Archives of Current Research International
2(1): 46-53, 2015, Article no.ACRI.2015.005

SCIENCEDOMAIN ***international***
www.sciencedomain.org

SCIENCEDOMAIN

Estimation of Generalizability Coefficient: An Application with Different Programs

Atilla Yelboga[1*]

[1]*Measurement and Evaluation Application and Research Center, Ankara University, Turkey.*

Author's contribution

The sole author designed, analyzed and interpreted and prepared the manuscript.

Article Information

DOI: 10.9734/ACRI/2015/17409
Editor(s):
(1) R. M. Chandima Ratnayake, Department of Mechanical and Structural Engineering and Materials Science, University of Stavanger, Norway.
Reviewers:
(1) Anonymous, University of Science and Technology, Taiwan.
(2) Vic W. K. Mok, Department of Applied Science, Hong Kong Institute of Vocational Education, VTC, Hong Kong, China.
(3) Anonymous, Kwame Nkrumah University of Science and Technology, Ghana.
Complete Peer review History: http://www.sciencedomain.org/review-history.php?iid=925&id=41&aid=9395

Original Research Article

Received 12th March 2015
Accepted 2nd May 2015
Published 25th May 2015

ABSTRACT

The main objective of making investigation about the field of psychology, education and behavioral sciences is to achieve a consistent measurement results. When carrying out the evaluation, many researchers use the classical test theory and other researchers use the generalizability theory. Generalizability theory is used less than the classical test theory by researchers. The main reason is generalizability theory analysis cannot perform by popular statistical programs. In this study, generalizability theory analysis was performed with different statistical programs such as GENOVA, EduG, with SPSS and SAS. In the study, GENOVA, EduG, SPSS and SAS programs outcomes were discussed as the results obtained in the study of alternative decisions in terms of variance components, partial and absolute errors and the results obtained generalizability coefficients. After the research results, there was no significant differences between the results obtained with GENOVA, EduG, SPSS and SAS program.

Aims: SPSS and SAS programs are statistical programs and generally used by researchers which help to perform various analyzes. As for GENOVA and EduG program, they are only written for making G theory analysis. It is considered that widely used statistics programs such as SPSS and SAS programs are written for making the analysis of a G theory. Therefore, when making G theory analysis, there is no need for different software like GENOVA and EduG program. However, it should be noted that it is required to take the same results in different programs due to consistency

**Corresponding author: Email: ayelboga@gmail.com;*

of results. The purpose of this study, examination of G theory analysis with different programs GENOVA, EduG ve SPSS-SAS and prediction of $E\rho^2$ and Φ coefficents.

Study Design: Data are collected from two different departments of a company with a proficiency exam (tasks). In this study, two different situations were examined.

Methodology: The first situation is crossed pattern [person(p) x task(t) x evaluator(e)] that includes an evaluation of the adequacy for 10 personnel in 4 different tasks and it is made by 2 evaluators. In other words, every staff made all the tasks and the way to do all of the tasks which was made by staff also evaluated by 2 evaluators. The second situation is mixed pattern [person(p) x (task(t): evaluator(e))] that includes an evaluation for each task in different way for 10 personnel in 3 different tasks and it is scored by 4 different evaluators. In G theory analysis to reach $E\rho^2$ and Φ coefficients, it must be made analysis which is appropriate for G theory (pattern) by GENOVA, EduG, SPSS and SAS programs.

Results: When comparing results, analysis are made with using two different designs in generalizability theory and also they are made with GENOVA, EduG, SPSS and SAS programs. (a) The estimated variance of the G study results are generated(p): person; (t): task; (e): evaluator; main effect with (pt): person x task; (pe): person x evaluator; (te): task x evaluator; (pte,c): person x task x evaluator common effects and (p):person; (t):task;(t:e): task : evaluator; (pt): person x task; (pt:e): person x task: evaluator, (b) G study is estimated with $\sigma^2(\delta)$ relative and $\sigma^2(\Delta)$ of the absolute error variance, (c) K study is estimated with $E\rho^2$ and Φ coefficients, (d) After alternative D study, number of different evaluators and tasks obtained from the combination of variance and Φ and $E\rho^2$ coefficients found to be consistent with each other.

Keywords: Generalizability theory; Genova; EduG; SPSS; SAS.

1. INTRODUCTION

The main objective of making investigation about the field of psychology, education and behavioral sciences is to achieve a consistent measurement results. It can be seen that, when assessing the reliability of measurement results, most common approaches are classical test theory, item response theory and generalizability theory. However, the classical test theory is more preferred than the other approaches by researchers. Searchers avoid using this approach because the analysis of both item response theory and generalizability theory are not possible to do with popular statistical programs.

In classical test theory, reliability is usually defined as the correlation between parallel tests [1]. Classical test theory is based on the idea of producing a single reliability coefficient of a group of parallel observations which is the only real points with each observation or test scores [2]. Currently, the most widely used reliability methods are reliability estimates (ex: test retest, alternative forms, consistency coefficient and inter-rater reliability) which are provided with the classical test theory [3].

Generalizability (G) theory [4] is a different approach and a process of determining the limits of generalizability of the results obtained from the measurement made. In other words, it evaluates the conclusions of the assessment process situations (ex: different persons, places and times and one of the measurement characteristics is how good to measure) which can be applied [5].

G-Theory can also assess the reliability of classical approach, but reveals that in a limited way [5]. In classical test theory, the reliability of the calculation method is different. Also, it takes different names according to mean of reliability and the only error source which is handled [6]. When calculating reliability with test-retest method which is a meaning of consistency, the source of error is time. As for calculating reliability with parallel (equivalent) form methods which means equivalency or consistency, forms will be handled as a source of error. The calculations of the reliability in terms of internal consistency, some methods are used like Cronbach alpha, Rulon, Flanagen, Mossier, Horst and they take into consideration substances as a source of error [7,8]. Because of this, it is taking into account the different sources of error in calculating the reliability of the results of classical test theory actually which shows the contradictions of the classical test theory [4].

1.1 G-Theory Theoretical Framework

G theory is a theory that can be used to determine the amount of variance attributable to different error sources [5]. In G theory, a coefficient called the generalizability coefficient is calculated. This factor does not reinterpreting the concepts of reliability, although similar to the reliability coefficient in the classical test theory.

In G theory, there are two types of studies which are called Generalizability (G) and Decision (D) Studies. G Studies are arranged to separate the measurement errors with the specific sources of variability. When regulating G theory, people must primarily to determine the universe of potential sources of variability and measurement sample to generalize. In D study, trying to minimize the error which is a specific purpose in a made to measure obtained in G study. The reliability of the measurements made in the work being evaluated in G study, decision study which is a work from the generalizability study movement organized to collect data based on the decision of the motion generalizability study [1,7,8].

In D study, two different reliability coefficient which are generalizability coefficient (G or Eρ^2) and dependability coefficient (Phi or Φ) can be calculated for different measurement scenarios [1,7-9].

Alternative K Studies can be edited to reduce the error variance which takes part in Eρ^2 and Φ coefficients. In G-Theory, this kind of research is called Alternative D Studies [1,7- 9].

It can be seen that, the basic assumptions of the G- Theory was revealed by [4] and has been investigated by different researchers (ex: [7,8,10-17]) in different aspects in the process. Beside this, some researchers (ex: [18]) define G theory to the psychometric approach which is parallel to the development of psychometric research.

In Turkey, although there are qualified researches (ex: [19-30]) on the G-Theory, it can be seen that their numbers are limited according to the abroad studies. It is observed that, the old time studies were usually calculated with GENOVA (A Generalized Analysis of Variance System) program [31], in new studies, SPSS and EduG programs are used partially. In the preparation of this study, in our country, there was no study about statistics programs like SAS and MATLAB which is used in G theory.

The first program which is written for G theory analysis is GENOVA program. mGENOVA and urGENOVA programs follow this. Also, another program for make G theory analysis is EduG program. Some researchers (ex: [32]) made studies to use G theory analysis for SPSS (Statistical Package for the Social Sciences) and SAS (Statistical Analysis System) programs with using syntax by [7].

GENOVA program which is used in G theory analysis, developed with using the fortran programming language by [31,33]. As for EduG program, it was developed by [33]. GENOVA program is delivered free by American College Testing Program. Regarding EduG program, it is delivered free by EDUCAN Inc. and IRDP.

The SPSS and SAS are statistical software programs which are distributed. SPSS program was established in 1967. It is used in applications like statistical analysis, data mining and predictive analytics. As regards SAS, it has been used with a project at North Carolina State University in 1976.

Unlike G theory, SPSS and SAS programs make different statistical analysis process. Although SPSS and SAS programs easily make many statistical analyses, they cannot make G-Theory analysis practically like GENOVA and EduG programs. GENOVA and EduG programs are developed only for G theory analysis. G theory analysis can made with the help of syntax by [32] without interfering with the contents of the SPSS and SAS programs.

2. AIM

SPSS and SAS programs are statistical programs and generally used by researchers which help to perform various analyzes. As for GENOVA and EduG program, they are only written for making G theory analysis. It is considered that widely used statistics programs such as SPSS and SAS programs are written for making the analysis of a G theory. Therefore, when making G theory analysis, there is no need for different software like GENOVA and EduG program. However, it should be noted that it is required to take the same results in different programs due to consistency of results. The purpose of this study, examination of G theory analysis with different programs GENOVA [31], EduG [33] ve SPSS-SAS [32] and prediction of Eρ^2 and Φ coefficents.

3. METHODS

In the following, data collection and data analysis sections are given for research on the screeing types that examines different statistical programs in G theory.

3.1 Data Collection

Data are collected from two different departments of a company with a proficiency exam (tasks). In this study, two different situations were examined.

The first situation is crossed pattern [person(p) x task(t) x evaluator(e)] that includes an evaluation of the adequacy for ten personnel in four different tasks and it is made by two evaluators. In other words, every staff made all the tasks and the way to do all of the tasks which was made by staff also evaluated by two evaluators.

The second situation is mixed pattern [person(p) x (task(t): evaluator(e))] that includes an evaluation for each task in different way for ten personnel in three different tasks and it is scored by four different evaluators.

3.2 Data Analysis

In G theory analysis to reach Eρ^2 and Φ coefficients, it must be made analysis which is appropriate for G theory (pattern) by GENOVA, EduG, SPSS and SAS programs.

3.2.1 Pattern 1

The equations used to calculate the analysis of variance person(p) x task(t) x evaluator(e) is given for pattern in Table 1. (p): person; (t): task; (e): evaluator; (pt): person x task; (te): task x evaluator; (pte,e): person x task x evaluator were used for this pattern. A modeling has been done in crossed pattern within the tasks (p x t x e).

3.2.2 Pattern 2

The equations used to calculate the analysis of variance [person(p) x task(t): evaluator(e)] is given in Table 2. (p): person; (t): task; (e): evaluator; (pt): person x task; (te): task x evaluator; (pte,e): person x task : evaluator were used for this pattern. A modeling has been done in nestled pattern within the materials (p x t: e).

GENOVA, EduG, SPSS and SAS programs are used for calculating correlations of Eρ^2(G) and Φ(Phi) for the reliability of scores on two different patterns.

4. FINDINGS

The aim of this study is first in line variance components were calculated by GENOVA, EduG, SPSS and SAS programs and later, for [(p) x (t) x (e)] and [(p) x (t): (e)] patterns, Eρ^2 (G) ve Φ (Phi) coefficient were estimated.

Table 1. Generalizability theory [(p) x (t) x (e)] pattern variance components

Variance source	Variance components
person	$\sigma_p^2 = \dfrac{MS_p - MS_{pe} - MS_{pt} + MS_{pte,e}}{n_e n_t}$
task	$\sigma_t^2 = \dfrac{MS_t - MS_{pt} - MS_{et} + MS_{pte,e}}{n_t n_e}$
evaluator	$\sigma_e^2 = \dfrac{MS_e - MS_{pe} - MS_{et} + MS_{pte,e}}{n_p n_t}$
person x task	$\sigma_{pt}^2 = \dfrac{MS_{pt} - MS_{pte,e}}{n_p}$
person x evaluator	$\sigma_{pe}^2 = \dfrac{MS_{pe} - MS_{pte,e}}{n_e}$
task x evaluator	$\sigma_{te}^2 = \dfrac{MS_{te} - MS_{pte,e}}{n_p}$
error	$\sigma_{pte,e}^2 = MS_{pte,e}$

Note: σ^2: variance; MS: Mean Square; n: data; (p): person; (t): task; (e): evaluator; (pt): person x task; (pe): person x evaluator; (te): task x evaluator; (pte,e): person x task x evaluator

Table 2. Generalizability theory [(p) x (t): (e)] pattern variance components

Variance source	Variance components
Person	$\sigma_p^2 = \dfrac{MS_p - MS_{pe}}{n_t n_e}$
Evaluator	$\sigma_e^2 = \dfrac{MS_e - MS_{(te)} - MS_{pe} + MS_{pt:e}}{n_p n_t}$
Task:evaluator	$\sigma_{t:e}^2 = \dfrac{MS_{t:e} - MS_{pt:e}}{n_p}$
Person x evaluator	$\sigma_{pe}^2 = \dfrac{MS_{pe} - MS_{pt:e}}{n_t}$
Error	$\sigma_{pt:e}^2 = MS_{pt:e}$

Note: σ^2: variance ; MS: Mean Square ; n: data; (p): person; (t): task; (e): evaluator; (t:e): task: evaluator; (pe): person x evaluator; (pt:e): person x task: evaluator

The obtained variance components by GENOVA, EduG, SPSS and SAS programs are given in the Table 3. As shown in Table 3; df: degree of freedom, SS: Sum of Squares; MS: Mean Square; σ^2: variance; (p): person; (t): task; (e): evaluator; (pt): person x task; (pe): person x evaluator; (te): task x evaluator; (pte,e): person x task x evaluator calculated. The most important issues are the percent variance calculation obtained. The percentage variance of the persons were found to be 20,0% in GENOVA, EduG, SPSS and SAS programs. Meanwhile, task and evaluator variability explains 20,0% of the total variance, respectively 15,9% and 0,3%. When considering the main effect variance, the results obtained are consistent with each other. According to this result, approximately 35% of the variance was found that the variation between individuals (in four tasks and two evaluators). Similarly, examined (p): person; (t): task; (e): evaluator main effects with (pt): person x task; (pe): person x evaluator; (te): task x evaluator joint effects, the results obtained are consistent with each other. (pte,e): person x task x evaluator effect of showing the joint effect of the three sources of variability, yielded similar results.

As shown in Table 4; df: degree of freedom, SS: Sum of Squares; MS: Mean Square; σ^2: variance; (p): person; (t): task; (t:e): task : evaluator; (pt) : person x task; (pt:e): person x task : evaluator calculated by GENOVA, EduG, SPSS and SAS programs. The percentage variance of the persons were found to be 10,8% in GENOVA, EduG, SPSS and SAS programs. On the other hand, (p): person; (t:e): tasks : evaluator; (pt): person x task; (pt:e): person x task : evaluator variability of the total variance ratios are consistent with each statement in GENOVA, EduG, SPSS and SAS programs.

G theory with D study interpreted similar to the reliability factor calculated on classical test theory $E\rho^2$ (G) calculating coefficients is possible. $E\rho^2$ (G) is used for relative decisions. In addition, Φ coefficient is calculated for absolute decisions in G theory. Φ coefficient is a coefficient which can be calculated on classical test theory.

bxgxd pattern obtained for $E\rho^2$ (G) and Φ (Phi) coefficients are compatible with each other by GENOVA, EduG, SPSS and SAS programs. Predicted coefficients of the GENOVA, EduG, SPSS ve SAS programs are the same as the results obtained are shown in a single table (Table-5). According to G theory, 3-point scale study that estimated by D study, $E\rho^2$ (G) coefficient of 0.640, while the Φ coefficient seems to be estimated as 0.554.

Applying different tasks and different alternatives and combinations assessors work with alternative D study, relative and absolute error variance and the coefficient of $E\rho^2$ (G) and Φ has produced consistent results in GENOVA, EduG, SPSS and SAS programs.

As shown in Table 6, for [bx(g:d)] pattern, the coefficients obtained by increasing the number of task $E\rho^2$ (G) = 0.681 and Φ = 0,601. Similarly, the coefficients obtained by increasing the number of evaluator $E\rho^2$ (G) = 0,684 ve Φ = 0,592.

Alternatively D study results, applying different tasks and different evaluator alternatives, the $E\rho^2$ (G) and Φ coefficients obtained are consistent with each other in GENOVA, EduG, SPSS ve SAS programs.

5. RESULTS AND DISCUSSION

When comparing results, analysis are made with using two different designs in generalizability theory and also they are made with GENOVA, EduG, SPSS and SAS programs.

Table 3. [(b) x (g) x (d)] pattern variance components

mGENOVA ve EduG (Variance components)						**SPSS ve SAS (Variance components)**					
	df(α)	**SS(α)**	**MS(α)**	**σ^2(α)**	**%**		**df(α)**	**SS(α)**	**MS(α)**	**σ^2(α)**	**%**
p	9	62,200	6,9111	0,5528	20,0	p	9	62,200	6,911	0,553	20,0
t	3	37,450	12,4833	0,4417	15,9	t	3	37,450	12,483	0,442	15,9
e	1	3,200	3,2000	0,0074	0,3	e	1	3,200	3,200	0,007	0,3
pt	27	56,300	2,0852	0,5750	20,8	pt	27	56,300	2,085	0,575	20,8
pe	9	12,050	1,3389	0,1009	3,6	pe	9	12,050	1,339	0,101	3,6
te	3	7,500	2,5000	0,1565	5,7	te	3	7,500	2,500	0,156	5,7
pte, e	27	25,250	0,9352	0,9352	33,8	pte, e	27	25,250	0,935	0,935	33,8

Note: df : degree of freedom; SS: Sum of Square; MS: Mean Square; σ^2: variance; (p) :person; (t): task; (e):evaluator; (pt): person x task; (pe): person x evaluator; (te): task x evaluator; (pte,e): person x task x evaluator

Table 4. [(p) x (t) : (e)] pattern variance components

mGENOVA ve EduG (Variance components)						SPSS ve SAS (Variance components)					
	df(α)	SS(α)	MS(α)	$\sigma^2(\alpha)$	%		df(α)	SS(α)	MS(α)	$\sigma^2(\alpha)$	%
p	9	92,6667	10,2963	0,4731	10,8	p	9	92,667	10,296	0,473	10,8
t	2	48,2000	24,1000	0,3252	7,4	t	2	48,200	24,100	0,325	7,4
t:e	9	79,7000	8,8556	0,6475	14,8	t:e	9	79,700	8,856	0,648	14,8
pt	18	83,1333	4,6185	0,5596	12,8	pt	18	83,133	4,619	0,560	12,8
pt:e	81	192,8000	2,3802	2,3802	54,3	pt:e	81	192,800	2,380	0,238	54,3

Note: df : degree of freedom; SS: Sum of Square; MS: Mean Square; σ^2: variance; (p): person; (t): task; (t:e): task : evaluator; (pt): person x task; (te): task x evaluator; (pt : e): person x task : evaluator

Table 5. [(p) x (t) x (e)] pattern alternative D studies results

$\sigma^2(\alpha)$	n_t	3	4	5	3	4	5	3	4	5
	n_e	1	1	1	2	2	2	3	3	3
$\sigma^2(\delta)$		0,604	0,478	0,403	0,398	0,311	0,259	0,329	0,255	0,211
$\sigma^2(\Delta)$		0,811	0,635	0,530	0,575	0,445	0,367	0,496	0,381	0,312
$E\rho^2$		0,478	0,536	0,578	0,581	0,640	0,681	0,627	0,684	0,724
Φ		0,405	0,465	0,511	0,490	0,554	0,601	0,527	0,592	0,639

Note: n_t and n_e.: number of task and evaluator; $\sigma^2(\delta)$: relative error variance; $\sigma^2(\Delta)$: absolute error variance; $E\rho^2$: generalizability coefficient; Φ: dependability coefficient.

Table 6. [(p) x (t):(e)] pattern alternative D studies results

$\sigma^2(\alpha)$	n_t	2	3	4	2	3	4	2	3	4
	$n_{t:e}$	3	3	3	4	4	4	5	5	5
$\sigma^2(\delta)$		0,676	0,451	0,338	0,577	0,385	0,289	0,518	0,345	0,259
$\sigma^2(\Delta)$		0,947	0,631	0,473	0,821	0,547	0,410	0,745	0,497	0,373
$E\rho^2$		0,412	0,512	0,583	0,450	0,551	0,621	0,477	0,578	0,646
Φ		0,333	0,428	0,500	0,366	0,464	0,535	0,388	0,488	0,559

Note: n_t and $n_{t:e}$.: number of task and evaluator; $\sigma^2(\delta)$: relative error variance; $\sigma^2(\Delta)$: absolute error variance; $E\rho^2$: generalizability coefficient; Φ: dependability coefficient

a. The estimated variance of the G study results are generated (p): person; (t): task; (e): evaluator; main effect with (pt): person x task; (pe): person x evaluator; (te): task x evaluator; (pte,c): person x task x evaluator common effects and (p):person; (t): task;(t:e): task : evaluator; (pt): person x task; (pt:e): person x task : evaluator
b. G study is estimated with $\sigma^2(\delta)$ relative and $\sigma^2(\Delta)$of the absolute error variance,
c. K study is estimated with $E\rho^2$ and Φ coefficients,
d. After alternative D study, number of different evaluators and tasks obtained from the combination of variance and Φ and $E\rho^2$ coefficients found to be consistent with each other.

GENOVA, EduG, SPSS and SAS programs are made G theory analyses which are basically written by the same algorithm. Thus, it is expected that the results are consistent with each other. In a study, [21] compared the GENOVA and SPSS analysis results. As a result of research, $E\rho^2$ and Φ coefficients was close to each other but not exactly the same which was obtained with GENOVA and SPSS programs. There is a suggestion that it may have been due to rounding in the calculation. It was seen that [21]'s conclusions reached was consistent as this study. In addition to this, it can be said that G theory analysis can be freely used in EduG and SAS programs. In G theory, the results which were obtained with EduG and SAS programs are consistent like GENOVA and SPSS programs.

6. CONCLUSION

Classical test theory is a theory that still preferred by researchers and maintains its popularity. However, with the help of one analysis, a common and single reliability coefficient cannot be obtained in the classical test theory [25,27]. The differences between estimation of different methods illustrate the limitations of classical test theory. Despite this, one of the important reasons that G theory is not widely used than the classical test theory is the programs which will be used in the analysis are not widely used as SPSS and SAS programs [32].

Nevertheless, researchers who do work on the G theory, using one of the programs such as GENOVA, EduG, SPSS and SAS programs are given the following suggestions;

1. While GENOVA and EduG programs have no commercial purposes with free delivery, SPSS and SAS programs are licensed trademarks programs. Users can download and use for free EduG and GENOVA programs on their computers, but SPSS and SAS programs are required user licenses to run the program effectively.
2. Researchers who think to use GENOVA program in G-Theory analysis have to learn program commands and their syntax to write appropriate commands (see [31]) which is proper to schedule.
3. EduG program was developed only for G theory analysis and also its visual interface is more understandable than GENOVA program. That is to say, it provides ease of use to users.
4. In SPSS and SAS program, it is sufficient that data which is written by [32], entered data syntax or read data from another file. If researchers prefer to use SPSS or SAS programs in G theory analysis, they do not need to learn command or rule.
5. Data input is different in GENOVA, EduG, SPSS and SAS programs. The use according to the researchers, think that the program will be required to input data.

COMPETING INTERESTS

Author has declared that no competing interests exist.

REFERENCES

1. Crocker LM, Algina L. Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston; 1986.
2. Lord FM, Novick MR. Statistical theories of mental test scores. New Jersey: Addison-Wesley; 1968.
3. Webb NM, Rowley GL, Shavelson RJ. Using generalizability theory in counseling and development. Measurement and Evaluation in Counseling and Development. 1988;21:81-90.
4. Cronbach L, Gleser GC, Nanda H, Rajaratnam N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley; 1972.
5. Aguinis H, Henle CA, Ostroff C. Measurement in work and organizational psychology. In N. Anderson, D. S. Ones, Sinangil HK, Viswesvaran C. (Eds.), Handbook of industrial, work, & organizational psychology. 2001;27-50. London/Amsterdam: Sage.
6. Baykul Y. Measurement education and psychology. Ankara: Cem Printing; 2000.
7. Brennan RL. Generalizability theory. New York/Berlin: Springer-Verlag; 2001.
8. Shavelson RJ, Webb NM. Generalizability theory: A primer. Newburry Park CA: Sage; 1991.
9. Shavelson RJ, Webb NM. Generalizability theory. In K. Kemp-Leonard (Ed.), Encyclopedia of social measurement Oxford, UK: Elsevier. 2004;99-105.
10. Brennan RL. (Mis) conceptions about generalizability theory. Educational Measurement: Issues and Practice. 2000a;19:5-10.
11. Brennan RL. Performance assessment from the perspectives of generalizability theory. Applied Psychological Measurement. 2000b;24:339-353.
12. Hoyt WT, Melby JN. Dependability of measurment in counseling psychology: An introduction to generalizability theory. Counseling Psychologist. 1999;325-352.
13. Goodwin LD. Interrater agreement and reliability. Measurement in Psychical Education and Exercises Science. 2001;5(1):13-14.
14. Kieffer KM. Why generalizability theory is essential and classical test theory is often inadequate. In B. Thompson, (Ed.), Advances in social science methodology. Stanford, CT: JAI. 1999;5:149-170.
15. Marcoulides GA. Generalizability theory: Picking up where the rasch IRT model leaves off? In Embretson SE, Hershberger SL (Eds.), The new rules of measurement: What every psychologist and educator should know. Mahwah, Nj: Erlbaum. 1999;129-152.
16. Marcoulides GA. Generalizability theory. In Tinsley H, Brown SD. (Eds.), Handbook of applied multivariate statistics and mathematical modeling San Diego. Academic Press. 2000;527-551.
17. Strube MJ. Reliability and generalizability theory. In Grimm LG, Yarnold PR. (Eds.), Reading and understanding more

multivariate statistics Washington, DC: American Psychological Association. 2000; 23-66.
18. Furr MR, Bacharach VR. Psychometrics: An introduction. Thousand Oaks, California: Sage; 2008.
19. Atilgan H. Comparison of generalizability theory and multivariate models sourced rasch. Unpublished doctoral dissertation. Hacettepe University. Ankara; 2004.
20. Atilgan H, Tezbasaran AA. An investigation on consistency of G and Phi coefficients obtained by generalizability theory alternative decisions study for scenarios and actual cases. Eurasian Journal of Educational Research. 2005;18:28-40.
21. Guler N. Generalizability theory and comparison of the results of G and D studies computed by SPSS and GENOVA packet programs. Education and Science. 2009;34(154):93-103.
22. Guler N, Gelbal S. Studying reliability of open ended mathematics items according to the classical test theory and generalizability theory. Educational Sciences: Theory & Practice. 2010;10(2).
23. Kan A. Effect of using a scoring guide on essay scores: Generalizability theory, Perceptual and Motor Skills. 2007;105: 891-905.
24. Kan, A. Effect of scale response format on psychometric properties in teaching self-efficacy. Eurasian Educational Research Journal, 2009;34:215-228.
25. Yelboga, A. The assessment of reliability with generalizability theory: An application in industrial and organizational psychology. Studies in Psychology. 2008;28:35-54.
26. Tasdelen G, Kelecioglu H, Guler N. Nedelsky and standard-setting methods angoff comparison with the theory of generalizability of the obtained cutting points. Journal of Measurement and Evaluation in Education and Psychology. 2010;1(1):22-28.
27. Yelboga A, Tavsancil E. The examination of reliability according to classical test and generalizability on a job performance scale. Educational Sciences: Theory & Practice. 2010;10(3):1825-1854.
28. Anil D, Buyukkidik S. A sample application with generalizability theory in four facet mixed usage patterns. Journal of Measurement and Evaluation in Education and Psychology. 2012;3(2):291-296.
29. Yelboga, A. Dependability of job performance ratings according to generalizability theory. Education and Science. 2012, 37(163), 157-164.
30. Ozberk EH, Gelbal S. Comparison of different variance components estimation methods used in the decision study generalizability theory. Journal of Measurement and Evaluation in Education and Psychology. 2014;5(2):91-103.
31. Crick JE, Brennan RL. Manual for GENOVA: A generalized analysis of variance system. (American College Testing Technical Bulletin 43). Iowa City: ACT, Inc; 1983.
32. Mushquash C, O'Connor BP. SPSS and SAS programs for generalizability theory analyses. Behavior Research Methods. 2006;38(3):542-547.
33. Cardinet J, Johnson S, Pini G. Applying generalizability theory using EduG. New York: Routledge Academic; 2009.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=925&id=41&aid=9395