DOCUMENT RESUME

ED 270 486 TM 860 369

AUTHOR Allal, Linda
TITLE Extensions of Generalizability Theory through the Principle of Symmetry.
PUB DATE Apr 86
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (70th, San Francisco, CA, April 16-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Data Collection; Educational Research; *Factor Analysis; Factor Structure; Foreign Countries; *Generalizability Theory; Item Analysis; Item Banks; *Measurement Techniques; Program Evaluation; *Research Methodology; Research Problems; Standardized Tests; *Symmetry

ABSTRACT

This paper discusses the theoretical scope and practical applicability of generalizability (G) theory through the principle of symmetry. Major ideas are summarized and factors hindering applications of G theory in research conducted in French-speaking Europe are presented. The principle of symmetry affirms that any factor of a design can be selected as an object of measurement and that the G theory operations defined for one factor can be transposed in the study of other factors. The principle allows the extension of G theory to situations based on complex factorial designs and involves multiple purposes of measurement in three major directions: (1) consideration of all types of facets; (2) analysis of multifaceted populations; and (3) development of a general framework for analysis. Widespread application of G theory is unlikely to occur until specialists in program evaluation develop procedures for integrating the collection and analysis of quantitative data with the application of quantitative methods of investigation. One of the most potentially useful applications of generalizability theory is the procedures it provides for using data from an initial study to determine improvements of the design to be used in subsequent research or in decision making. (PN)

AERA Annual Meeting, San Francisco, 16-20 April 1986

Symposium on "New Developments in Generalizability Theory : International Perspectives"

EXTENSIONS OF GENERALIZABILITY THEORY THROUGH THE PRINCIPLE OF SYMMETRY

Linda Allal
University of Geneva

The work I will be presenting has been carried out by a group of three persons affiliated with three different institutions. The group includes, in addition to myself, Jean Cardinet, director of research at the institute of research and documentation attached to the public departments of education in the French-speaking region of Switzerland, and Yvan Tourneur, professor at the University of Mons in Belgium.[1] Cardinet, the senior researcher of our group, has an interest in psychometric theory that dates back to his doctoral studies in the 1950's with Thurstone, followed by work with Cronbach during a second period spent in the United States in the 1960's. Yvan Tourneur began work on the generalizability model and its relationship to other measurement models in his doctoral dissertation presented at the University of Mons in 1974. My own interest in G theory derives less from previous work on psychometric and statistical models than from my belief that measurement in education - whether in the area of classroom or curriculum evaluation - needs and deserves improvement, and can benefit from the techniques generalizability analysis has to offer. A common interest of our group lies in our desire to provide research and evaluation in French-speaking Europe with stronger methodological foundations. More concretely, we are all three confronted with questions regarding the design and implementation of studies of how schools (children, teachers, curricula, methods, etc.) function and we would all three like to find tools that help provide more accurate and valid answers to these questions.

Our work has been primarily oriented toward an effort to extend the theoretical scope and the practical applicability of generalizability theory through the "principle of symmetry". This principle was first presented in an article in the Journal of Educational Measurement (Cardinet, Tourneur & Allal, 1976); its implications have since been developped in a series of publications, principally Cardinet, Tourneur & Allal (1981/1982), Cardinet & Allal (1983), Cardinet & Tourneur (1985).

In this symposium presentation, I will first summarize the major ideas developed in our publications and will then mention several factors which presently hinder applications of generalizability theory in research conducted in French-speaking Europe, and perhaps elsewhere.

The principle of symmetry

In the initial formulation of generalizability theory by Cronbach and his associates (Cronbach, Rajaratnam & Gleser, 1963; Cronbach, Gleser, Nanda & Rajaratnam, 1972), it is assumed that characteristics of persons, or groups of persons, are the object of measurement, while other factors, such as items, testing occasions, correctors, etc., are sources of measurement error. This adoption of the classical aim of psychometrics is reflected in the fact that the term *facet* is applied to all factors of the data collection design except the factor "persons". The principle of symmetry affirms that any factor of a design can be selected as an object of measurement, and that the G-theory operations defined for one factor can be transposed in the study of other factors (Cardinet *et al*, 1981, p. 184). If we consider a very simple design in which a random sample of persons is crossed with a random sample of test items, we have, according to Cronbach *et al.* 's terminology, a 1-facet design aimed at measuring the traits of persons while generalizing over randomly sampled levels of the facet "items". The principle of symmetry accepts this first measurement aim as a longstanding and legitimate concern of psychometricians, but adds a second possibility of potential interest to educational researchers: i.e., the measurement of achievement levels attained for different items (or groups of items) while generalizing over randomly sampled levels of the facet "persons".

Beyond the case of simple transposition, the principle of symmetry allows the extension of G theory to situations based on complex factorial designs and involving multiple purposes of measurement. We will take as an illustration a design of the type used in surveys of educational achievment or in curriculum evaluations: a sample of pupils (P) is nested in another factor of interest, for example, school districts (D); both of these factors are crossed with a sample of test items (I), which in turn are nested in two crossed factors of classification: instructional objectives (O) and content chapters (C). The data collected with this design may be of interest to several different groups of decision makers. School administrators may be interested in comparing the levels of achievement of different school districts in order to determine how ressources (e.g., funds for remedial activities) should be allocated. Curriculum constructers on th other hand are more likely to be interested in comparisons of achievement scores for different categories of items in order to identify aspects of the instructional materials that should be modified. The principle of symmetry implies application of G theory procedures to test the adequacy of the data collection design for each of the above aims of measurement. As shown in Table 1, the allocation of the variance components for the estimation of the generalizability parameters will be quite different in the two cases. For instance, the variance component for "districts" which constitutes universe-score variance in the first case becomes a component of error variance in the second case. Obviously, it is unlikely that a given design will serve several divergent aims equally well; priority in generalizability analysis should undoubtedly be given to goals determined before data collection occured. But, as the scale and cost of surveys and evaluations increase, the possibility of multiple measurement aims should not be overlooked. In some cases additional aims may emerge after the initial design of the study has been determined, or goals may be defined by data users independently of the considerations that led to data collection. To take the above example: even if the study was initially designed to furnish data to the curriculum constructers (case 2) and a generalizability analysis was conducted accordingly, it could never-the-less be useful to estimate margins of error for comparisons along the facet "districts", if only as a means of warning school administrators against possible misuse of the data in the event that such comparisons are found to be highly unreliable.

Implications of the principle of symmetry

The principle of symmetry has led us to work on extensions of generalizability theory in three major directions.

1. **consideration of all types of facets** : In the initial formulation of G theory by Cronbach and associates (1972), it was assumed that the objects of measurement (persons, or groups of persons) are randomly sampled from an infinite population, whereas the conditions of measurement may be constituted by fixed facets, in addition to random facets. The principle of symmetry led us to the conclusion that the three modes of sampling of factor levels as defined in analysis of variance, i.e., purely random, finite random and fixed, should be considered in the framework of generalizability theory not only with respect to the conditions of measurement (sources of error), but also with respect to the objects of measurement. Work carried out by Brennan (1983) was helpful in pointing out the estimation problems encountered when dealing with fixed and finite random facets. After several revisions and enlargements of our initial proposals, we have now devised a general framework (Cardinet & Allal, 1983) which extends generalizability analysis to designs in which both the objects and the conditions of measurement may be formed by any combination of crossed and nested facets, whatever their mode of sampling.

2. **analysis of multifaceted populations.** We adhere to the idea, expressed by Cronbach *et al* (1972), that the analysis of a multifaceted universe of observation - which entails the definition of the sources of error, the estimation of their contributions to error variance, the search for

means to reduce their effects - is a more interesting and important aspect of generalizability theory than the mere calculation of generalizability coefficients. The principle of symmetry leads us to apply a similar approach to the analysis of the other dimension of a measurement design, i.e., the population of the objects of measurement. In many, perhaps most, instances, the population under study is in reality constituted by several facets, e.g., pupils are nested in classes, in treatments, in socio-economic status, etc., or, symmetrically, if "objectives" are under study, they are defined by several crossed or nested facets of a table of specifications. By analyzing the relative contributions of the sources of variation which enter into the universe-score variance, it is possible to identify undesirable components (sources of "bias") and consider modifications of the design to eliminate or reduce their effects (Cardinet *et al.*, 1981). Thus, our proposals make it possible to extend generalizability theory's procedures of "optimization" to situations in which a multifaceted population is studied in a multifaceted universe of observation.

3. development of a general framework for analysis : Since we were interested from the beginning in making generalizability theory a useful tool for educational researchers in general, and not just an object of pleasure for measurement specialists, we have given considerable importance to the development of a computational framework for generalizability analysis that can be easily applied in virtually any measurement situation. This has meant, among other things, defining a framework that is (a) accessible to any researcher having a basic knowledge of analysis of variance and (b) compatible with the output of current computer programs for ANOVA. These aims have led us to certain choices that differ from those of other specialists on G theory. In our framework of analysis, we clearly separate the first phases of analysis of variance (leading to the estimation of the variance components), from the subsequent phases that are specific to generalizability theory, i.e., the estimation of universe-score variance, error variances and generalizability coefficients. This separation allows us to define a general algorithm for the estimation of generalizability parameters that can be easily applied to any design, however complex.[2] This algorithm eliminates the need for the derivation of formulas for each new design and is particularly advantageous when dealing with situations involving multiple measurement aims. We have retained the well-known Cornfield & Tukey (1956) model for the phases of analysis of variance of our framework, but because this model's definition of variance is not in all cases compatible with the general algorithm we propose for estimating generalizability parameters, we have introduced a correction factor that must be applied whenever a variance component includes a fixed or finite random facet in its primary subscript. The justifications of this procedure are dealt with in more detail in a paper currently being prepared (Cardinet, Tourneur & Allal, in preparation).

Problems of application

Although extensions of generalizability theory through the principle of symmetry, as well as developments introduced by other researchers, provide the basis for a wide range of applications in the field of educational measurement, there have been few attempts thus far to apply the theory in operational research conducted in French-speaking Europe. This can be explained in part by the fact that until the very recent publication of Cardinet and Tourneur's book (1985), there was no comprehensive, well-structured presentation of generalizability theory in the French language. However, even in the Anglo-American context, if one refers to indicators such as the frequency of articles in major journals, or the number of sessions on the topic at annual AERA meetings, it is surprising to note that applications of G theory remain at a relatively low level, compared to what might be expected after nearly 15 years of dissemination and refinement of the theory. In this second part of my presentation, I will mention two factors which, I believe, have hindered application of generalizability theory, at least in the French-speaking European context, but which if modified could permit more widespread use in the future.

1) Conceptions of program evaluation

Theoretically, generalizability theory, as extended by the principle of symmetry, should be a very valuable tool for the planning and implementation of evaluations of educational programs (new curricula, innovations in instructional methods, etc.). Current techniques for generalizability analysis can handle virtually any data collection design that an evaluation team might devise, providing that the data can be quantified. In recent years, in French-speaking Europe and perhaps elsewhere, the adoption of more qualitative, interactive and/or ethnographic methods of program evaluation has often resulted in the abandonment of techniques for quantitative measurement of learning outcomes. Justified criticisms of these techniques have often led to the unjustified conclusion that if quantitative measurement is avoided the problems that it entails (such as how to determine reliability, validity, etc.) can be safely ignored. New conceptions of program evaluation as a process of "enlightened accomodation" based on transactions among various interest groups, as advanced in a recent publication by Cronbach and others (1981), are often seen as having made irrelevant, or unanswerable, questions such as "What are the effects of the program on children's learning, and how accurately can these outcomes be measured?" In the European context, the notion of "accomodation", whether enlightened or not by scientific data corresponds quite closely to the prevailing "naive" theories of how assessments are conducted. Many European researchers are thus quite willing to embrace qualitative methods, in part because they were never really convinced of the usefulness of the techniques of quantitative measurement and analysis which dominated the evaluation models coming from the United States in the 1960s-70s.

More widespread application of generalizability theory is unlikely to occur until specialists in program evaluation develop procedures for <u>integrating</u> the collection and analysis of quantitative data with the application of qualitative methods of investigation. In another recent, individually authored book, Cronbach (1982) points out the need for articulating quantitative and qualitative approaches to program evaluation. He also criticizes the ignorance that often underlies rejection of quantitative techniques of analysis:

> The unsophistocated student of statistics may think that "error variance" is a residual, junk-heap category unworthy of explanation; the experienced investigator knows better. After the F ratios are neatly tabulated, he settles down to figure out what error variance means. A fully quantitative study of a

prespecified hypothesis leaves plenty of room for roving curiosity (Cronbach, 1982, p. 301).

In our view, generalizability theory can perhaps best be seen as a means of "getting inside" the traditional boxes of both "error" and "true" variance in order to find out what are their most important constitutants. But, convincing demonstrations of how to conduct a generalizability analysis in the context of program evaluation are still largely lacking. Even Cronbach devotes little space his 1982 book (pp. 267-68) to a discussion of the uses of G theory in program evaluation.

2) Measurement infrastructure

One of the most potentially useful aspects of generalizability theory is the procedures it provides for using data from an initial study to determine improvements of the design to be used in subsequent research or in decision making, i.e., in Cronbach *et al's* terms, the passage from G study to D study. Although generalizability analysis can be carried out in "one-shot" studies as a means of determining *a posteriori* whether the design already used was in fact adequate for the study's goals, the real usefulness and power of the theory can be best exploited in situations of recurrent monitoring of educational outcomes. Up to now, the school systems of French-speaking Europe have been generally reluctant to introduce the sort of measurement infrastructure -- based on standardized tests or on item banks, and involving repeated testing of samples of children - that would permit extensive use of generalizability theory's procedures of optimization. Investment in this sort of infrastructure has been hindered in part by budgetary constraints but even more by the natural skepticism of education officials with respect to the usefulness of monitoring systems for the framing and implementation of educational policy. Development of such systems is unlikely to occur unless researchers make a greater effort to clarify the benefits that can be derived by policymakers and practitioners. A critical review of what has been gained from these systems in countries where they already exist would be useful.

In summary, although the usefulness of generalizability theory has been demonstrated in several areas of application by studies conducted in the United States (see review by Shavelson & Webb, 1981), in England (Johnson & Bell, 1985) and elsewhere, we are not yet at the stage where the theory has become a basic tool of educational research. Reflexion on obstacles to application takes us to considerations outside the theory itself, but is necessary if one wishes to foster increased use based on a better understanding of the underlying measurement issues and of the contextual factors affecting scientific practice.

Notes

[1] Addresses where Cardinet and Tourneur can be contacted:
- Jean Cardinet, Institut romand de recherches et de documentation pédagogiques, 43 Fbg. de l'Hôpital, 2000 Neuchâtel, Switzerland
- Yvan Tourneur, Faculté des sciences psycho-pédagogiques, Université de l'Etat, 21 place du Parc, 7000 Mons, Belgium

[2] A computerized version of our algorithm which fonctions on Apple IIc has been developed by François Duquesne (in press), and can be obtained from the University of Mons (above address).

References

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

Cardinet, J. & Allal, L. Estimation of generalizability parameters. In L.J. Fyans (Ed.), Generalizability theory: Inferences and practical applications. San Francisco: Jossey-Bass, 1983.

Cardinet, J. & Tourneur, Y. Assurer la mesure: Guide pour les études de généralisabilité. Bern: Peter Lang, 1985.

Cardinet, J. Tourneur, Y. & Allal, L. The symmetry of generalizability theory: Applications to educational measurement. Journal of Educational Measurement, 1976, 13, 119-135.

Cardinet, J. Tourneur, Y. & Allal, L. Extension of generalizability theory and its applications in educational measurement. Journal of Educational Measurement, 1981, 18, 183-204; and 1982, 19, 331-332.

Cardinet, J. Tourneur, Y. & Allal, L. Definition of the components of generalizability parameters, in preparation.

Cornfield, J. & Tukey, J. Average values of mean squares in factorials. Annals of mathematical statistics, 1956, 27, 907-949.

Cronbach, L.J. Designing evaluations of educational and social programs. San Francisco: Jossey-Bass, 1982.

Cronbach, L.J., Rajaratman, N. & Gleser, G. Theory of generalizability: A liberalization of reliability theory. British Journal of Mathematical and Statistical Psychology, 1963, 16, 137-163.

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

Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik, R.C., Phillips, D.C., Walker, D.F. & Weiner, S.S. Toward reform of program evaluation. San Francisco: Jossey-Bass, 1981.

Duquesne, F. Développement sur micro-ordinateur d'un programme pour l'étude de la généralisabilité. Scientia Paedagogica Experimentalis, vol. 23, no. 1, in press.

Johnson, S. & Bell, J.F. Evaluating and predicting survey effeciency using generalizability theory. Journal of Educational Measurement, 1985, 22, 107-119.

Shavelson, R.J. & Webb, N.M. Generalizability theory: 1973-1980. British Journal of Mathematical and Statistical Psychology, 1981, 34, 33-66.

Table i : Allocation of the variance component estimates for the estimation of generalizability parameters in a study with multiple measurement aims

Design : (P : D) x (I : O C), where P, D & I are purely random and O & C are fixed facets

	Aim of measurement	
Generalizability parameters	Differentiation of districts	Differentiation of objectives
universe-score variance*	D	O
relative error variance	P: D (P: D) x (I: OC) D x (I: OC)	I: OC (P: D) x (I: OC) D x (I: OC) O x D O x P: D
absolute error variance	components of relative error plus : I: OC	components of relative error plus : D P: D

*In the analysis framework of Cardinet, Tourneur & Allal (1981/82), the term "differentiation variance" is used.