Journal of Medical Statistics and Informatics
ISSN 2053-7662 | Volume 8 | Article 8

HOAJ
Herbert Open Access Journals

Methodology Open Access

Generalizability Theory:Demonstrating the Process and its Utility with EEG Measurements

Adrienne Kline[1,2*], Theresa Kline[3], Daniel Pittman[4], Bradley Goodyear[4] and Janet Ronsky[1,5]

*Correspondence: askline1@gmail.com

[1]Department of Biomedical Engineering, University of Calgary, Calgary, AB, Canada.
[2]Cumming School of Medicine, University of Calgary, Calgary, AB, Canada.
[3]Department of Psychology, University of Calgary, Calgary, AB, Canada.
[4]Department of Radiology and Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada.
[5]Department of Mechanical and Manufacturing Engineering, University of Calgary, Calgary, AB, Canada.

Abstract

Background: The purpose of this study is to demonstrate the utility of generalizability theory in assessing the reliability of EEG data. Generalizability theory and its relevance for measurement are described, followed by the steps to consider and decisions to be made when conducting a generalizability analysis.

Methods: Using an actual data set, how to conduct generalizability and decision analyses using IBM ® SPSS ® are outlined. Specifically, the beta frequency data collected from the C1 electrode from 16 participants across 60 trials and 2 time periods as they moved their right leg in a stepping motion are used for the demonstration. Data analysis decisions such as number and type of facet, relative versus absolute G-coefficients, follow-up decision-study information, data set-up and analytic commands are all described in detail.

Results: Outputs from the analyses are discussed in terms of their meaning and implications. These include the reliability of each facet, variance accounted for and best next steps for improving data reliability.

Conclusion: Advantages of using the generalizability model are discussed as well as suggestions of various generalizability programs available for use.

Keywords: Generalizability, reliability, EEG

Introduction

The use of generalizability theory in the pursuit of assessing reliability of measures is becoming more prevalent in a number of research areas including center of pressure data collected with a force platform [1], ankle-complex laxity measures [2], electromyogram power measures [3], and EEG interpretations [4]. This approach allows for a more fulsome understanding of the sources of variance than does the simpler test-retest approach often used in EEG research [5-10]. Despite its growing use, a comprehensive step-by-step "how to" demonstration of generalizability analysis using available computer programs has yet to be provided. This study aims to fill that gap and in doing so, encourage its use.

Generalizability analysis (G-study) estimates the amount of variance that can be attributed to various facets in a dataset. Facets in a study could be participants, trials, repeated time measures, coders, etc. G-study analysis follows a relatively straightforward ANOVA approach, which ascribes the amount of variance in the data set to source, including all interactions and error terms. Via a series of equations using information in the ANOVA source table, variance components for each facet are generated, and based on these components the overall generalizability of the design is computed. This overall generalizability coefficient (G coefficient) can be interpreted as a reliability coefficient, similar to that of the intra-class correlation coefficient (ICC). The variance components themselves are of interest because it allows the researcher to determine from which facet, or facets, the most variance in the data arises.

For example, assume we have measures of beta EEG frequency bands for 16 different subjects, across 60 trials at 2 time periods separated by 2 weeks. Further assume that the G-coefficient is calculated to be 0.90. One important interpretation is that the researcher can claim that these observed measures provide a good estimate of what would be expected to be obtained by another set of EEG beta frequency measures taken under similar circumstances (i.e., reliability). If most of the variability in the EEG beta measures is due to subject differences, then this may be expected, as we would usually anticipate between-subject variability in EEG beta measures across different people. However, if much of the variability in the data set is due to time-periods then this might be of concern, particularly if the researcher planned to average the data across trials and time periods. This would mean that there is systematic variance across the two time intervals such that they provide markedly different measurements (e.g., a training effect, machine calibration, etc.). The methodology of the study may then be modified to address this issue.

The G-study is typically followed up with a decision-study (D-study). Using the variance components generated in the G-study, a series of "what if" questions are answered with a D-study. For example, for the study described above, it is possible to ask: What would the G-coefficient be if we used 20 trials instead of 60 trials? What would the G-coefficient be if we used 3 time points for data collection instead of 2 trials? and so on.

The history, rationale and equations associated with generalizability theory have not been described because they are presented in excellent formats elsewhere [11-14]. Instead, the purpose of this study is to demonstrate how generalizability theory can be applied to a set of data using the IBM ® SPSS ® statistical package. Included in the presentation is the data set, the steps involved in the analysis, and interpreting the outputs. A caveat is that the data in this example are fully crossed (i.e., are non-nested) and completely balanced (i.e., no data points are missing). The equations and interpretations become much more complex if the data do not meet these specifications and are beyond the scope of this demonstration. However, once the basics of the design presented here are mastered, the logic can be extended to more complex models.

Materials and Methods

Ethics statement

Participants took part individually in the study after signing a written informed consent document. The study was approved by the University of Calgary Conjoint Health Research Ethics Board (ID: REB15-1473).

Sample

The sample size was 16. All participants were males between the ages of 19 and 31 (mean age=24.7, SD=3.2), had no history of knee or hip injury, were in good health, and had no reported neurological deficits.

Procedure

The methodology used to obtain the data will not be described in detail, as that is not the purpose of this demonstration. Participants were fitted with a 64 electrode NEUROSCAN® EEG cap where location of the electrodes followed the 10-20 electrode international placement system. Participants lay in a supine position and performed alternating left and right stepping motions at a pace of 50 steps/minute while observing a computer-generated image of a human walking. They repeated this stepping motion 60 times for each leg. Their heads were immobilized using compressible foam cushions and their feet were strapped to pedals that allowed for flexion and extension at the hip, knee and ankle joints, sliding along a near frictionless track for each foot. A pulley system tethered to each pedal applied weight to simulate the effects of gravity on the lower body [15]. Participants completed the data collection at time 1 and then returned for the same process two weeks later at time 2.

Data

The data used for purposes of this study were collected from the C1 electrode and were part of a larger study [16]. They represent only the data associated with the beta frequency for right leg flexion/extension collected at time 1 and time 2. Data were sampled at 1000Hz and analyzed using customized software developed in *Matlab 2017b* (Mathworks™, Natick, Massachusetts, USA). Data were DC offset corrected, bandpass filtered between 5 and 55 Hz, and referenced to the global average of all 64 channels. Epochs of data were generated based on the timing of the visual stimuli onset/offset for each individual 'step'. Table 1 shows the first three and last three columns of C1 beta frequency EEG data for subjects 1, 2, 15, and 16. The complete data set for this electrode and frequency band is provided in the supplementary materials (Kline_C1_right_executed_stepping.pdf).

Analysis and Results

Step 1: The Facets and Data Set-up

In this data there were three facets of interest: participants (*p*, n=16), trials (*t*, n=60) and time periods (moments) (*m*, n=2). Prior to analyzing the data, several questions regarding the data and the use of the findings need to be answered. The first is, which is the facet of differentiation and which are the facets of generalization [11]? Doing so is useful because it assists in setting up the data in a manner conducive to answering the most pressing reliability question.

For example, assume the question of a data set is: "What is the generalizability of the measures for participants collapsed across trials and time periods?" The focus is on the participants' measures and how they generalize. When this is the case the data are set up so that one "row" of data is assigned to each participant (p). The facets of generalizability, then, are the trials (t) and time periods (m). These, then, are the "columns" and are set up so that the trials (fastest moving facet) are

Table 1. Truncated data set where m=moments (time points) and t=trials.

Participant	m1_t1	m1_t2	m1_t3	…	m2_t58	m2_t59	m2_t 60
1	-945	-1012	-747	…	-2041	-2510	-2924
2	-1481	-2298	-2698	…	-1242	-1823	-822
…	…	…	…	…	…	…	…
15	-886	-2177	-1860	…	-1195	-961	-1065
16	171	-579	-387	…	-197	-429	64

oriented from left to right across the columns (as in Table 1). Let's change the question. Assume there is an expected training effect from time 1 to time 2. Variability in the measures across time periods would present itself in a high variance component allocated to the effect of time period (m). In this case the data would be set up so that time period becomes the facet of differentiation (each time period would be a row). The columns would represent the participants (p) and trials, (t) with trials changing fastest moving across the columns. The calculated generalizability coefficient would answer the question "What is the generalizability of the measures for time periods collapsed across trials and participants?"

Another important facet question is whether each of the facets is fixed or random. A fixed facet is one where the conditions of the design are the only ones in which the researcher is interested, whereas a random effect is one where the conditions of the design are assumed to be a random sample from a larger population to which the researcher wishes to generalize the findings. The facet of differentiation is always considered to be a random effect, as the purpose of a reliability analysis is to generalize to a population of some sort. In the first example above, participants are considered to be a random sample from a larger population. In the second example, the time periods are considered to be a random sample of time periods from the universe of time periods.

Not surprisingly, generalizability coefficients based on fixed effects are higher than those based on random effects. However, it is unusual to have fixed effects in research, given its mandate to generalize findings, and so in this demonstration is will be assumed all the effects (p, t, and m) are random.

Step 2: The Generalizability Coefficient

There are two variations on the G-coefficient. It is important to decide which one is appropriate, as both will be calculated. The first is the *relative G-coefficient*. This is analogous to the traditional reliability coefficient in classical test theory and reflects how well the values on the facet of differentiation maintain their *relative* rank order across columns of generalizability facets. This is also sometimes referred to as a "consistency index". This level of reliability may be appropriate, if, for example, one is interested in taking an average of a set of ratings. Assume for a moment that there are 5 research articles rated by two different raters on a scale of 1 to 10. The data are shown in Table 2. It can be observed that the raters are consistent in terms of the relative ranking of the articles. When these ratings are averaged across the raters, the ratings of the articles' rank ordering remains the same. For many studies that use average scores to relate with another variable of interest, it suffices to have relative agreement between the raters.

Table 2. Sample data set of article ratings.

Article	Rater 1	Rater 2	Average Rating
1	1	6	3.5
2	2	7	4.5
3	3	8	5.5
4	4	9	6.5
5	5	10	7.5

On the other hand, there may be a reason that an *absolute* level of agreement is needed. Here in addition to consistency in rank order across generalizability facets, there also needs to be agreement on the elevation levels of the measures. This would be assessed with the *absolute G-coefficient* - sometimes referred to as an "agreement index", "index of dependability", or "phi coefficient". This is important when the numerical value of the measurement is used to make decisions, such as a fixed cutoff score for entry into a professional program. In such a situation, training raters to be both consistent and in agreement would play an important role in the perceived fairness of a decision process. Both the relative and absolute G coefficients will be produced in the analysis. Which one to interpret depends on the nature of the investigation.

It is important at this point to draw a parallel to the intra-class correlation coefficient that also derives from generalizability theory [12]. The intra-class coefficient is based on a 2-facet design, where there are often persons (facet 1, the facet of differentiation) crossed with items on a test (facet 2, the facet of generalizability). As with the generalizability coefficient, the intra-class can be assessed with a fixed or random facet of generalizability and can be a reported as a measure of consistency or agreement. Interpretations of the intra-class coefficient follow those of the generalizability coefficient. Thus, the intra-class is subsumed under the broader classification of generalizability coefficients. A specialized program, however, is required when the number of facets under investigation is more than two.

Step 3: The Decision Study (D-study)

As was mentioned earlier, once the variance components from the G-study are generated, a series of "what if" questions are answered with a D-study. Using a series of equations, more *relative G-coefficients* and *absolute G-coefficients* are calculated. These interpolate or extrapolate the expected values of the calculated coefficients when the number of observations for each facet are assumed to be reduced or increased. This is very useful to understand how best to improve the reliability of the study by adding observations to some of the most problematic facets. It also provides information on the minimum number of observations required for an acceptable level of generalizability for facets that may be very resource-intensive to collect.

Step 4: Generalizability Program Analysis

Initial descriptions of generalizability analyses have the data set up in "long" format – with each measured observation assigned a single row [11,13]. In these data set-ups, the objects in the facet of differentiation, such as participants, are coded with a numeric value and are repeated down a single column. The facets of generalizability are each given a single column with a numeric value representing a level of that facet. Finally, there is a single "dependent variable" column for the observed measurements. In this demonstration data set trials would be coded 1, 2, 3, …60 and time points would be coded 1 and 2. This approach would be advantageous in that a fully crossed random-effects ANOVA can be calculated and the ANOVA table produces easily-identifiable sums of squares and mean squares for each main, interaction, and error effect. More recently, however, computer programs have been developed specifically for generalizability theory and the data are set up in the more traditional-looking repeated measures "wide" format as in Table 1.

The program *SPSS* (IBM Corporation, Endicott, NY, USA) was used to analyze the data. The syntax for the program (G1.sps) has been designed and published for easy access for use with SPSS, SAS, and *Matlab* (Mathworks, Natick, Massachusetts, USA), free of charge [17]. One of the drawbacks of this routine is that it is restricted to analysis of three or fewer all random facets. However, for purposes of this demonstration these restrictions are met. Once the process is understood other programs that have fewer restrictions can be readily used.

Step 5: Specifying the Generalizability Model and Running the Program

A sample control file (Table 3) is provided that runs the data from Table 1 [17].

Table 3. Control file.

Command	*Explanation of Command*
setprintback = off.	*Stops printing syntax to the output window*
matrix.	*Defines the matrix to be used in the analysis*
GET scores / file = * /variables = m1_t1 to m2_t60 / missing = omit.	*Defines the variables to use in the analysis, indicates that the file to be used is the open, active window (*), the variable names, and that any missing variables are omitted. In this data set no variables are missing so the subcommand could be deleted*
compute nfacet1 = 60.	*Defines the number of trials and is the most rapid-changing facet in the data set*
compute nfacet2=2.	*Defines the number of moments and is the slower changing facet in the data set*
compute type = 3.	*Indicates that the type of data for this generalizability analysis follows a fully-crossed, all random facet design*
compute dfacet1 = {20,40,60,80}.	*Requests in the D-study that generalizability coefficients for 20, 40, 60 and 80 trials be produced*
compute dfacet2 = {1,2,3}.	*(Requests in the D-study that generalizability coefficients for 1,2, and 3 moments be produced. These requests in the D-study will be fully crossed in the output*
computegraphdat = 3.	*Plots the D-study results for the G coefficients; indicate "4" here if you want the phi coefficients plotted. Although this was requested, the graph is not shown in the results*

The rest of the G1.sps syntax is the same for all models and can be copied and pasted from the program available at:https://people.ok.ubc.ca/brioconn/gtheory/G1.sps

Step 6: Interpreting the ANOVA table and Variance Components

The ANOVA table reported in the output (Table 4) is similar to one that can be obtained from running the data in "long" form in a full factorial model with the measurements noted as the "dependent variable" and each of the p, t, and m variables listed as random factors.

The degrees of freedom, sums of squares and mean squares are all obtained in the usual ANOVA fashion. However, the additional "Variance" column is obtained using an algorithm and matrix formulae [18]. The values represent the variance components associated with each facet in the model. The "Proport." column is generated by taking the sum of the Variances and then dividing each Variance component by that sum. It provides insight into where the variance in the data is arising. In this example, 25% of the variance in the data is due to between individual participants and is expected in that there is random variability across persons. There is very

Table 4. ANOVA and Variance Table.

One or more negative variance estimates have been set to zero

ANOVA Table:

Effect	df	SS	MS	Variance	Proport.
P	15	476100143	31740010	190551	.25
F1	59	112877999	1913186	48753	.07
F2	1	6181057	6181057	0	.00
P*F1	885	336976780	380765	11861	.02
P*F2	15	132749607	8849974	141549	.19
F1*F2	59	19432368	329362	0	.00
P*F1*F2	885	315982208	357042	357042	.48

little variance (7%) due to trials and virtually none due to the moment facet (time intervals). This means that the EEG beta values, when averaged across participants and trials were consistent across the two moments of measurement. The two-way interactions show very little of the variance (2%) is due to systematic variation in Participant X Trials. A larger proportion (19%) is due to systematic Participant X Moments variation. Again, as expected, the participants themselves are responsible for a considerable amount the variance in the data set. However, most of the variance is due to the three-way interaction as well as other sources or error, which cannot be disentangled (48%). Again, this is usual, as not all variation can be controlled.

Table 5. D-Study Output.

D-Study:

Entered D-Study values for Facet 1:

20 40 60 80

Entered D-Study values for Facet 2:

1 2 3

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

D-Study G Coefficients

	1	2	3
20	.54	.70	.78
40	.56	.72	.79
60	.56	*.72*	.79
80	.57	.72	.80

D-Study Phi Coefficients

	1	2	3
20	.54	.70	.77
40	.56	.71	.78
60	.56	*.72*	.79
80	.57	.72	.79

Step 7: Interpreting the G and phi values

The G-coefficients printed are the Generalizability (Relative Generalizability, Consistency), in this case .72 and Phi (Absolute Generalizability, Agreement), in this case .72. The G- coefficients are typically higher than the phi coefficients due to the more stringent assumptions about the phi coefficients. However, the difference between them is negligible in this case, indicating that both that rank-order as well as absolute agreement level across time periods for these data are similar.

Both values are at an acceptable level, above .70 for research purposes, as suggested by traditional reliability approaches [19]. They indicate that these data are likely to be reproduced under the same circumstances, and that generalization can be made with some confidence regarding these scores to other data sets collected under similar circumstances. Due to the negligible roles played by the trials and moment facets, collapsing the data across these facets would be justified. Doing so would provide a robust estimate using a single EEG beta measurement at the C1 electrode for each participant.

Step 8: Interpreting the D-Study

The section of the output associated with the D-Study is shown below in Table 5. The obtained G (.72) and phi (.72) indices are italicized in the matrices, as these correspond to the 60 trials and 2 moments used in the initial methodology. Interpolating we can see that if we had used a method with our 16 participants, 2 time periods and 40 trials, the generalizability of the data would be G=.72 and phi=.71. This would reduce the time to collect the data, and would reduce the reliability of the data only a small amount. Extrapolating to increase the number of trials (80) and add another moment (3) to the methodology, the generalizability of the data would be G=.80 and phi=.79. This would improve the reliability of the data, but require a large increase in resources to collect the data. Instead, as noted earlier, collapsing the time 1 and time 2 data into an average measure for participants would improve the reliability of the data. These G or phi coefficient data can be plotted on a graph to visualize the information presented in the matrices. However, the graph is not reproduced here.

Discussion

It is well accepted that at a minimum, the data collected and used in a study should be reliable. While many studies continue to use traditional measures of test-retest reliability or now more frequently, the intra-class, these measures have constraints on the type of data that can be analyzed. One issue that frequently is raised is 'What is a "good" generalizability coefficient?' There is no correct response to this question, but some acceptable limits for intra-class coefficients have been proposed: values less than 0.50 are poor, between 0.50-0.75 are acceptable, between 0.75-0.90 are moderate, and greater than 0.90 are excellent [20]. If reporting of generalizability coefficients becomes usual practice such standards will evolve.

Generalizability theory provides a much-needed alternative to the more traditional approaches for several reasons. First, more than two facets (where participants are considered a facet) are handled within generalizability theory, whereas the other methods are restricted to two. Given that many studies incorporate multiple facets that would have an effect on reliability (e.g., coders, time points, etc.), generalizability models are the preferred option. Second, both absolute and relative generalizability coefficients are provided. It is helpful to have both so that the researcher can make a decision as to which one is most appropriate depending on the nature of the data and the use to which the data will be put. Third, the sources of variance contributing to the unreliability of the data are specified, providing needed information as to how to improve the reliability of a data-gathering methodology. Fourth, extrapolation and interpolation of the indices through increasing or decreasing facet points are estimated. This provides guidance in terms of what the effects on the reliability would be based on various types of changes to the data collection procedure.

A limitation of this demonstration is that the G1.sps analysis is restricted to data sets with three or fewer random facets and no missing data, but does allow for nested facets. A more comprehensive program that handles more facets and complex designs (G2.sps) is also available [17]. Other generalizability programs are also readily and freely available. These include GENOVA [18] that handles up to six different facets in a given study, and EDUG [21] that has a relatively straightforward user interface.

Conclusion

With the wide availability of statistical programs to run generalizability analyses, it is comparatively easy for researchers to include multi-faceted reliability information regarding their data. The process can be easily adapted to other research contexts where there is more than one facet that can impact data reliability. Given the central role of reliability in data measurement and its predictive utility, it is strongly encouraged that researchers adopt generalizability theory and include the results in published studies.

Additional files

Data Sheet

Competing interests

The authors declares that they have no competing interests.

Authors' contributions

Authors' contributions	AK	TK	DP	BG	JR
Research concept and design	✓	--	--	✓	✓
Collection and/or assembly of data	✓	--	✓	--	--
Data analysis and interpretation	✓	✓	✓	--	--
Writing the article	✓	✓	--	--	--
Critical revision of the article	✓	✓	✓	✓	✓
Final approval of article	✓	✓	✓	✓	✓
Statistical analysis	✓	✓	--	--	--

Acknowledgements and funding

Assistance with data collection and input were provided by Calin Gaina Ghiroaga. This work was supported in part by Alberta Innovates Technology Futures (AITF); Canada Institutes of Health Research (CIHR); National Science and Engineering Research Council (NSERC).

Publication history

EIC: Jimmy Efird, East Carolina University, USA.
Received: 03-June-2020 Final Revised: 24-July-2020
Accepted: 27-July-2020 Published: 08-Aug-2020

References

1. Doyle RJ, Ragan BG, Rajendran K, et al. **Generalizability of stabilogram diffusion analysis of center of pressure measures**. *Gait Posture* 2008;**27**(2);223-230.
2. Heitman RJ, Kovaleski JE, Pugh, SF. **Application of generalizability theory in estimating the reliability of ankle-complex laxity measurement**. *J Athl Train.* 2009;**44**(1):48-52.
3. Wastell DG, Barker GR. **Intraclass correlations: A two-facet study and some comments on the concept of reliability**. *Bull. Psychonomic Soc.* 1988;**26**(6):583-586.
4. Grant AC, Abdel-Baki SG, Weedon J, et al. **EEG interpretation reliability and interpreter confidence: A large single center study**. *Epilepsy Behav.* 2014;**32**(Mar):102-107.
5. Benbadis SR, LaFrance WC, Papandonatos GD, et al. **Interrater reliability of EEG-video monitoring**. *Neurol.* 2009;**73**(11):843-846.
6. Fernandez T, Harmonya T, Rodriguez M. et al. **Test-retest reliability of EEG spectral parameters during cognitive tasks: I absolute and relative power**. *Int J Neurol.* 1993;**68**(3-4): 255-261.
7. Harmonya T, Fernandez T, Rodriguez M, et al. **Test-retest reliability of EEG spectral parameters during cognitive tasks: II coherence**. *Int J Neurol.* 1993;**68**(3-4):263-271.
8. Salinsky MC, Oken BS, Morehead L. **Test-retest reliability in EEG frequency analysis**. *Electroenceph Clin Neurol*. 1991;**79**(5):382-392.
9. Thatcher RW. **Validity and reliability of quantitative electroencephalography (qEEG).** *J Neurother*. 2010;**14**(2):122-152.
10. Williams LM, Simms E, Clark CR, et al. **Test-retest reliability of a standardized neurocognitive and neurophysiological test battery**. *Int J*

Neurol. 2004;**115**(12):1605-1630.

11. Cardinet J, Tourneur Y, Allal L. **The symmetry of generalizability theory: Applications to educational measurement**. *J Educ Measur*. 1976; **13**(2):119-135.
12. Cronbach LJ, Nageswari R, Gleser GC. **Theory of generalizability: A liberation of reliability theory**. *Brit J Stat Psychol.* 1963;**16**(2):137-163.
13. Evans WJ, Cayten CG, Green PA. **Determining the generalizability of rating scales in clinical settings**. *Med Care*. 1981; **19**(12):1211-1220.
14. Hoyt CJ. **Test reliability estimated by analysis of variance**. *Psychometrika*. 1941;**6**(3):153-160.
15. Kline A, Goodyear B, Pittman D, Ronsky J. **A functional MRI-compatible apparatus for investigations of brain activity during simulated walking – A pilot study.** *J Biomed Eng Med Dev.* 2010; **5**(2):1-5.
16. Kline, A. *Multimodal Imaging of Cortical Networks Controlling Lower Limb Locomotion: Towards the Development of Brain-Computer Interfaces* [dissertation]. Calgary: University of Calgary; 2018.
17. Mushquash C, O'Connor BP. **SPSS and SAS programs for generalizability theory analyses**. *Behev Res Method.* 2006;**38**(3):542-547.
18. Crick JE, Brennan RL. *Manual for GENOVA: A Generalizability Analysis of Variance System* (American College Testing Technical Bulletin No. 43). Iowa City, IA:ACT;1983.
19. Nunnally JC, Bernstein IH. *Psychometric Theory* (3rd ed.). New York:McGraw-Hill;1994.
20. Koo TK, Li MY. **A guideline of selecting and reporting intraclass correlation coefficients for reliability research**. *J Chiropr Med.* 2016;**15**(2):155-163.
21. Swiss Society for Research in Education Working Group. *EDUG User Guide*. Neuchatel, Switzerland:IRDP;2010.

Citation:
Kline A, Kline T, Pittman D, Goodyear B and Ronsky J. **Generalizability Theory:Demonstrating the Process and its Utility with EEG Measurements**. *J Med Stat Inform*. 2020; **8**:3.
http://dx.doi.org/10.7243/2053-7662-8-3