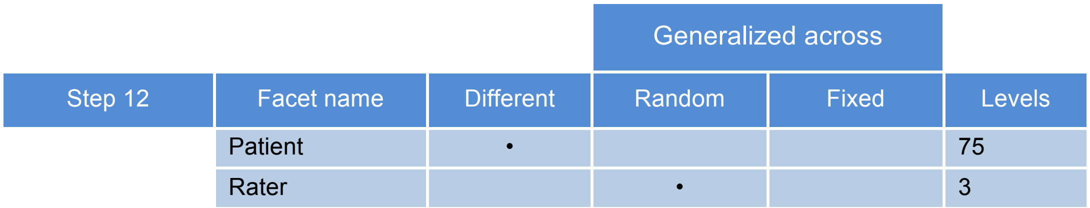
			Design	
Step 3	Subj. Population	Abbrev	crossed	nested
	Patients	p	•	

Step 4	Number of facets	1

			Design	
Step 5	Facet name	Abbrev	crossed	nested
	Raters	r	•	

Step 8	p	75
	r	3

The G Study output automatically generated on Screen 12 would look like:



			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Patient	•			75
	Rater		•		3

Note that the computed G coefficient is for the average of all raters. To calculate Inter-rater reliability for a single rater, you enter "1" as levels for Rater, and rerun.

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Patient	•			75
	Rater		•		1

GENERAL TIP:

Often people distinguish between agreement on nominal variables , which should be analyzed with Kappa or Weighted Kappa, and reliability with measured variables, which can be analyzed with ANOVA methods and intraclass correlations. However, Fleiss and Cohen (1963) showed the two methods are mathematically identical. This means that you can use the power of G theory even with data like 1= Dead, 2 = Alive. See Health Measurement Scales, 4th ed. pp.187-188.

Design 1.2 Questionnaire

The researcher administers a questionnaire on “learning style” with 25 questions and “Strongly Agree’ → Strongly Disagree” and 7 point scales to a sample of first year medical students (n = 125). He analyses the data to calculate the Internal Consistency reliability (Cronbach’s α)

Again, this can be handled with classical test theory, however we will cast it in G theory framework. The facet of differentiation is “Student” (s) with 125 levels and the facet of generalization is “Item” (i) with 25 levels. Typically the data would be laid out on a spreadsheet with 125 lines, and 25 columns. Input screens would look like

			Design	
Step 3	Subj. Population	Abbrev	crossed	nested
	Students	s	•	

Step 4	Number of facets	1

			Design	
Step 5	Facet name	Abbrev	crossed	nested
	Items	I	•	

Step 8	s	125
	I	25

The design is formally equivalent to the previous design. The G Study output automatically generated on Screen 12 would look like:

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Student	•			125
	Item		•		25

However, in this case, no further analysis is necessary. Internal consistency is the reliability of the average score or total score across all items (Health Measurement, pp.89-93), which is the G coefficient computed automatically. We could then do D Studies varying number of items to determine the effect n reliability.

GENERAL TIP:

Although use of parametric methods like ANOVA for Likert type scales comes under frequent criticism, there is strong evidence that these methods are completely acceptable. See Health Measurement, p.53

Design 1.3 Teacher Rating

A researcher examines the reliability of teacher ratings. The analysis is based on the total score over 5 items, with 5 point “Agree” → “Disagree” responses. There are 5 teachers involved in the study, with each teacher responsible for a different section. Varying numbers of students completing the ratings – Teacher 1 – 12 students; Teacher 2 – 17 students; Teacher 3 – 9 students; Teacher 4 – 15 students; Teacher 5 – 22 students

This design introduces a new concept – *nested* facets. Student (s) is nested in Teacher (t); since each student can appear with only one teacher. The design is also *unbalanced* – different numbers of students per teacher.

In laying out the data, it is important to note that, while each row in the spreadsheet will likely contain the 5 ratings of each student, in contrast to the previous examples, the facet of differentiation is not equivalent to the row. We are differentiating Teachers, and Student now is a *rater* of the teacher, so Student is the facet of generalization. **Because G String identifies data by location in the data base, not identifier, all records for each teacher must appear in sequence in the data base.**

The input screens would now look like:

			Design	
Step 3	Subj. Population	Abbrev	crossed	nested
	Teacher	t	•	

Step 4	Number of facets	1

			Design	
Step 5	Facet name	Abbrev	crossed	nested
	Student	s		•

Step 6
You declare the nature of the nesting in Screen 6, by dragging "s" (on the left) to "t" (on the right)

Step 8	t	5				
	s	12	17	9	15	22

Note the differing number of levels for student at each level of teacher The G Study output automatically generated on Screen 12 would look like:

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Teacher	•			5
	Student:Teacher		•		13.7

Note the fractional number of levels of Student. This is because the harmonic mean is used for these calculations (See page 28) . You can proceed to do D studies, to determine the relation between number of raters and reliability by simply overwriting the "Levels" in Student and recalculating.

2) Two Facet Designs

DESIGN 2.1 Raters and Items

To examine the reliability of the abstract review process for a recent conference, the Chair assembled 30 abstracts at random, and had 5 judges rate each abstract on 4 items -- Creativity , Methodological Rigour, Analysis, Practical Relevance, each with 5 point Poor → Excellent scales.

This is a straightforward two facet, crossed design. However, it is critical to recognize that the "object of measurement" is not a person (the rater) but the abstract. The data must be laid out with with raters grouped within abstracts – that is, Abs 1 -- Rater 1, Abs1 -- Rater 2, Abs 1 -- Rater 3, Abs1 -- Rater 4, Abs1 -- Rater 5, Abs2 -- Rater 1, Abs 2 -- Rater 2, Abs2 -- Rater 3, etc. These may occur on the same or separate lines (which is handled in Screen 7) but must occur in this sequence.

The input screens would now look like:

			Design	
Step 3	Subj. Population	Abbrev	crossed	nested
	Abstract	a	•	

Step 4	Number of facets	2

			Design	
Step 5	Facet name	Abbrev	crossed	nested
	Rater	r	•	
	Item	I	•	

Step 8	a	30
	r	5
	I	4

The G Study output automatically generated on Screen 12 would look like:

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Abstract	•			30
	Rater		•		5
	Item		•		4

G String automatically computes the G coefficient corresponding to the average score over 5 raters and 4 items (dividing error variances by 5, 4, or 20). You can also modify this screen to calculate the G theory equivalent of inter-rater reliability and internal consistency (alpha, α). To do this, the general strategy is to set the facet of interest as a random facet and set the other facets as fixed facets. You then modify the number of levels of the facets. The basic idea is that the number of levels of each facet is the number of observations that will be used to average the error variance, either of random or fixed facets.

Thus, if you wish to examine inter-rater reliability, “i” is set as fixed. Then the number of levels of “r” is set to 1, since, as described in Example 1.1, you want to compute the reliability of a single rater. If you want to compute inter-rater reliability of the total score, no. of levels of “i” remains at 4; if you want to compute the inter-rater reliability for a single rating, “i” is set to 1. The possibilities, then, are:

Inter-rater – one item:

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Abstract	•			30
	Rater		•		1
	Item			•	1

Inter-rater – average score:

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Abstract	•			30
	Rater		•		1
	Item			•	4

Internal consistency (α):

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Abstract	•			30
	Rater			•	1
	Item		•		4

Average inter-item correlation:

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Abstract	•			30
	Rater			•	1
	Item		•		1

GENERAL TIP

It is always important to be very careful in determining which facet represents the “object of measurement” or equivalently, the facet of differentiation. As in the example above, it is not always the people who are completing the questionnaire. Serious errors can result. Further, the data may be analyzed with different facets of generalization, depending on the question. (See Health Measurement, p. 241 for an example)

DESIGN 2.2 Questionnaire with Multiple Subscales

A researcher assesses quality of life for a cohort of patients (n=50) with multiple sclerosis using a quality of life scale with 3 subscales – Physical – 20 items; Social – 12 items; Emotional – 7 items. She examines internal consistency from the single administration

The study is quite common. Essentially, from the single administration, you can examine internal consistency within scale and between scales. The facet of differentiation is "Patient" (p) with 50 levels; there are two facets of generalization: Subscale (s) (3 levels) and item nested in subscale (i:s) , (20, 12 and 7 levels). The data would typically have one line per patient, with 39 observations on each. Input would look like:

			Design	
Step 3	Subj. Population	Abbrev	crossed	nested
	Patient	p	•	

Step 4	Number of facets	2

			Design	
Step 5	Facet name	Abbrev	crossed	nested
	Scale	s	•	
	Item	I		•

Step 6: drag "I" from left to "s" on right.

Step 8	p	50		
	s	3		
	I	20	12	7

The G Study output automatically generated on Screen 12 would look like:

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Patient	•			50
	Scale		•		2.6
	Item:Scale		•		13

Note the unusual number of levels for both Scale and Item:Scale. These formulae are described on page 28.

The G coefficient represents the internal consistency of the overall scale consisting of the 3 subscales with variable number of items. You can then compute various other combinations, similar to the D study manipulations in the previous example.

1. Generalizability across scales:

   Set Scale Random, Item Fixed. Set number of Levels for scale = 1, leave Items: Scale at 13. This then computes the average correlation between scale scores.

2. Generalizability across items within scale:

   Set Scale Fixed, Item:Scale random. Set number of levels for Scale = 1, leave Items: Scale at 13. This then is the *average* internal consistency within each subscale.

   However, generally, one would report the internal consistency of each scale individually since the number of items and the specific items vary across scales

   To do this, you would do separate runs for each subscale, using item as the only facet of generalization, as in Design 1.2, and using the feature of Screen 9 to change the starting point.

3. Overall internal consistency, independent of subscales:

   Simply rerun as Design 1.2 , with Item having 39 levels.

Note that it is difficult to compare alphas derived from different scales as alpha is sensitive to the number of items in the scale.

3) Multiple Facet Designs

The introduction of additional facets involves additional complexity, but no new concepts. The critical steps are to first identify object of measurement, then label the various additional facets in the design, identify which are nested and which are crossed, and then ensure that the sequence of data in the spreadsheet lines up with the intended design.

4) Stratification Facet Designs

One other class of designs that is very common in generalizability studies in medical education. Particularly for performance tests like OSCE’s and Oral examinations, it is very common to run the examination at multiple sites over several days. In these circumstances, each subject can be said to be nested in a particular “Stratum” of a stratification facet (Day, Site). To complicate things further, it is very common to change raters, or in the case of OSCE’s, to also change the specific stations to ensure test security. Thus, both Participant (p) and possibly Station and Rater are nested in one or more “Stratification” variables – Site, Day, Circuit.

GENERAL TIP

G String and urGENOVA are not currently capable of dealing with designs when a facet of generalization is nested in a stratification facet. The next version (GString V) will have this capacity.

DESIGN 4.1

You are running an OSCE which is taking place in two different hospitals. Students (p) are randomly assigned to one hospital or the other. At each hospital the same 12 stations are used. Three circuits are run at hospital A; for a total of 36 students and 4 circuits at Hospital B, for a total of 48 students. Each station has a station – specific checklist with anywhere from 12 to 27 items.

This is a very typical OSCE setup identifying the facets from slowest (supraordinate) to fastest (subordinate). The first stratification variable is Hospital (h) with 2 levels, then Circuit:Hospital (c:h) with 3 and 4 levels. Then Participant: Circuit and Hospital (p:c:h). Crossed with this is Station (s) and Item:Station (i:s).

Data need to be laid out consistently with this hierarchy, likely with one physical record per applicant or per station. As before, caution must be exercised to ensure that the records are grouped according to this hierarchy.

The Screens will now look like:

			Design	
Step 3	Subj. Population	Abbrev	crossed	nested
	Participant	p		•

Step 4	Number of facets	4

			Design	
Step 5	Facet name	Abbrev	crossed	nested
	Hospital	h	•	
	Circuit	c		•
	Station	s	•	
	Item	i		•

Step 6: drag “c” to “h”, “p” to “c:h” and “i” to “s”.

Step 8	h	2											
	c:h	3	4										
	p:c:h	12	12	12	12	12	12	12	12	12	12	12	12
	s	12											
	I:s	14	22	17	...	....	...	...	..	..	..	..	..

The G Study output automatically generated on Screen 12 would look like:

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Particip. per Circuit	•			12
	Hospital				2
	Circuit per Hospital				3.4
	Station		•		10.7
	Item per Station		•		18.5

Note that a) Hospital and Circuit do not have an asterisk. This signifies that they are stratification facets. b) the number of levels for Station and Item:Station contain fractions, which reflects the unbalanced design (p. 27).

The resulting G coefficient is the overall test reliability. D studies can be conducted using the strategies discussed previously to examine the average inter-station correlation (S random, I fixed, n(s) = 10.7) or the internal consistency among items within station (I random; S fixed n(i) = 18.5).

What about the stratification facets? Basically, any variance due to the stratification facet represents a bias, so that one circuit or hospital is, on average, harder or easier than another. The hope or expectation is that these variances will be small. If participants are judged relative to others in the same stratum, this variance is of no consequence, as reflected in the G coefficient for "Relative Error". However, if absolute interpretation is placed on scores, variance due to strata is a source of error. Therefore, it has to be included in the Absolute Error" calculation.

DESIGN 4.2

You are running an OSCE which is taking place with residents at two levels. Residents (r) are either PGY1 (36 residents) or PGY4 (48 residents). Residents go through the OSCE 12 at a time, with all residents at each level together. Each station has a station – specific checklist with anywhere from 12 to 27 items.

This design is deliberately set up to be identical in layout to the previous study. The only difference is the meaning attached to one stratification facet. In the previous example, Hospital was the supraordinate facet, and the expectation (or hope) was that this would contribute no variance. Any variance due to Hospital was treated as error variance which would confound interpretation of scores.

Thus the Absolute Error coefficient best represented the overall generalizability. In the present case, the expectation is that difference in educational would be large, amounting to a test of construct validity. The statistical test can be easily extracted from the G String ANOVA table.

Differences with level should NOT be treated as error. Nevertheless, Education should be in the design, since the G coefficient is then determining the ability of the test to differentiate among residents *within* an educational level, which is completely appropriate. Omitting this facet would then result in a G coefficient that is biased upwards, since differences between levels now become incorporated in variance due to facet of differentiation (resident).

5) Nested Designs

There is one final class of designs that is very common in generalizability studies in medical education. This is the situation where there are multiple and variable numbers of ratings on the object of measurement, with rating nested in the object of measurement (g:d) designs. One example is teacher ratings, where students in each class rate their teacher. Student is nested in Teacher, and numbers of students will likely vary. Peer assessments of practicing physicians, called "360 degree evaluation" or "multi-source feedback" is another -- different peers with different numbers of observations for each physician. Typically these are not the only two facets, since often ratings are on multi-item questionnaires, so the design would be Peer nested in Doctor crossed with Item. Another common variant is the so-called "mini-CEX" where each student is observed on a number of occasions by her supervisor(s), and again, typically each student has different supervisors.

Frequently these designs can have very many observations. One study involved over 1000 physicians rated by 17,000 peers. Another was based on a teacher evaluation system at a large university and had 65,000 observations on 1700 teachers. To handle these studies in previous versions of G String is very tedious as the number of observations in each nest had to be entered manually. However, with G IV, all one need do is assign a unique index to each teacher or physician and another unique index to each rater, creating two column variables. G IV will read these indices and automatically create the correct number of levels in each nest.

There is one common variant of this design. Frequently the same rater may be involved in multiple ratings of the subject. For example, with students in community clinical rotations, each student may receive multiple observations and ratings **from the same rater.** This I handled in G String simply by creating a third "sequence" index which is unique for each rating, so that the design become g2:g1:d (Sequence:Rater:Student).

While this design can be analyzed, ***extreme caution must be exercised in interpretation.*** The problem is that with multiple ratings from each rater, rater variance (lenient -- stringent) is confounded with subject variance. In the extreme case, where each subject is rated by one rater, different for each subject, rater and subject variance are completely confounded. One can obtain high G coefficients, but the value is biased upwards since this results from variance due to rater and variance due to subjects.

As a heuristic rule, G String issues a cautionary message if the average number of (nested) raters per subject is less than 3.

GENERAL TIP

With designs where facets of generalization (raters) are nested in facet of differentiation, exercise extreme caution in situations where there are multiple observations from individual raters

DESIGN 5.1

You are collecting data from your undergraduate program to assess teacher effectiveness. You have 7 undergraduate courses, with numbers of students varying from 12 to 145. (Although this is not strictly true, assume in this example that students are different in each course. These ratings are done after random lecture, so ratings are available for varying numbers of lectures per teacher. The form has 11 items.

This is a g3xg1:g2:d study, where the facet of differentiation is Teacher, the facets of generalizations are Lecture, Student and Item. Typically, there would be one physical record for each rating with 11 ratings. To analyze in G IV, the ratings should be identified with 3 indices -- Teacher, Lecture and Student, in that sequence. **Data must be sorted in ascending order on each of these indices.**

The Screens will now look like:

			Design	
Step 3	Subj. Population	Abbrev	crossed	nested
	Teacher	t	•	

Step 4	Number of facets	3

			Design	
Step 5	Facet name	Abbrev	crossed	nested
	Lecture	L		•
	Student	S		•
	Item	I	•	

At Step 4 and 5, a "Column" box will also appear on the right. You will indicate in what column on the record the index for Teacher, Lecture and Student is located. For the Item facet, which is multiple observations on each record, you can either leave Column blank and enter number of items at Step 8, or insert "-1" and G IV will compute the number of items. If there are items within scales, on the same record, you can simply enter the number of levels of each at Step 8.

Step 6: drag “l” to “t”, “s” to “l:t”.

At Step 8, G IV will automatically generate the number of levels for t, l and s (and I if Column is -1)

The G Study output automatically generated on Screen 12 would look like:

			Generalized across		
Step 12	Facet name	Different	Random	Fixed	Levels
	Teacher	•			7
	Lecture: Teacher		•		3.2
	Student: Lecture:Teacher		•		17.9
	Item		•		11

Note that the number of levels for Lecture and Student contain fractions, which reflects the unbalanced nested design (p. 27).

Caution:

Once again, we emphasize the potential for bias in the design as a result of confounding between rater and teacher (g facet and d facet). If, for example, ratings of all lectures for each teacher were done by a single paid student in the class, then rater variance is confounded with teacher variance and coefficients are uninterpretable.

Appendix 3.

As an aid in trouble shooting, we provide here a summary of all error messages of G String IV. Each error message carries a specific error code in {}. These identify uniquely, at which location of the code an error was detected.

Error of experimental design:

```
{E 10} Facets 'Facet 1' and 'Facet 2' are confounded. You won't get valid
results!
```

Your experiment is poorly designed. You don’t have a sufficient number of nested data in your study to resolve the confounding between it and the nested facet. G_String will deliver results, but they are meaningless.

Errors of design specification:

```
{D 10} Pattern should not be empty!
```

You have to define a design pattern for each nesting level. This error is fatal.

```
{D 20} G String IV doesn't handle a subcomponent of type 'x:y:z'.

{D 21} G String IV doesn't handle a subcomponent of type 'x:y:z'.

{D 22} At present, we don't handle effects of the type 'x:y:z'.

{D 24} At present, we don't handle effects of the type 'x:y:z'.

{D 25} G_String can't handle this level of complexity at present.{x:y:z}
```

These error messages all mean the same; they have been detected at various stages of calculation. G String IV can not handle this specific design complexity. Maybe, at a later stage we will figure out how to do it and will update the program. This error is fatal.

```
{D 30} You must have exactly one facet of differentiation!

{D 31} You must have exactly one facet of differentiation!
```

Under normal circumstances, you should not get this error, since following the steps of G_String will automatically prevent it. A corrupted, re-use control file, though, could give rise to this error. This error is fatal.

```
{D 40} Error in naming facets; typically duplication.
```

Each facet requires a distinct one character abbreviation. This error is fatal.

```
{D 50} The facet of differentiation can only be nested in a facet of
stratification.
```

Under normal circumstances, you should not get this error, since following the steps of G_String will automatically prevent it. A corrupted, re-use control file, though, could give rise to this error. This error is fatal.

Errors involving the control file:

```
{C 10} Control file is not well formed!
```

In order for G String IV to re-use an existing control file, it has to be formed according to fixed rules (see page 23 of the manual for an example). Specifically, the “Comment” tag of the line specifying the facets must be terminate by a ‘%’ character, i.e. “COMMENT%” rather than “COMMENT”. When you use a control file generated by G_String_III or later, it is automatically in the correct format. This error is fatal.

Errors involving the data file

```
(F 10} Datafile 'file name' is not readable.
```

The format of the file specified is not recognized as a data file format for either G_String or urGENOVA. This error is usually due to specifying the wrong file. This error is fatal.

```
{F 20} Data don't match facet specifications.
```

The facet specification doesn’t correspond to the structure of the data file. Maybe, the asterisk was set to the wrong level (step 7). This error is fatal.

```
{F 30} Insufficient records to calculate grand mean! Empty line ‘xxx’.

{F 31} Data file does not contain sufficient data.
```

Either, you require too many datapoints, or you dropped some data from your data file. This error is fatal.

```
{F 32} Your data file is missing  ‘xxx’ values. They have been replaced with
the grand mean.

{F 33} Your data file is missing  ‘xxx’ values. They have been replaced with
the grand mean.
```

Thes messages indicate that the structure of the data file is correct, but you have empty data cells. G_String will replace missing values with the grand mean, which is ok, if only a small percentage of cells is involved, and they are more or less randomly distributed through your data file. Otherwise you have to rethink your design, in order to avoid systematic errors.

```
{F 40} Unable to convert 'String' to decimal number.
```

You may have mixed up your files, or left the column titles in the data file. G_String expects a numerical value, not characters. This error is fatal.

Internal errors:

```
{M 10} Crossed facets must have integer levels.
```

G_string expects that integer levels rather than fractional levels are specified for crossed facets. This error is fatal.

```
{M 20} Wrong averaging type 'X'!
```

This error should not normally occur. G_String selects the appropriate averaging types according to rules listed in the manual and in Brennan. Theoretically, there could be internal errors which would call up an incorrect averaging type. This error is fatal.

Errors transmitted from urGENOVA:

```
{U 10} urGENOVA error: ‘message’
```

If urGENOVA fails for any reason, it emits an error message which is displayed by G_String. These errors are usually fatal.

Bibliography

Brennan, R.L. Generalizability Theory. New York, Springer, 2003.

The "bible" of G theory.This manual contains a number of references to Brennan.

Streiner, D.L. , Norman, G. R. Health Measurment Scales: A practical Guide to their Development and Use. (4th ed.) Oxford, OUP, 2008.

An approachable treatment of G theory, although it differs somewhat from the formalism used here.

Shavelson, P, Webb, N. Generalizability Theory: A Primer. Thousand Oaks CA, SAGE, 1991.

Another classic book, intermediate in complexity between Streiner and Brennan.