Chapter 6

GENERALIZABILITY THEORY

Hakan ATILGAN[1]

INTRODUCTION

In education, different measurements are conducted to determine individuals' characteristics in cognitive, affective and psychomotor domains. In order to carry out these measurements, a number of measurement instruments are used such as control lists, rating scales, rubrics, tests, observations forms etc. However, measurements conducted in social sciences and one of the sub-fields of it, education, and in natural sciences as well cannot be precisely accurate. Many factors affecting measurement lead to errors in the results of the measurement. Due to these errors intervening with the measurement results from different sources, the results are not equal to the true score of the characteristic being measured. True score cannot be measured directly most of the time and X-E is called true score, where X is the measurement result (or observed score) and E is the measurement errors intervening with the result from different sources (or error score) (Gullksen, 1950; Lord & Novick, 2008). In measurement studies, the expectation is that the measurement result is close to the true score of the characteristic being measured (Baykul, 2000). This is only possible on condition that the error intervening with the measurement results is low. Therefore, the primary aim of measurement studies is to produce measurement instruments that can attain the true score of the measured characteristic as much as possible and to use the scores to be obtained from the measurement by making them as error-free as possible. The degree to which measurement results are free of measurement errors is defined as reliability (American Educational Research Association, American Psychological Association & National Council on Measurement in Education, 2014).

Historically, various procedures have been developed to estimate reliability within the scope of Classical Test Theory. These reliability estimation procedures are given different names and have different meanings based on the sources of measurement errors. Test-retest reliability depends on the correlation between the scores obtained from two performances of a single test held at different times and shows the stability of test scores over time. In test-retest reliability estimation, the source of error is the time between the two administrations of the test

[1] Assoc. Prof. Dr., Ege University, email: hakan.atilgan@ege.edu.tr

$$\Phi = \frac{\sigma_p^2}{\sigma_p^2 + \Delta} \tag{17}$$

Alternative D studies are used to optimize the number of conditions of each facet to reach the desired reliability level (Brennan, 2001; Shavelson & Webb, 1991). K studies can be designed to reduce the error variances in G and Φ coefficients. Similar to the Spearman-Brown prediction coefficient in the Classical Test Theory or the mean standard error in sampling theory, alternatives can be sought to reach higher G and Φ scores and reduce error variances by writing different combinations of item and rater numbers instead of the item and rater numbers represented as n_i and n_r in Equation 14 and Equation 15 (Atılgan, 2004).

REFERENCES

Allal, L. (1990). Generalizability theory. J. a. Edited by: Walberg içinde, *The international encyclopaedia of educational evaluation* (s. 274-279). Oxford: Pergamon Press.

American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standarts for Educational and Psychological Testing.* Washington, DC: American Educational Research Association.

Atılgan, H. (2004). *Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılmasına İlişkin Bir Araştırma. [A Research on Comparisons of Generalizability Theory and Many Facets Rasch Measurement].* Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü (Unpublished Phd Thesis).

Atılgan, H. (2008). Using Generalizability theory to assess the score reliability of the Special Ability Selection Examinations for music education programmes in higher education. *International Journal of Research & Method in Education*, 31(1), 63-76.

Atılgan, H., Kan, A., & Aydın, B. (2017). *Eğitimde Ölçme ve Değerlendirme [Measurement and Evaluation in Education].* Ankara: Anı Yayıncılık.

Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması [Measurement in Edıcation and Psychology: Clasical Test Theory and Aplication].* Ankara: ÖSYM.

Brennan, R. L. (1983). *Elements of generalizability theory.* Iowa: American College Testing Program.

Brennan, R. L. (2001). *Generalizability Theory.* New York: Springer-Verlag.

Burt, C. (1936). The analysis of examination marks. P. H. (Eds.) içinde, *the marks of examiners* (s. 245-314). Londan: Macmillan.

Cardinet, J., Johnson, S., & Pini, G. (2010). *Applying Generalizability Theory using EduG.* New York: Routledge.

Crocker, L., & Algina, J. (2008). *Introduction to Classical and Modern Test Theory.* Mason, Ohio: Cengage Learning.

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

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberization of reliability theory. *British Journal of statistical Psychology*, 16, 137-163.

Ebel, R. L. (1951). Estimation of the reliability of ratings. *Psychometrika*, 16, 407-424.

Gleser, G. C., Cronbach, L. J., & Rajaratnam, N. (1965). Generalizability of scores influenced by multiple sources of variance. *Psychometrika*, 30(4):395-418.

Gullksen, H. (1950). *Theory of Mental Tests.* New York: John Wiley & Sons, Inc.

Hoyt, C. (1941). Test Reliability Estimated by Analysis of Variance. *Pychometrica*, 6(3), 153-160.

Lindquist, E. F. (1953). *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin.

Lord, F. M., & Novick, M. R. (2008). *Statistical Theories of Mental Test Scores.* IAP - Information Age Publishing Inc.

Medley, D. M., & Meitzel, H. E. (1963). Measuring classroom behavior by systematic. N. L. Gage içinde, *Handbook of research on teaching.* Chicago: Rand McNally.

Rajaratnam, N., Cronbach, L. J., & Gleser, G. C. (1965). Generalizability of stratified-parallel tests. *Psychometrika*, 30(1), 39-56.

Shavelson, R. J., & Webb, N. M. (1991). *Generalizability Theory: A Primer.* Newbury Park, CA: Sage Publicıuiona, Inc.