

G_String_VI

User Manual

by Ralph F. Bloch

&

Geoff Norman

McMaster University

Hamilton, September 2021

© 2021 by Ralph Bloch & Geoff Norman, Hamilton, Ontario, Canada

Table of Contents

Introduction.....	3
Background.....	3
Download and Installation.....	5
Author's Reflections.....	5
Illustrative Example.....	5
User Interface: Analysis.....	6
User Interface: Synthetic Data Sets.....	15
Empirical vs Synthetic Data Set.....	18
Data File Format.....	19
Auxiliaries.....	19
Troubleshooting.....	20
Limitations.....	21
Bibliography.....	21
Index.....	22

Introduction

G_String_VI is a software program with a graphical user interface to code data entry and compute generalizability coefficients based on variance component estimates generated by urGENOVA. It can also create synthetic data sets based on user-provided designs and variance coefficients. It was designed and coded by Ralph Bloch in cooperation with Geoff Norman as part of a project originally commissioned by the Medical Council of Canada and subsequently developed further.

G_String_VI is written in Java 8 on the Linux platform. It can run on Macintosh or PC computers under Windows, Linux or iOS operating systems.

urGENOVA was written by R.L. Brennan at the University of Iowa. It is included in the installation package for G_String_VI, and operates in the background. It is a traditional command line program written in ANSI C; users needed to 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.

Background

Generalizability Theory is complex, and readers interested in a deeper understanding should consult the resources listed in the Bibliography. We will here only explain some essential terms superficially. For novices in G theory the YouTube video <https://www.youtube.com/watch?v=uLMrHCTP65Q> provides a good eye opener, and the AMEE Guide No. 68 may be particularly helpful.

The purpose of a generalizability study is to estimate the reliability of a specific behavioral measurement with the intent of generalizing its findings, and to identify potential sources of measurement error.

Behavioral measurements require defined methods to elicit and collect responses from subjects of interest to a set of stimuli such as statements on a questionnaire, observations of performance, or almost any psychometric measurement. The responses will typically vary widely, depending on the individual subject, as well as other factors such as the rater, the specific question, or other sources of error. G theory is a statistical strategy to identify and quantify these various sources of error.

The variability can be expressed mathematically as a variance, which generalizability analysis subdivides into a variance component attributable to the subjects and variance components associated with various errors associated with the measurement process.

The variance component attributed to the subjects is designated $\sigma^2(\tau)$, while the variance component resulting from all the other aspects of the measurement is referred to as $\sigma^2(\delta)$ or $\sigma^2(\Delta)$ respectively. The δ -component is called relative or norm referenced, the Δ -component is called absolute or criteria referenced.

The ratio between the subject variance component and the relative total variance is called the 'generalizability coefficient', while the ratio between the subject variance component and the absolute total variance is called the 'index of dependability':

$$E\rho^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)} \text{ Generalizability Coefficient}$$

$$\Phi = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\Delta)} \text{ Index of Dependability}$$

The individual aspects of the measurement design are customarily referred to as 'facets'. The facet corresponding to the subjects under investigation is referred to as 'facet of differentiation', the other aspects are called 'facets of generalization'. If the facet of differentiation is nested in some other facet (for example, students within classrooms) the nesting facet (classroom) is called a 'facet of stratification'.

$\sigma^2(\tau)$, $\sigma^2(\delta)$, $\sigma^2(\Delta)$, $E\rho^2$, and Φ are calculated by G_String_VI from the variance components corresponding to the various facets and their appropriate combinations, provided automatically by Brennan's urGenova contained in G_String_VI. The actual calculations are detailed in the printout.

G theory is based on the linear regression model of statistics. The analysis solves for the parameter variances in a data set collected under a specific experimental design.

G_String_VI also allows the researcher to construct synthetic data sets that would result from a specified design, assuming postulated variances. Being able to create synthetic data sets is not only useful in teaching G Analysis, it also allows for experimentation, thus promoting a deeper understanding of G Theory.

Users should be warned not to misrepresent synthetic data sets as empirical data. G_String_VI synthetic data sets carry a pseudo-random signature that allows detection of their synthetic nature by statistical methods.

Download and Installation

G_String_VI and its manual are available for download from McMaster University at:

https://healthsci.mcmaster.ca/merit/research/g_string_vi

Before the program can be used, it needs to be properly installed depending on the operating system of the machine it has to run on. Detailed instructions are available on the web site.

Author's Reflections

G_String started about 10 years ago as a humble data entry tool for urGenova. Over the years it has gradually become more stable and capable, largely due to helpful feedback from the user community. It also required frequent adaptations to changes in operating systems and unusual behaviors of the same operating system in different locales.

There is no reason to assume that this need is likely to change over the next years. But the author's age makes it unlikely that he will be around for very much longer.

Unless some users unite to take initiative and accept the baton, G_String will gradually fizzle – which seems a pity.

I am willing to assist the take over by fleshing out the source code documentation, and supporting whoever will take on the stewardship. It needs a group of enthusiasts willing to collaborate, and should include statistical, programming and management skills – not necessarily all within one person.

If you are interested, please write to me at ralph@papaworx.com.

Illustrative Example

In the following we run G_String_VI through an actual study, courtesy of Bao Quoc Le MD. MSc. from The Center for Advanced Training in Clinical Simulation (ATCS), University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam.

A class of 99 medical students was divided into 12 groups. Six groups each were tested over two days. On each day 2 instructors rated the students of three groups each. Each student had to complete 40 items, each rated on a pass/fail scale.

This example illustrates many important features of G_String_VI.

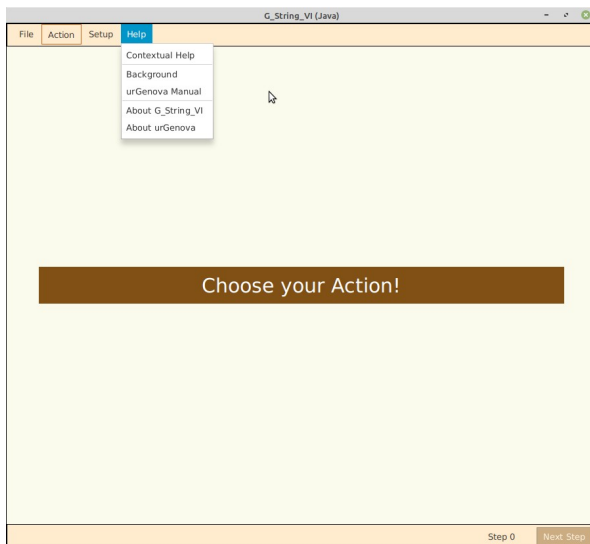
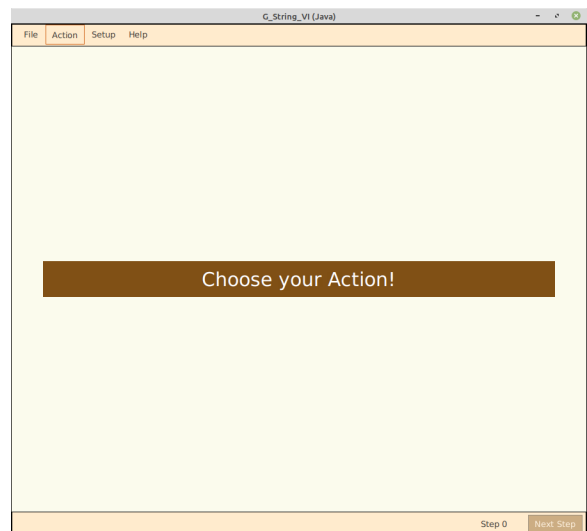
User Interface: Analysis

The basic user interface of G_String_VI stays constant during use.

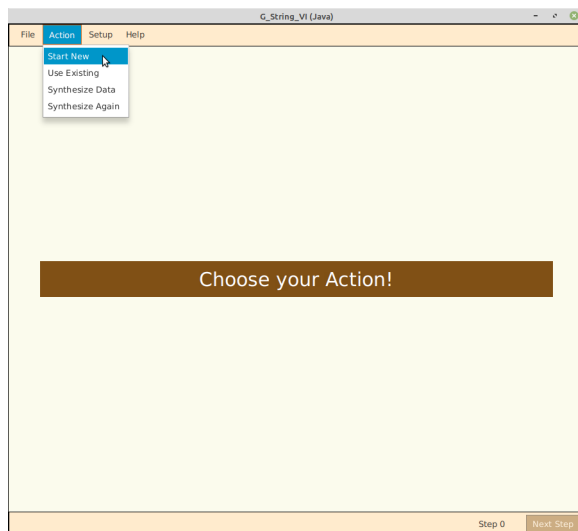
Use the menu bar at the top of the G_String application window to select the desired operation.

All information relevant to the actual operations are shown in the central panel. Throughout G analysis or data synthesis the user is guided step by step. Context-specific help is available via the Help option on the menu bar.

The bottom bar enables users to step through the operation.

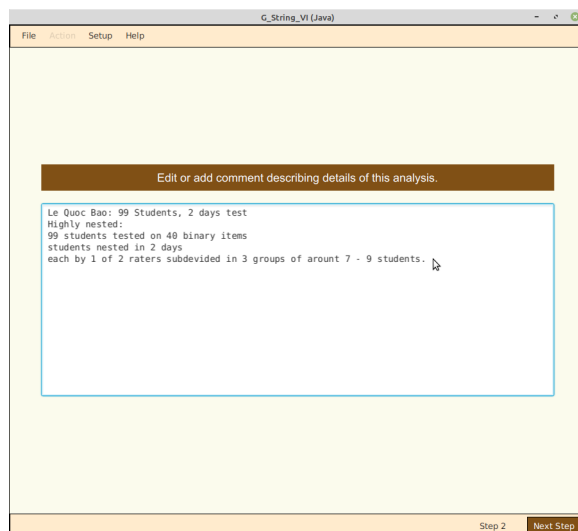
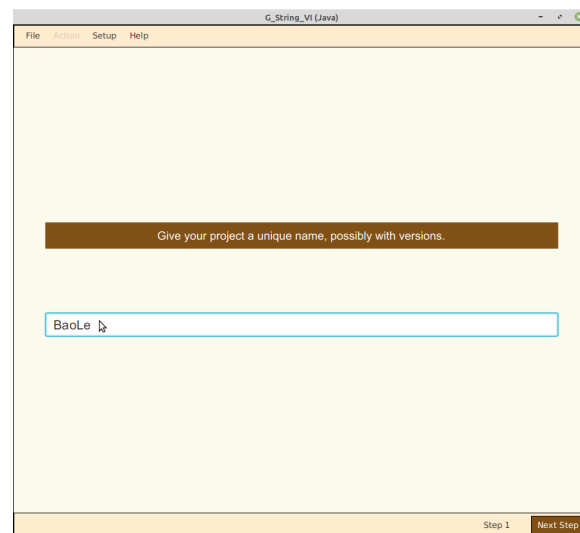


Besides providing access to context specific help, the Help menu also leads to background information, the manual for urGenova, as well as some administrative data relating to G_String_VI and urGENOVA.



There are two ways to start a G Analysis. Usually one chooses the 'Start new' from the 'Action' menu option. In this case all data and control language has to be entered by the user. The control language is entered in response to queries from G_String. Data are typically entered from a spreadsheet. If, however, users have a project in progress, they can select 'Use Existing'. Thus users can select an existing control file, and step through already completed forms, changing and adding information as required.

Step 1, one enters a suitable title for the project as an aid in identifying the data later. For illustration purposes we are using actual data provided by Dr. Bao Quoc Le with his kind permission.



Step 2 allows the addition of descriptive text, which contributes context information to assist the interpretation of results.

Note: None of the lines in this descriptive text field may begin with one of the reserved keywords: GSTUDY, COMMENT, OPTIONS, EFFECT, FORMAT, ANCHORS, or VARIANCES.

Specify subject and number of facets.

Subject	Label	Nesting
		crossed nested
Student	s	<input type="radio"/> <input checked="" type="radio"/>

Select number of facets (excl. subject): 4

Step 3 [Next Step](#)

In **step 3** the designation for the facet of differentiation - 'student' - has to be entered, together with a unique, identifying lower case label - 's'.

A toggle identifies whether the subject appears crossed or nested in the particular design.

And finally, the number spinner allows the user to select how many facets of generalization are involved.

In **step 4**, the facets of generalization are entered next, including their designation, a label and whether they are crossed or nested.

Now specify each of the remaining facets.

Facets	Label	Nesting
		crossed nested
Day	d	<input checked="" type="radio"/> <input type="radio"/>
Rater	r	<input type="radio"/> <input checked="" type="radio"/>
Group	g	<input type="radio"/> <input checked="" type="radio"/>
Item	i	<input checked="" type="radio"/> <input type="radio"/>

Step 4 [Next Step](#)

Set order of facets and identify the main effect indicated by change of record (Star).

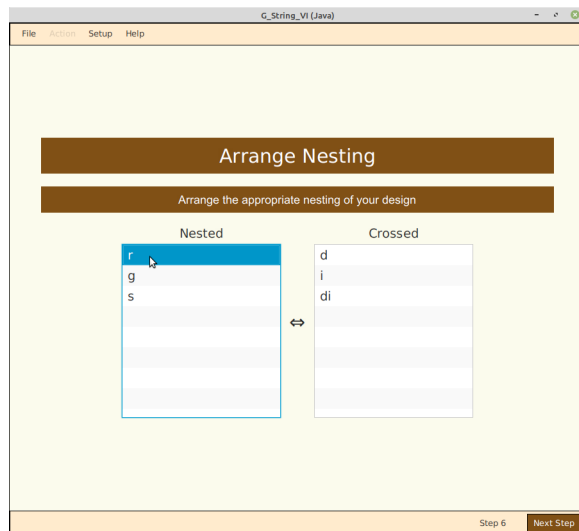
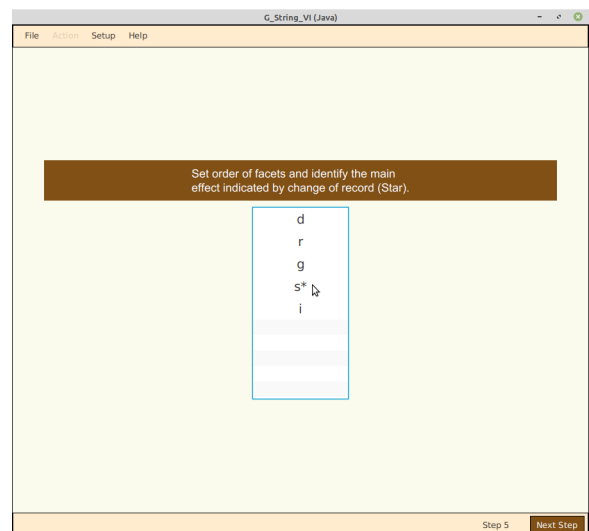
s*	
d	
r	
g	
i	

Step 5 [Next Step](#)

In **step 5** the facets have to be arranged in the same order as they appear in the data file.

In this case, since 'Student' is nested in 'Group', the letter 's' must be grabbed and moved down to 'g'. The asterisk indicating 'change of line' stays with 's'; otherwise one needs to click on the label where the line changes.

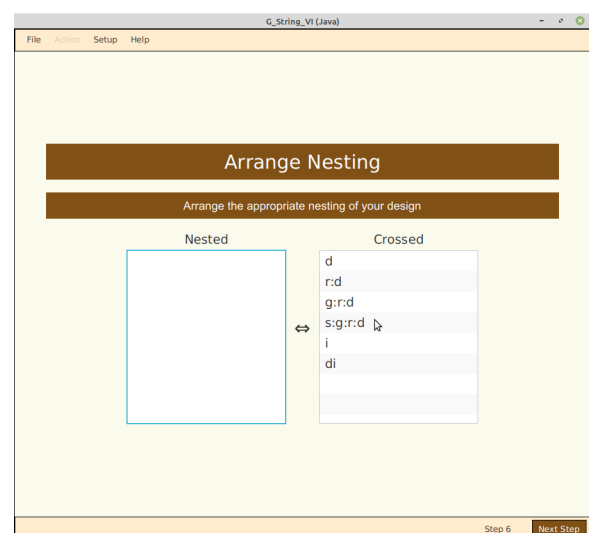
Now that the order of facets is correct, corresponding to the order of the data file, and each student's scores appear on a separate line.

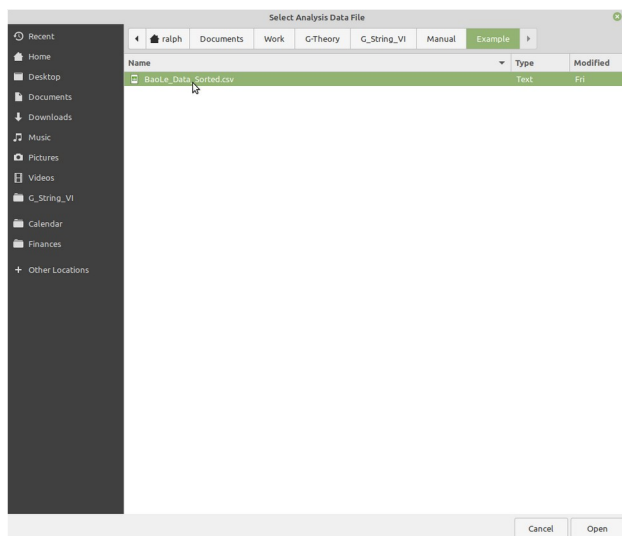


Step 6 establishes correct nesting by grabbing the corresponding label on the left, and dragging it over to the right, dropping it on the facet it is nested in.

This process has to be repeated until the desired nesting order has been set.

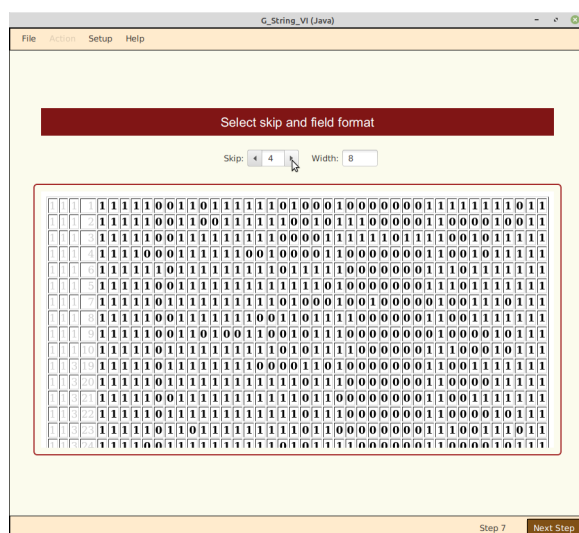
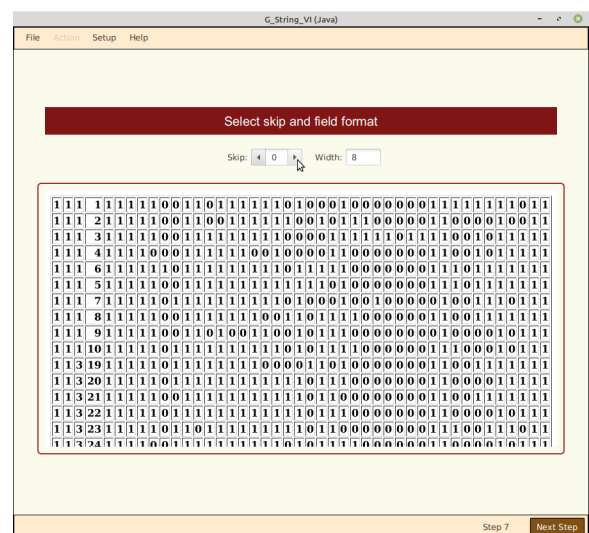
Once, no more orphan nested facets are left on the left, one can proceed to the next step.





Step 7 opens the actual data file. The correct formatting of the data file is described in section ‘Data File Format’.

In this case the first four fields on each line indicate the order index for each facet. They are not used explicitly for the urGenova calculations, therefore we set the spinner to skip the first four fields. Field width can be left at ‘8’.



Now that the relevant data have been selected, one can proceed to the next step.

In **step 8** one specifies sample size for all five facets, one after the other.

In this example, the study extends over two days ('d').

The screenshot shows the 'Set sample sizes for facet 'd'.' dialog box. It contains a table with the following structure:

Sample Sizes	
1	2

The 'Next Step' button is visible at the bottom right.

The screenshot shows the 'Set sample sizes for facet 'r' in d.' dialog box. It contains a table with the following structure:

Sample Sizes	
2	2

The 'Next Step' button is visible at the bottom right.

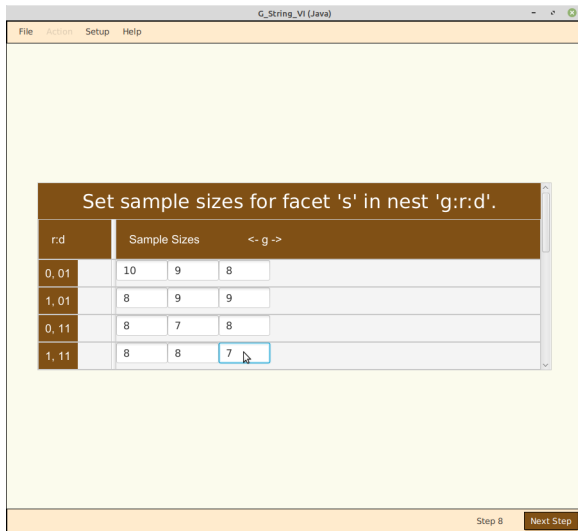
For each of the two days there are two raters ('r').

Each rater sees three groups of students ('g') on his day.

The screenshot shows the 'Set sample sizes for facet 'g' in nest 'r:d'.' dialog box. It contains a table with the following structure:

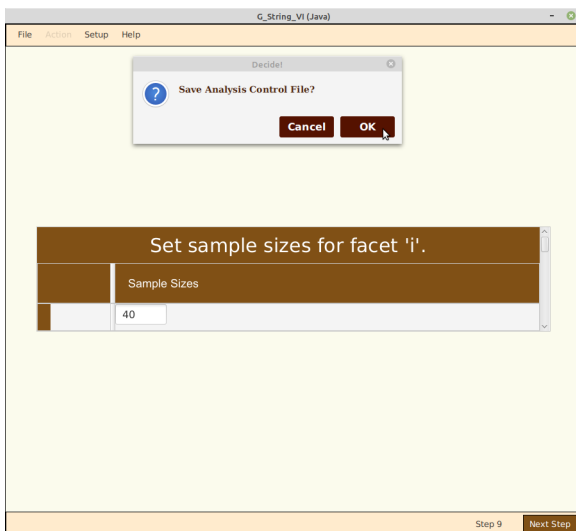
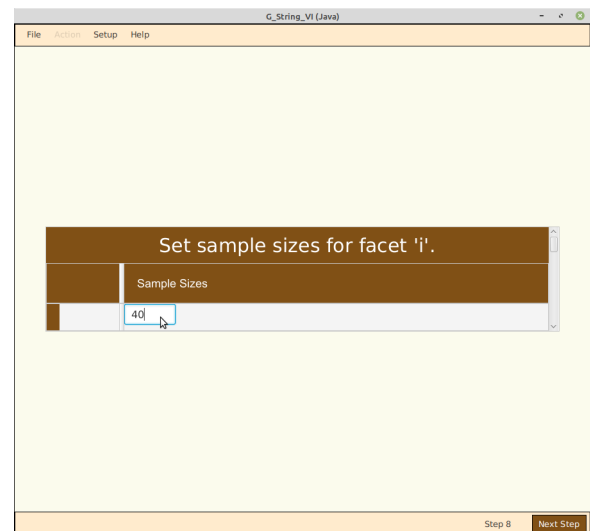
d	Sample Sizes	
0	3	3
1	3	3

The 'Next Step' button is visible at the bottom right.



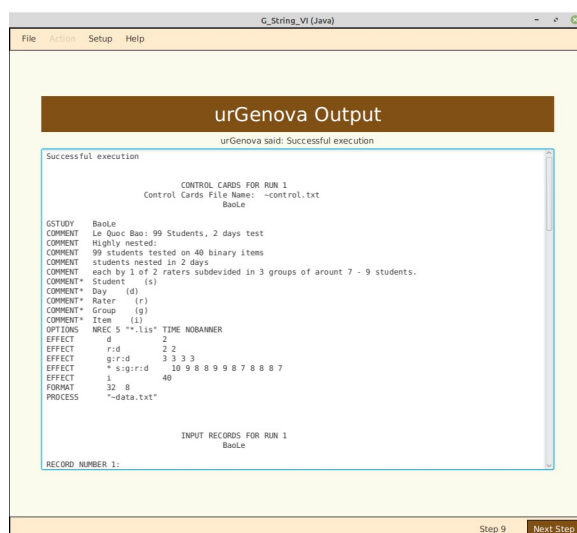
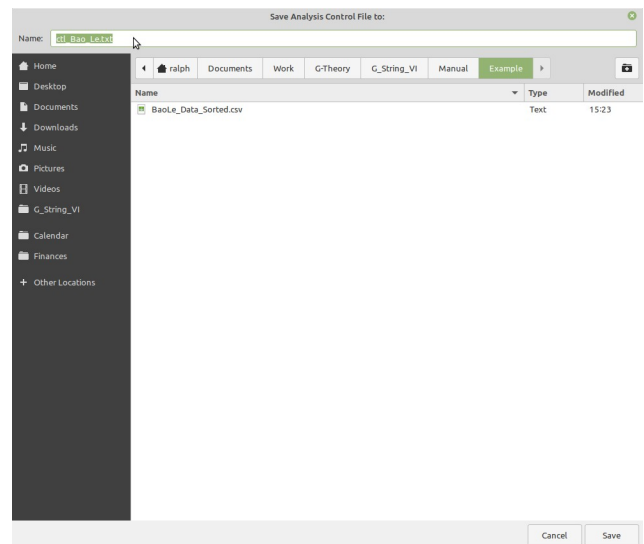
But not every group contains the same number of students. If we add them all up, there are 99 students ('s').

Every student, finally, has to complete all 40 items ('i') in the assigned time. 'Items' is crossed with the rest.



Now we have the option of saving the completed Analysis Control file, to be reused or modified at a later stage. This can save a lot of time in the future.

Just give it a suitable file name, and place it in the directory of choice.

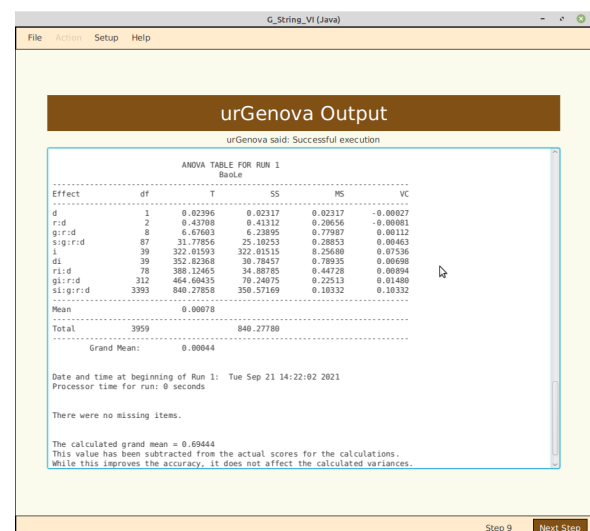


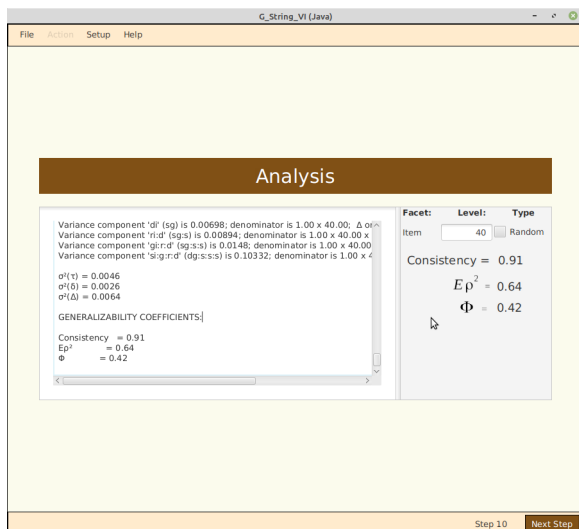
In **step 9** we are ready to interpret the results.

The screen confirms that urGenova could successfully process the data entered. The scrollable window shows the standard urGenova output.

It repeats the control input data and shows calculated mean values.

As we scroll down further the output shows a table of the calculated variance components. From here on G_String_VI takes over again to calculate the generalizability coefficients.

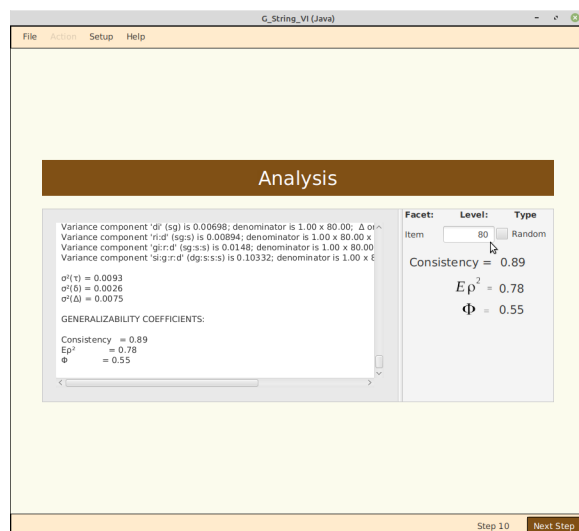




Finally **step 10**: urGenova arrives at the variance components by statistical estimations. Some of the components can come out negative, which is formally nonsensical for squared quantities. The convention is to set these components to zero for further calculations. The 'Consistency' measure indicates the relative sum of all positive components. If 'Consistency' falls significantly below 1.0, results should be taken with caution.

The calculations of the generalizability coefficients are detailed in the output. Test results can be considered reliable, if $E\rho^2$ reaches at least 0.8. For high impact exams, a threshold value of 0.9 is desirable.

To the right of the scrolling window is a calculator for D-Studies. Let's see what happens, if we double the number of items to 80: the generalizability coefficient rises from 0.64 to 0.78.



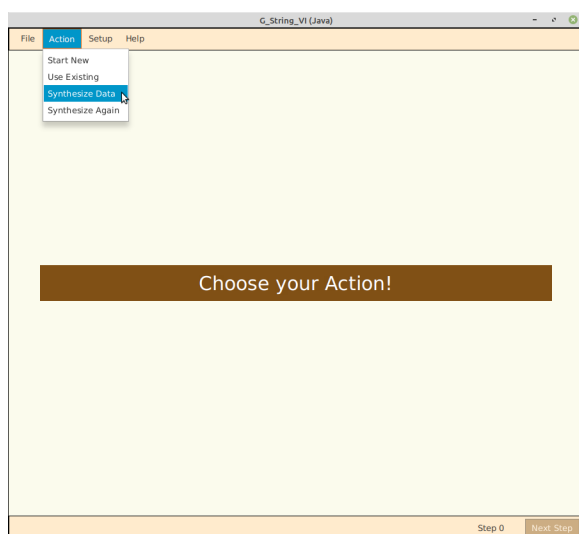
Finally it is wise to save the results of the analysis in a text file (File → Save Results).

User Interface: Synthetic Data Sets

New in Version G_String_VI is the ability to construct synthetic data sets based on individual research designs with realistic variance components, using a similar graphical user interface as that provided for the Analysis.

Such synthetic data sets can be fed directly to Analysis. They provide at least two benefits:

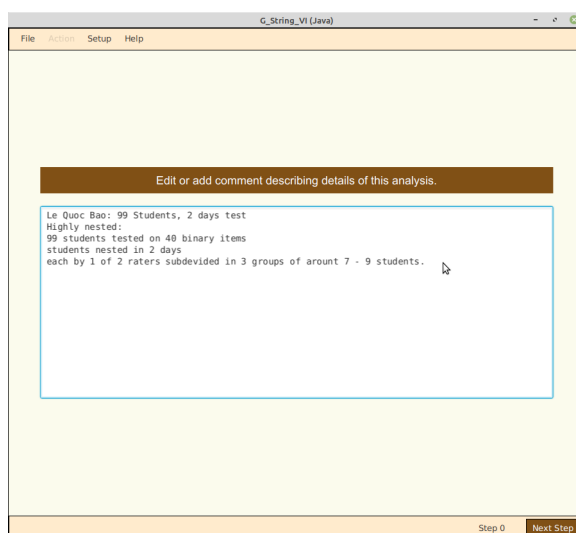
- 1 Instructors can painlessly construct data sets for students to practice G analysis skills.
- 2 Researchers can try out various research architectures, based on identical subject variances and experimental designs in order to optimize reliability of their potential findings.

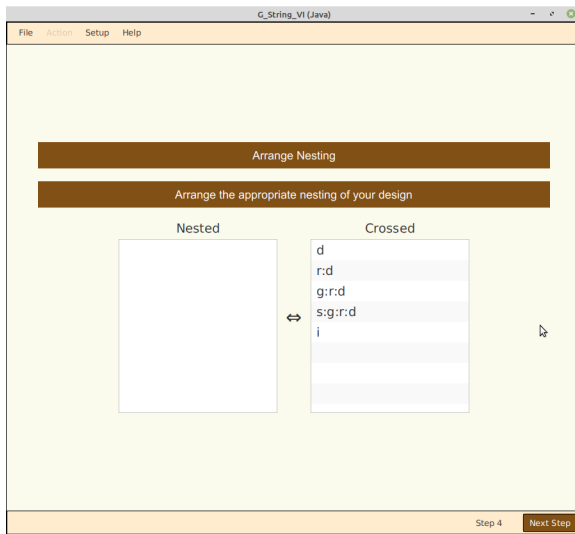


Choose Action → 'Synthesize Data' in the top menu bar to start the process. In analogy to Analysis, Synthesis can also be invoked based on a previously constructed script by selecting 'Synthesize Again'.

As in Analysis, the user starts by entering a short verbal description on purpose, context, and design.

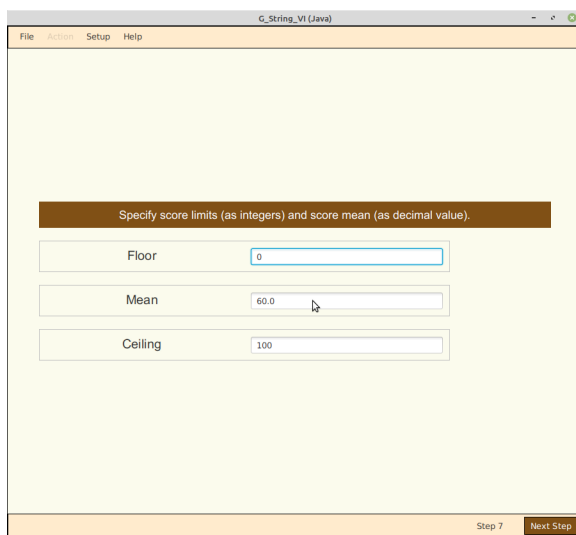
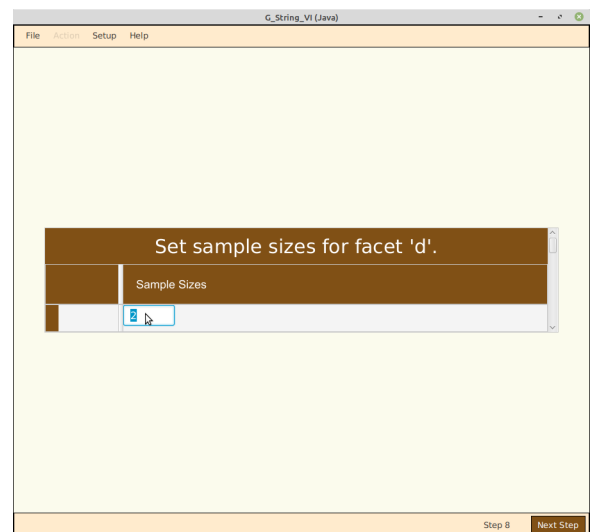
This can often help later in clarifying some issues.





The next four steps are identical to the ones in Analysis, and therefore will not be described here.

Since the purpose of 'Synthesis' is to generate a data set file, the next step requires us to specify sample sizes for all facets identically to the Analysis. We will skip here the steps after 'Day' in our chosen example.



After we have finished specifying all the sample sizes, we have to specify the desired score range in integers from the minimum (floor) to maximum value (ceiling), and the expected mean value as decimal value in between.

In this example we simply expand the score scale from the Analysis by a factor 10.

This is probably the most complex entry form. From the design data you have entered, G_String_VI has determined all possible facet combinations and interactions.

In this example there are nine variance components to be specified. Since variances are quadratic forms, we have to multiply the original variance components from the results of Analysis by a factor 100.

Configuration	Levels	Variance
d	2	2.0
r:d	4	5.0
g:r:d	12	3.0
s:g:r:d	99	10.0
i	40	3.0
i x d	80	6.0
i x r:d	160	5.0
i x g:r:d	480	5.0
i x s:g:r:d	3960	10.0

Configuration	Levels	Variance
d	2	0.0
r:d	4	0.0
g:r:d	12	0.1
s:g:r:d	99	0.4
i	40	7.0
i x d	80	0.7
i x r:d	160	0.9
i x g:r:d	480	1.5
i x s:g:r:d	3960	10.0

You now have the option of saving the Synthesis control file for reuse in various 'what if' experiments. You will also have to save the newly generated, synthetic data set.

Don't forget to save the synthetic data.

Next G_String_VI informs you how the Synthesis went. Remember that this is a Monte-Carlo type operation. The Normal Distribution goes from $-\infty$ to $+\infty$, so some results will be outside the allowed range and have to be constrained. As a consequence the resulting variance components will be somewhat smaller than the input values.

Don't forget to save the synthetic data.

Configuration	Levels	Variance
d	2	2.0
r:d	4	5.0
g:r:d	12	3.0
s:g:r:d	99	10.0
i	40	3.0
i x d	80	6.0
i x r:d	160	5.0
i x g:r:d	480	5.0
i x s:g:r:d	3960	10.0

Empirical vs Synthetic Data Set

Let us now compare the results from analyzing the original, empirical binary data set with those of the synthetic, expanded score range data set:

Parameter	empirical	synthetic, expanded
Score, floor	0	0
Score, ceiling	1	10
Score, mean	0.694	6.37
Variance Components:		
d	-0.00027	-0.04190
r:d	-0.00081	0.05636
g:r:d	0.00112	-0.00393
s:g:r:d	0.00463	0.27099
i	0.07563	3.52117
di	0.00698	0.80609
ri:d	0.00894	0.30097
gi:r:d	0.01480	0.64473
si:g:r:d	0.10332	6.30737
$\sigma^2(\tau)$	0.0046	0.2710
$\sigma^2(\delta)$	0.0026	0.1577
$\sigma^2(\delta)$	0.0064	0.3459
Generalizability		
Coefficients		
Consistency	0.91	0.93
$E\rho^2$	0.64	0.63
Φ	0.42	0.44

Remember that, while only 12 (3 + 9) fixed parameters were specified for the synthesis, G_String_VI contributed 4837 random values to generate the synthetic data set. It is, therefore, not surprising that intermediate results of the synthesis diverge significantly from those of the original, empirical values. But the final generalizability coefficients agree surprisingly well, considering that we have even expanded the scoring range.

Data File Format

G_String_VI accepts data in comma (csv), or tab (tsv) separated format. Scores have to be integer values. Leading index columns are permitted, but have to be screened out in step 6. No leading title or column header lines are permitted.

The best way to prepare data set files is in a spreadsheet, and then export the properly formatted data in either .csv or .tsv format.

Auxiliaries

From anywhere in the program one can start over by selecting **File → Start Over**.

From anywhere in the program one can exit by selecting **File → Exit**.

To change Diagnostic Log levels select **Setup → Log Level**.

The graphical appearance and functionality is largely controlled by a set of preference parameters. These can be viewed, changed, saved to an external file, or imported from an external file by selecting **Setup → Preferences**. Casual users are advised to leave the preferences unchanged. Users who want to experiment with preferences should first save the original state in an external file, so they can always restore G_String_VI to its default configuration..

While stepping through Analysis or Synthesis users are supported by context-specific help screens callable by selecting **Help → Contextual Help**.

Although contained in the manual, background information is available by selecting **Help → Background**.

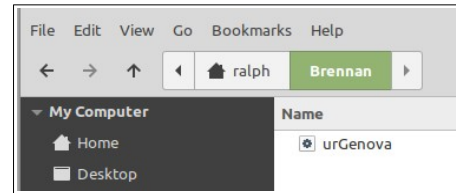
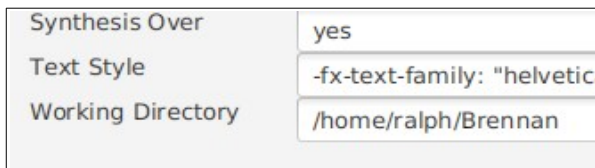
Users who wish the full functionality of urGenova can select **Help → urGenova Manual**, and change preference options accordingly.

To find the version and build date of G_String_VI select **Help → About G_String_VI**.

To inspect the copyright constraints Robert Brennan has stipulated, select **Help → About urGenova**.

Troubleshooting

1. If new users fail to properly install G_String_VI according to their specific operating system, and perform **Setup**, as explained on the download site, the program will not start. Make sure that you followed the installation instructions exactly.
2. Clicking Setup → **Preferences** → **Change Preferences** should show a named directory in the 'Working Directory' window. The actual directory must contain the program 'urGenova' ('urGenova.exe' for Windows).



3. If it all fails, send a detailed email to ralph@papaworx.com.

Additional resources and tutorials are available at:

<https://papaworx.com/G-School>

Limitations

G_String_VI has no limitation in number of permitted crossed facets or sample size. Nesting has been tested up to level 5 deep, but we are not aware of structural limitations beyond.

Bibliography

Brennan, R.L. Generalizability Theory. New York, Springer, 2003. (*The “bible” of G theory. This manual contains a number of references to Brennan*).

Cronbach, L.J., Gleser, G.C., Nanda, H., & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley.

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

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

Bloch R., Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. Med Teach. 2012;34(11):960-92

Wikipedia: Generalizability Theory.

Bloch R. G-Theory, an Introduction. YouTube: <https://www.youtube.com/watch?v=uLMrHCTP65Q>

Alphabetical Index

analysis.....	4, 6, 14, 15
Analysis.....	2, 4, 6, 7, 12, 15, 16, 17, 19
Analysis, and therefore will not be described here.....	16
ANSI C.....	3
Brennan.....	3, 4, 19, 21
control file.....	3, 7, 17
Control file.....	12
facet.....	21
facet of differentiation.....	4, 8
facet of stratification.....	4
facets.....	4, 8, 9, 11, 16, 21
facets of generalization.....	4, 8
facets of generalization'.....	4
G theory.....	3, 4, 21
G Theory.....	4
generalizability coefficient.....	3, 4, 13, 14, 18
Generalizability Coefficient.....	4
Generalizability theory.....	21
Generalizability Theory.....	3, 21
help.....	19
Help.....	19
index.....	10, 19
Index.....	4
index of dependability.....	4
Index of Dependability.....	4
linear regression model.....	4
manual.....	19, 21
Manual.....	19
measurement.....	3, 4, 21
nested.....	4, 8, 9
reliability.....	3, 15
sample size.....	11, 16, 21
spreadsheet.....	7, 19
synthesis.....	6, 18
Synthesis.....	15, 16, 17, 19
urGenova.....	19
variance component.....	3, 4, 13, 14, 15, 17
Variance Component.....	18
variance components.....	3, 4, 13, 14, 15, 17
Variance Components.....	18