DOCUMENT RESUME

ED 146 215 TM 006 608

AUTHOR Brennan, Robert L.
TITLE Generalizability Analyses: Principles and Procedures. ACT Technical Bulletin No. 26.
INSTITUTION American Coll. Testing Program, Iowa City, Iowa. Research and Development Div.
REPORT NO ACT-TB-26
PUB DATE Sep 77
NOTE 98p.; Revision of a paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS *Analysis of Variance; *Mathematical Models; Measurement; *Measurement Techniques; Orthogonal Rotation; Problems; *Reliability; Research Methodology; Sampling; Standard Error of Measurement; *Statistical Analysis; Test Interpretation; True Scores
IDENTIFIERS *Generalizability Theory

ABSTRACT

Rules, procedures, and algorithms intended to aid researchers and practitioners in the application of generalizability theory to a broad range of measurement problems are presented. Two examples of measurement research are G studies, which examine the dependability of some general measurement procedure; and D studies, which provide the data for substantive decisions. Major emphasis is given to the estimation of G study variance components, and to the estimation and use of D study variance components for different objects of measurement and different universes of generalization. D studies in which the universe of generalization contains facets that are either fixed or essentially infinite are discussed, as well as D studies that involve sampling from a finite universe. A notational system is introduced to facilitate the discussion; and each rule, procedure or algorithm is illustrated using designs that involve varying types and degrees of complexity. (Author/MV)

ACT TECHNICAL BULLETIN NO. 26

Generalizability Analyses: Principles and Procedures

by

Robert L. Brennan

The American College Testing Program

The Research and Development Division

The American College Testing Program

P. O. Box 168, Iowa City, Iowa 52240

September, 1977

Abstract

This paper presents "rules," procedures, and algorithms intended to aid researchers and practitioners in the application of generalizability theory to a broad range of measurement problems. Major emphasis is given to the estimation of G study variance components, and to the estimation and use of D study variance components for different objects of measurement and different universes of generalization. Consideration is given to D studies in which the universe of generalization contains facets that are either fixed or essentially infinite, as well as D studies that involve sampling from a finite universe. A notational system is introduced to facilitate the discussion; and each "rule," procedure, or algorithm is illustrated using designs that involve varying types and degrees of complexity.

Table of Contents

Glossary of Symbols

Symbol	Definition
x	"Crossed with."
:	"Nested within."
p, c, i, s, t, o	A facet; or a specific condition of a facet.
P, C, I, S, T, O	A set of conditions for a facet; or the sample mean for a set of conditions for a facet.
α	Generic symbols for <u>any</u> component or source of variance in a G or D study.
γ	Facet that serves as object of measurement for some D study.
$\mu_{\alpha}^{\sim}$	Score effect for the component α.
μ_{α}	Mean score for the component α.
$\overline{X}_{\alpha}^{\sim}$	Observed score analogue of $\mu_{\alpha}^{\sim}$.
$\overline{X}_{\alpha}$	Observed score analogue of μ_{α}.
e	Random error.
$\mathcal{E}$	Expected value.
$\sigma^2(\alpha)$	Random effects variance component for α (given sampling from an <u>infinite</u> universe).
n	G study sample size for a facet.
N	Size of universe of admissible observations for a facet in the G study.

Symbol	Definition
$\underline{n}'$	D study sample size for a facet.
$\underline{N}'$	Size of universe of generalization for a facet in the D study.
$\underline{MS}(\alpha)$	Mean square for α.
$\underline{EMS}(\alpha)$	Expected mean square for α.
$\sigma^2(\tau)$	Universe score variance for a D study.
$\sigma^2(\Delta)$	Variance of differences between observed scores and universe scores.
$\sigma^2(\delta)$	Variance of differences between observed deviation scores and universe scores expressed in deviation form.
$\sigma^2(\varepsilon)$	Variance of differences between universe scores and regression estimates of universe scores.
$\mathcal{E}\sigma^2(X)$	Expected observed score variance.
$\mathcal{E}\rho^2$	Generalizability coefficient.
$\underline{V}$	Main effect index in γ.
$\underline{F}$	Set of facets that are fixed in a D study.
$\underline{R}$	Set of facets that are randomly sampled in a D study.
$\underline{f}(\alpha)$	See Equation 4.
$\underline{d}(\alpha\|\gamma)$	See Equation 13.

Introduction

Classical test theory provides a very simple structural model of the relationship between observed, true, and error scores. However, the simplicity of the model necessitates some rather restrictive assumptions if the model is to be applied to real data. Generalizability theory liberalizes and extends classical test theory in several important respects. For example, the theory of generalizability does not necessitate the classical test theory assumption of "parallel" tests; rather, generalizability theory employs the weaker assumption of "randomly parallel" tests. Also, classical test theory assumes that errors of measurement are sampled from an undifferentiated univariate distribution. By contrast, generalizability theory allows for the existence of multiple types and sources of error through the application of analysis of variance procedures, or, more specifically, through the application of the general linear model to the dependability of measurement. Consequently, generalizability theory is applicable to a broad range of testing and evaluation studies that arise in education and psychology.

Background and Terminology

The basic theoretical foundation for generalizability theory can be found in papers by Cronbach, Rajaratnam, and Gleser (1963) and Gleser, Cronbach, and Rajaratnam (1965). These papers were followed by an extensive explication of generalizability theory in a monograph by Cronbach, Gleser, Nanda, and Rajaratnam (1972) entitled, The Dependability of Behavioral Measurements.

However, the use of analysis of variance approaches to reliability issues did not begin with the publications of Cronbach and his colleagues, even though it is they who have most clearly and completely formulated reliability issues in analysis of variance terms. Over 35 years ago Burt (1936), Hoyt (1941), and Jackson and Ferguson (1941), discussed analysis of variance approaches to reliability. Subsequent contributions were made by Alexander (1947), Ebel (1951), Finlayson (1951), Loveland (1952), and Burt (1955). Also, Lindquist (1953), in the last chapter of his experimental design text, discussed in considerable detail the estimation of variance components in reliability studies. In fact, in several respects work by Burt and Lindquist appears to anticipate the development of generalizability theory. Additional evidence of the role of analysis of variance in reliability issues can be seen in the work of Webster (1960) and Medley and Mitzel (1963) not long before the original publication by Cronbach et al. (1963) of the theory of generalizability.

Although generalizability theory borrows its statistical models and research designs from analysis of variance, there are some changes in emphasis, terminology, and interpretation. For example, in analysis of variance, the magnitudes of variance components sometimes receive direct attention (see, for example, Vaughan & Corballis, 1969), but the ultimate goal is usually a test (or tests) of statistical significance. In generalizability theory interest is focused primarily on the magnitude of variance components and, to some extent, generalizability coefficients. Tests of statistical significance receive less direct emphasis.

In generalizability theory, any observation on some object of measurement (e.g., school, class, student) is assumed to be sampled from a universe of observations. While universe and population are logically equivalent terms, here the word population is reserved for the object of measurement, and the word universe is reserved for the conditions under which observations are made. Any observations from the universe can be characterized by the conditions under which the observation is made. The set of all possible conditions of a particular kind is called a facet.

Generalizability theory also emphasizes the distinction between G studies, which examine the dependability of some general measurement procedure, and D studies, which provide the data for substantive decisions (Rajaratnam, 1960). "For example, the published estimates of reliability for a college aptitude test are based on a G study. College personnel officers employ these estimates to judge the accuracy of data they collect on their own applicants (D study)" (Cronbach et al., 1972, p. 16). The primary purpose of the G study is to estimate components of variance, which may then be used in a variety of D studies. The G study and the D study may be the same study, or they may be different studies using the same design. Generalizably, however, G studies are most useful when they employ complex designs and large sample sizes to provide stable estimates of as many variance components as possible.

Based on the difference between a G study and a D study, Cronbach et al. (1972) make a further distinction between the universe of admissible observations and the universe of generalization.

> The test developer or other investigator who carried out a G study takes certain facets into consideration and, with respect to each facet, considers a certain range of conditions. The observations encompassed by the possible combinations of conditions that the G study represents is called the universe of admissible observations. We may also speak of the universe of admissible conditions of a certain facet. A decision maker, applying essentially the same measuring technique, proposes to generalize to some universe of conditions all of which he sees as eliciting samples of the same information. We refer to that as the universe of generalization. The G study can serve this decision maker only if its universe of admissible conditions is identical to or includes the proposed universe of generalization. Different decision makers may propose different universes of generalization. A G study that defines the universe of admissible observations broadly, encompassing all the likely universes of generalization, will be useful to various decision makers. (p. 20)

Overview

In this paper "rules," procedures, and algorithms are presented that involve a notational system, analysis of variance considerations, G studies, and D studies. In addition, all "rules," procedures, and algorithms are illustrated using designs that involve varying levels and types of complexity.

The notational system used here differs, in some respects, from that used in the Cronbach et al. (1972) monograph on generalizability theory. The primary difference is that the notational system for variance components used in this paper does not necessitate specifically reporting the effects that are confounded in a design that involves nesting. Nevertheless, this notational system does implicitly "carry the meaning" of a nested component. In most other respects, the notation used by Cronbach and his colleagues has been maintained or minimally altered.

The terminology used in some analysis of variance literature is not always the same as the terminology employed by Cronbach and his colleagues in discussing generalizability theory. For example, the word "facet" in generalizability theory has approximately the same connotation as "main effect" in much of the analysis of variance literature. Also, the word "component" in Cronbach et al. (1972) is basically synonymous with the word "effect" in some analysis of variance literature. One of the purposes of this paper is to help practitioners familiar with analysis of variance literature to understand and apply generalizability theory. Therefore, some terminological compromises are made here. Generally, the terminology employed is that of Cronbach et al. (1972); but exceptions do occur, especially in initial sections that primarily treat analysis of variance considerations. When terminological ambiguities arise an attempt is made to resolve them, or at least clarify them.

The major portion of this paper is devoted to a consideration of "rules," procedures, and algorithms for performing G studies and D studies. Particular emphasis is given to the estimation of G study variance components, and to the estimation and use of D study variance components for different objects of measurement and different universes of generalization. Most of the discussion treats D studies in which the universe of generalization contains facets that are either fixed or essentially infinite. However, consideration is also given to D studies that involve sampling from a finite universe of generalization.

There are some restrictions placed upon the treatment of generalizability analysis in this paper. In particular, with minor exceptions, only orthogonal analysis of variance designs are considered; i.e., designs that do not involve missing data and/or unequal size subgroups. Also, all designs and studies involve only one dependent variable; i.e., this paper treats univariate generalizability theory, as opposed to multivariate generalizability theory (see Cronbach et al., 1972, Chapter 10). Finally, the "rules," algorithms, and procedures are not intended to cover, in depth or breadth, the extensive treatment of generalizability theory provided by Cronbach and his colleagues. Rather, this paper is intended to provide researchers and practitioners with a set of procedures to facilitate the application of generalizability theory to a broad range of measurement problems.

A Notational System and Analysis of Variance Considerations for G Studies

The first steps in performing a G study involve the usual initial procedures for an analysis of variance; namely, defining the model and determining sums of squares, degrees of freedom, and mean squares for each of the effects in the G study design. These issues are usually treated in experimental design texts in the context of specific designs. Here, rules and algorithms are provided that are applicable to a large class of orthogonal, or balanced, designs.

Notation for ANOVA Designs

Using the symbols "x" to mean "crossed with" and ":" to mean "nested within," most common analysis of variance designs can be represented by a suitable sequence of effect indices and symbols. In this paper, five different designs will be used for illustrative purposes: p x i, p x i x o, p x (i:s), (p:c) x i, and (p:c) x (i:s:t). [In Cronbach et al. (1972), p x i x o is called Design VII, p x (i:s) is Design V-A, and (p:c) x i is essentially Design V-B.] The indices in these designs can be interpreted as referring to a person (p), a class (c), an item (i), a subtest (s), an occasion (o), and a test (t). For example, (p:c) x i can be interpreted as meaning that persons are nested within classes, and both persons and classes are crossed with items. The use of these specific identifying words for each index is maintained throughout this paper; however, it is the nature of the design that is under consideration--not the names associated with the indices.

These designs have been chosen for two reasons. First, they involve different types and degrees of complexities in applying the "rules," and procedures which will be presented. Second, these designs are typical of the kinds of designs that do occur in testing and evaluation studies. Most of the classical results from test theory come from a consideration of the basic design for persons crossed with items, $p \times i$. The design $p \times i \times o$, which Cronbach et al. (1972) treat in great detail, is a simple extension of this basic design. In many realistic situations, however, some degree of nesting is present. For example, it is very common for items to be nested within subtests, as in the design $p \times (i{:}s)$. Also, in many testing studies, persons are nested within classes, as in the design $(p{:}c) \times i$. Finally, an extensive testing study may involve considerable nesting, as in the design $(p{:}c) \times (i{:}s{:}t)$.[1]

Main Effects and Interaction Effects

Figure 1 provides a Venn diagram representation for each of the illustrative designs. In these Venn diagrams, the mean square for a main effect is represented by a circle (of any size), and the mean square for an interaction effect is represented by the intersection of two or more circles. (The words "effect" and "component" are basically synonymous terms; however, we will use the term "effect" here because the phrases "main component" and "interaction component" are rare in ANOVA literature.)

Insert Figure 1 about here

Notation for Main Effects. A main effect can be represented by

$$\left\{\begin{matrix}\text{main}\\ \text{effect}\\ \text{index}\end{matrix}\right\} : \left\{\begin{matrix}\text{first}\\ \text{nesting}\\ \text{index(es)}\end{matrix}\right\} : \left\{\begin{matrix}\text{second}\\ \text{nesting}\\ \text{index(es)}\end{matrix}\right\} : \ldots ;$$

If the main effect is not a nested main effect, then it can be represented by the main effect index, only.

For example, in the design $p \times i$, the main effect for persons is denoted p, and the main effect for items is denoted i. In the design $(p{:}c) \times i$, the (nested) main effect for persons is p:c, where the main effect index is p, and the nesting index is c. Similarly, in the design (p:c) x (i:s:t), the (nested) main effect for items is i:s:t, the (nested) main effect for subtests is s:t, and the main effect for tests is t. In general, the number of main effects is equal to the number of indices in the symbolic representation of a design.[2]

In some monographs and textbooks, main effects are called treatments, factors, or facets. However, not all effects are easily interpretable as treatments, and the word "factor" is apt to cause confusion with factor analysis. Here the terms "main effect" and "facet" are used synonymously, unless otherwise noted.

Notation for Interaction Effects. Each interaction effect can be represented as a combination of main effects in the following manner:

$$\left\{\begin{matrix}\text{Combination of}\\ \text{Main Effect}\\ \text{Indexes}\end{matrix}\right\} : \left\{\begin{matrix}\text{Combination of}\\ \text{First Nesting}\\ \text{Indexes}\end{matrix}\right\} : \left\{\begin{matrix}\text{Combination of}\\ \text{Second Nesting}\\ \text{Indexes}\end{matrix}\right\} : \ldots$$

subject to the constraint that no index may appear more than once in any interaction effect.

Insert Tables 1 - 5 about here

Tables 1 - 5 list the main effects and interaction effects for each of the five illustrative designs using the notation defined above. Consider, for example, the design (p:c) x (i:s:t) in Table 5. The interaction of c and t is simply ct. The interaction of c and s:t is cs:t; that is, combinations of cs are nested within t (see Figure 1). Similarly, the interaction of p:c and i:s:t is pi:cs:t; that is, combinations of pi are nested within combinations of cs, which, in turn, are nested within t. Also, note that the interactions of p:c and c would be pc:c, but this possibility is ruled out by two occurrences of the index c.

Nested Effects and Confounding. Cronbach et al. (1972) usually use a sequence of confounded effects to identify any main effect or interaction effect that involves nesting. For example, if data for the design (p:c) x i were analyzed as if the design were the completely crossed design p x c x i, then the effects would be p, c, i, pc, pi, ci, pci; but some of these effects would be confounded. In particular, the main effect p:c, in the design (p:c) x i, represents the confounding of two of the effects, p and pc, from the design p x c x i. Similarly, the interaction effect pi:c represents the confounding of pi and pic.

Whenever a design involves nesting, there is some degree of confounding. In designs with more than one nested main effect, or more than one level of nesting, the representation of a nested effect by its confounded effects leads to considerable complexity. This is one reason for using the nesting operator in representing effects. Nevertheless, it is frequently useful to know which effects are confounded in a nested effect.

Using the notation introduced above, for any nested effect, the effects that are confounded are all combinations of indices in the effect that include the main effect index (or indexes). For example, in the design (p:c) x (i:s:t), the effect s:t represents the confounding of s and st. Similarly, the effect i:s:t represents the confounding of i, is, it, and ist; and the effect pi:sc:t represents the confounding of pi, pic, pis, pit, pisc, pict, pist, and pisct.

In general, for any nested effect, the number of effects that are confounded is:

$$2 \ \underline{\text{Exp}} \left\{ \begin{array}{c} \text{number of nesting} \\ \text{indices} \\ \text{in component} \end{array} \right\}.$$

For example, in the design (p:c) x (i:s:t), the effect i:s:t has two nesting indices (s and t); and, therefore, this effect has [2 Exp (2)], or four confounded effects. Similarly, the effect pi:cs:t has [2 Exp (3)], or eight confounded effects.

Degrees of Freedom. For any effect (main effect or interaction effect) that is not nested, the degrees of freedom are the product of the $(n - 1)$'s for the indexes in the effect, where n is the G study sample size associated with an index. For any nested effect, the degrees of freedom are:

$$\left\{ \begin{array}{c} \text{Product of } n\text{'s} \\ \text{for nesting indexes} \end{array} \right\} \times \left\{ \begin{array}{c} \text{Product of } (n - 1)\text{'s} \\ \text{for main effect indexes} \end{array} \right\}.$$

Degrees of freedom for the effects in each of the five illustrative designs are provided in Tables 1 - 5. For example, for the design (p:c) x (i:s:t) in Table 5, the degrees of freedom for the main effect s:t are $n_t(n_s - 1)$. Also, for the main effect i:s:t, the degrees of freedom are $n_s n_t (n_i - 1)$, and for the interaction effect pi:cs:t, the degrees of freedom are $n_c n_s n_t (n_p - 1)(n_i - 1)$.

Structural Models

Consider the design $(\underline{p}:\underline{c}) \times \underline{i}$. For this design, the observed score for person $\underline{p}$ in class $\underline{c}$ on item $\underline{i}$ can be represented by the structural model:

$$X_{\underline{pi}:\underline{c}} = \mu + \mu_{\underline{p}:\underline{c}}^{\sim} + \mu_{\underline{c}}^{\sim} + \mu_{\underline{i}}^{\sim} + \mu_{\underline{ci}}^{\sim} + \mu_{\underline{pi}:\underline{c}}^{\sim} + \underline{e} \; ; \qquad (1)$$

where μ = grand mean in the universe;

$\mu_{\underline{p}:\underline{c}}^{\sim}$ = effect for person $\underline{p}$ in class $\underline{c}$;

$\mu_{\underline{c}}^{\sim}$ = effect for class $\underline{c}$;

$\mu_{\underline{i}}^{\sim}$ = effect for item $\underline{i}$;

$\mu_{\underline{ci}}^{\sim}$ = effect for interaction of class $\underline{c}$ and item $\underline{i}$;

$\mu_{\underline{pi}:\underline{c}}^{\sim}$ = effect for interaction of person $\underline{p}$ in class $\underline{c}$ on item $\underline{i}$; and

$\underline{e}$ = random error

(Note that the structural model for each of the five illustrative designs is provided in the footnotes to Tables 1 - 5.)

Score Effects. Equation 1 provides a decomposition of the observed score $X_{\underline{pi}:\underline{c}}$ in terms of independently estimable effects which we will call score effects. Specifically, we will say that $\mu_{\alpha}^{\sim}$ is the score effect for the component α. Since the words "effect" and "component" are basically synonymous, one could also speak of the score component for the effect α; however, the author generally prefers the former verbal description because it avoids some verbal ambiguities in subsequent sections.

The usual assumptions concerning score effects, such as those represented in Equation 1, are well documented in the literature and in experimental design texts. First, each effect is assumed to be independent of every other effect. Second, in order to make the estimates of the effects unique, the expected value of each effect over any of its subscripts is set equal to zero. Consider, for example, the effect $\mu_{\underline{c}}\sim$ in Equation 1, and suppose we take a sample of $\underline{n}_{\underline{c}}$ classes from a universe of $\underline{N}_{\underline{c}}$ classes. The universe of classes is called the universe of admissible observations for the class facet. The second assumption implies that the sum of $\mu_{\underline{c}}\sim$ over the universe of $\underline{N}_{\underline{c}}$ classes is constrained to be zero, and the sum of the estimates of $\mu_{\underline{c}}\sim$ over the sample of $\underline{n}_{\underline{c}}$ classes is constrained to be zero. However, it is not necessarily true that the sum of $\mu_{\underline{c}}\sim$ over the sample of $\underline{n}_{\underline{c}}$ classes is zero. Finally, note that Equation 1 involves no assumptions about the distributional form of the errors.

Mean Scores. Associated with each score effect is a unique mean score. For any component α, the mean score is the expected value of the observed score over all indices not contained in α. Note that for any facet (i.e., index) the expected value is taken over the universe of admissible observations, and the symbol "$\mathcal{E}$" is used to define expectation. For example, from Equation 1:

$$\mathcal{E}_{\underline{i}} X_{\underline{pi:c}} = \mu + \mu_{\underline{p:c}}\sim + \mu_{\underline{c}}\sim = \mu_{\underline{p:c}}$$

That is, $\mu_{\underline{p:c}}$ is the expected value of $X_{\underline{pi:c}}$ over all items in the universe of admissible observations.

Note, in particular, the distinction between $\mu_{\underline{p:c}}\sim$ (score effect) and $\mu_{\underline{p:c}}$ (mean score). Notationally, a score effect always has a tilda ($\sim$) associated with it, and a mean score does not. Also, the term "mean score" in this context should <u>not</u> be confused with an <u>observed</u> mean score for a <u>sample</u>, or a universe score for a particular D-study, both of which are discussed in considerable detail later.

Using this notational system it is easy to express any mean score in terms of score effects. In general, for the component α,

$$\mu_\alpha = \mu + \left\{ \begin{array}{c} \text{Sum of score effects} \\ \text{for all components that consist} \\ \text{solely of indices in } \alpha \end{array} \right\} \quad (2)$$

For each component in the design (<u>p</u>:<u>c</u>) x <u>i</u>, Table 6 reports equations for mean scores in terms of score effects. Conversely, score effects can be expressed in terms of mean scores.

Insert Table 6 about here

<u>Algorithm 1</u>: <u>Expressing a Score Effect in Terms of Mean Scores</u>.
The following algorithm can be used with any design to express a score effect as a combination of mean scores. Let α be a component with <u>t</u> nesting indices and <u>m</u> main effect indices; then $\mu_{\alpha}\sim$, the score effect associated with α, is equal to:

<u>Step 0</u>: μ_α

<u>Step 1</u>: <u>Minus the mean scores</u> for components that consist of the <u>t</u> nesting indices and <u>m</u> - 1 of the main effect indices;

<u>Step 2</u>: <u>Plus</u> the <u>mean scores</u> for components that consist of the <u>t</u> nesting indices and <u>m</u> - 2 of the main effect indices;

Step i: Plus (if i is even) or Minus (if i is odd) the mean scores for components that consist of the t nesting indices and $m - i$ of the main effect indices;

.
.
.

The algorithm terminates with Step m; that is, with the mean score for the component containing only the t nesting indices. If there are no nesting indices in the component α, then it follows that Step m results in adding or subtracting μ.

Consider, for example, the component $pi{:}c$ in the design $(p{:}c) \times i$. This component has a single nesting index, c, and two main effect indexes, p and i. Step 1 in the algorithm results in subtracting $\mu_{p:c}$ and μ_{ci} from $\mu_{pi:c}$, because both $p{:}c$ and ci contain the nesting index, c, and $2 - 1 = 1$ main effect index. Step 2 results in adding μ_c to the result of Step 1, because c is the component that contains the nesting index, c, and $1 - 1 = 0$ main effect indexes. Therefore,

$$\mu_{pi:c}^{\sim} = \mu_{pi:c} - \mu_{p:c} - \mu_{ci} + \mu_{c}$$

For each component in the design $(p{:}c) \times i$, Table 6 reports equations for score effects in terms of mean scores.

Sums of Squares[3]

For each component α, the mean score μ_α has an observed score analogue, which we denote $\overline{X}_\alpha$. Similarly, $\mu_\alpha^{\sim}$ has an observed score analogue $\overline{X}_\alpha^{\sim}$. For example, in the design $(p{:}c) \times i$, $\overline{X}_c$ is the observed mean score over the sample of persons and items in class c, and $\overline{X}_c^{\sim}$ is the observed score effect for class $\underline{c}$. The relationship between μ_α and $\mu_{\alpha}\sim$ is identical to the relationship between $\overline{X}_\alpha$ and $\overline{X}_{\alpha}\sim$. That is, Algorithm 1 and Equation 2 are applicable to observed mean scores and observed score effects through replacement of μ_α, $\mu_\alpha\sim$, and μ by $\overline{X}_\alpha$, $\overline{X}_\alpha\sim$, and $\overline{X}$, respectively. In this terminology and notational system the "sums of squares" calculated in the performance of an analysis of variance are, more correctly, the "sums of squares" for observed score effects. There are two well-known, algebraically identical procedures for determining the sums of squares for observed score effects. The first procedure entails a direct application of the observed score effects. See, for example, the last column of Table 7 for the design ($\underline{p}$:$\underline{c}$) x $\underline{i}$.

Insert Table 7 about here

This procedure is, at least conceptually, the simpler of the two. However, a computationally easier procedure involves using the sums of squares for observed mean scores (to be distinguished from the sums of squares for observed score effects). Kirk (1968), among others, uses this second procedure extensively. In general, the sum of squares for observed mean scores, for the component α, is

$$[\overline{X}_\alpha] = \underline{f}(\alpha)\Sigma\ \overline{X}_\alpha^2 \quad ; \tag{3}$$

where the summation is taken over all indices in α, and $\underline{f}(\alpha)$ is the number of observations used to calculate the mean for any one of the levels of α. Specifically,

$$\underline{f}(\alpha) = \begin{cases} \text{1, if } \alpha \text{ includes all indices in the design; and, otherwise,} \\ \text{the product of the G study sample sizes } (\underline{n}\text{'s}) \text{ for the} \\ \text{indices } \underline{\text{not}} \text{ included in } \alpha. \end{cases} \tag{4}$$

The quantities $[\overline{X}_\alpha]$ for each of the components in the design (p:c) x i are reported in Table 7. Table 7 also provides the sums of squares for observed score effects expressed in terms of the quantities $[\overline{X}_\alpha]$. Note that the above terminology directly implies that $[\overline{X}_\alpha]\sim$ is the sum of squares for observed score effects, for the component α. Furthermore, Algorithm 1 and Equation 2 are applicable to the quantities $[\overline{X}_\alpha]$ and $[\overline{X}_\alpha]\sim$ through replacement of μ_α, $\mu_\alpha\sim$, and μ by $[\overline{X}_\alpha]$, $[\overline{X}_\alpha]\sim$, and $[\overline{X}]$, respectively.

From the above development it follows that the sum of squares (for observed score effects) associated with the component α is:

$$\underline{SS}(\alpha) = [\overline{X}_\alpha]\sim \text{, or} \tag{5}$$

$$\underline{SS}(\alpha) = \underline{f}(\alpha)\Sigma(\overline{X}_\alpha\sim)^2 , \tag{6}$$

where $\underline{f}(\alpha)$ and Σ have the same interpretation in Equation 6 that they have in Equation 3.

Equations 5 and 6 are applicable to calculating sums of squares associated with any component, whether or not it is nested. In addition, for any nested component, the sum of squares can be obtained by adding the sums of squares for the confounded effects. For example, in the design (p:c) x i (Figure 1 and Table 4), the component pi:c represents the confounded effects pi and pic, which are independently estimable in the design p x c x i. Therefore, to obtain the sums of squares for pi:c, the data can be treated as if they came from the design p x c x i, and the addition of the sums of squares associated with pi and pic results in the sums of squares for pi:c. This is a very useful procedure for performing a G-study having nested components, especially when available computer programs cannot directly accommodate nested designs.

G Study Considerations and the Estimation of Variance Components for the Random Effects Model

Whereas, classical analysis of variance procedures typically emphasize F-tests, generalizability theory emphasizes the estimation of variance components. According to the most recent edition of Standards for Educational & Psychological Tests (APA, 1974): the "estimation of clearly labeled components of score variance is the most informative outcome of a reliability study, both for the test developer wishing to improve the reliability of his instrument and for the user desiring to interpret test scores with maximum understanding" (p. 49).

Variance Components--Notation and Terminology

The variance component associated with the component α is, by definition, the variance of the universe score effect for the component α. Consider, for example, the design $p \times i$, which can be represented as:

$$X_{pi} = \mu + \mu_p^{\sim} + \mu_i^{\sim} + \mu_{pi}^{\sim} + e \text{;} \qquad (7)$$

where μ = grand mean in the universe,

$\mu_p^{\sim}$ = effect for person p,

$\mu_i^{\sim}$ = effect for item i,

$\mu_{pi}^{\sim}$ = effect for the interaction of person p with item i, and

e = random error.

The variance for the component p is denoted $\sigma^2(\mu_p^{\sim})$, which is abbreviated $\sigma^2(p)$. This is, $\sigma^2(p)$ is the variance of $\mu_p^{\sim}$, over all persons in the universe (or population) of admissible observations.

Similarly, $\sigma^2(\underline{pi})$ is the variance of the component $\underline{pi}$; or, more specifically, the variance of $\mu_{\underline{pi}}$ in the universe. However, $\sigma^2(\underline{pi})$ is confounded with random error variance. To account for this confounding, Cronbach et al. (1972) denote this variance component $\sigma^2(\underline{pi},\underline{e})$. Using the notational system discussed above, the component that consists of all indices in the design is always confounded with random error. Therefore, it is not necessary to explicitly indicate this confounding in the notation for variance components, and we will not do so here. As another example, consider the component $\underline{pi}{:}\underline{c}$ in the design ($\underline{p}{:}\underline{c}$) x $\underline{i}$ (see Equation 1 and Table 4). Here, the variance of this component is denoted $\sigma^2(\underline{pi}{:}\underline{c})$. Cronbach et al. (1972), however, represent this variance component by $\sigma^2(\underline{pi},\underline{pic},\underline{e})$, which explicitly indicates both the confounding resulting from the nesting of $\underline{pi}$ within $\underline{c}$, and the confounding of random error with $\underline{pi}{:}\underline{c}$.

For the design $\underline{p}$ x $\underline{i}$ (see Equation 7), the variance of $X_{\underline{pi}}$ over all persons and items is:

$$\sigma^2(X_{\underline{pi}}) = \mathop{\mathcal{E}}_{\underline{p},\underline{i}} (X_{\underline{pi}} - \mu)^2 \tag{8}$$

$$= \sigma^2(\underline{p}) + \sigma^2(\underline{i}) + \sigma^2(\underline{pi}) \,. \tag{9}$$

Since the variance components in Equation 9 are non-negative and independent, none of them can be greater than the maximum value of $\sigma^2(X_{\underline{pi}})$. If, for example, items are scored (0,1), then no variance component can be greater than 0.25, the maximum value of $\sigma^2(X_{\underline{pi}})$. In effect, each variance component in Equation 9 represents that part of $\sigma^2(X_{pi})$ uniquely attributable to the component. (This, of course, is not true for mean squares.) Furthermore, since X_{pi} is the observed score for a single person and a single item, the variance components in Equation 9 are for a single person, a single item, and a single person-item combination, respectively. It is both usual and highly advisable to report G study variance components for single observations based on sampling one condition of each facet. These G study variance components can be used easily in subsequent D studies that involve sampling any number of conditions of each facet.

There are several procedures that might be used to estimate variance components. For example, Cornfield and Tukey (1956), Cronbach et al. (19??), Millman and Glass (1967), and most experimental design texts (e.g., Kirk, 1968) discuss procedures for obtaining the expected value of mean squares in terms of variance components. The resulting set of equations can be solved to express estimated variance components in terms of mean squares (see Endler, 1966). Also, using these procedures, expected mean squares and estimated variance components can be obtained for models other than the random effects model. These procedures, however, are often more general and more complicated than the requirements of a generalizability analysis demand. For example, usually a G study does not directly require expected mean squares. Furthermore, it is usually best to perform a G study under the assumptions of a random effects model.

The terms "random," "fixed," and "mixed effects" are common in the context of analysis of variance, but they have been used less frequently in the context of generalizability theory. In the usual terminology of generalizability theory, a facet is random if conditions of the facet are randomly sampled from an infinite (or essentially infinite) universe of possible conditions for the facet.

Notationally, if $\underline{n}$ is the sample size for some facet, and $\underline{N}$ is the size of the universe for the facet, then the facet is random if $\underline{n} < N \rightarrow \infty$. A facet is fixed if $\underline{n} = \underline{N}$. If all facets are random, then the design is a random effects design. Similarly, if all facets are fixed, then the design is a fixed effects design. If some facets are fixed and some random, then the design is a mixed effects design. For a G study it is almost always best to estimate variance components under the assumptions of a random effects model. The variance components resulting from a random effects analysis of G study data can be used easily in subsequent D studies that employ random, fixed, or mixed models. The only important exception to this general rule involves random sampling from a finite universe, which is treated later.

Algorithm 2: Estimation of Variance Components for Random Effects Models.

Another procedure for estimating variance components entails the use of Venn diagrams (see Cronbach et al., 1972). This procedure (which is illustrated later) is quite useful when the random effects model is employed in a design that is relatively uncomplicated. However, the Venn diagram approach is rather difficult to use with more complicated designs. The following algorithm reflects the Venn diagram approach, but it does not require the use of Venn diagrams.

Assume that α is some component consisting of $\underline{k}$ indices. Here, it does not matter whether an index in α is nested or not. In general, the estimated value of the variance of the component α, for the random effects model is:

$$\hat{\sigma}^2(\alpha) = \frac{1}{\underline{f}(\alpha)} \begin{bmatrix} \text{some combination} \\ \text{of mean squares} \end{bmatrix}$$

where $\underline{f}(\alpha)$ has been defined in Equation 4, and the appropriate combination of mean squares is:

<u>Step</u> <u>0</u>: <u>MS</u>(α)

<u>Step</u> <u>1</u>: <u>Minus</u> the mean squares for all components that consist of the <u>k</u> indices in α and exactly <u>one</u> additional index (call the set of additional indices <u>A</u>);

<u>Step</u> <u>2</u>: <u>Plus</u> the mean squares for all components that consist of the <u>k</u> indices in α and any <u>two</u> of the <u>A</u> indices;

<u>Step</u> <u>3</u>: <u>Minus</u> the mean squares for all components that consist of the <u>k</u> indices in α and any <u>three</u> of the <u>A</u> indices;

.
.
.

<u>Step</u> <u>i</u>: <u>Plus</u> (if <u>i</u> is even) or <u>Minus</u> (if <u>i</u> is odd) the mean squares for all components that consist of the <u>k</u> indices in α and any <u>i</u> of the <u>A</u> indices;

.
.
.

The algorithm terminates when a step results in no mean squares added or subtracted.

For some components no steps are required. For example, the estimated variance of the component that contains all indices in the design is simply the mean square of that component. Also, except in quite complicated designs, it is rare that more than two steps are required to obtain the estimated variance component in terms of mean squares. The actual number of steps required for any component in any design is $q - k$, where q is the total number of indices in the design.

Tables 1 - 5 provide equations for estimating the variance of the components in each of the five illustrative designs, assuming the random effects model. Consider, for example, the component $\alpha = pi{:}c$ in the design $(p{:}c) \times i$. Since all indices in the design are included in α, $f(\alpha) = 1$ and Step 1 results in no mean squares subtracted from $MS(pi{:}c)$; therefore, $\hat{\sigma}^2(pi{:}c) = MS(pi{:}c)$.

For the component $\alpha = p{:}c$ in the same design, $f(\alpha)$ is simply n_i. Step 1 results in subtracting only $MS(pi{:}c)$ from $MS(p{:}c)$, since $pi{:}c$ is the only component in the design that contains α (i.e., $p{:}c$) and one additional index (i). Step 2 results in no mean squares added. Therefore,

$$\hat{\sigma}^2(p{:}c) = [MS(p{:}c) - MS(pi{:}c)]/n_i .$$

For the component $\alpha = c$ in the design $(p{:}c) \times i$, the product of the sample sizes for the indices not included in α is $n_p n_i$. Step 1 results in subtracting both $MS(p{:}c)$ and $MS(ci)$ from $MS(c)$. Step 2 results in adding $MS(pi{:}c)$. Step 3 results in no mean squares subtracted. Therefore,

$$\hat{\sigma}^2(c) = [MS(c) - MS(p{:}c) - MS(ci) + MS(pi{:}c)]/n_p n_i .$$

Insert Figure 2 about here

Figure 2 uses Venn diagrams to illustrate the estimation of the variance of the three components discussed above. In the Venn diagram approach, a mean square for a main effect is represented by a circle; a mean square for an interaction is represented by the intersection of two or more circles; and a variance component is represented by a part of a circle that usually looks like a phase of the moon. More specifically, a part of a circle represents $\underline{f}(\alpha)\hat{\sigma}^2(\alpha)$. The Venn diagram approach to determining estimates of variance components is quite useful for relatively simple designs, such as p x i and (p:c) x i. However, this approach is not possible with some complicated designs, and this approach is difficult to employ with designs that involve considerable nesting, such as the design (p:c) x (i:s:t).

Algorithm 2 provides an estimate of the magnitude of a variance component--not its statistical significance. Even if a variance component is not statistically significant, it is an unbiased estimate, and it is better to use it, than to replace it with zero (Cronbach et al., 1972). Nevertheless, estimated variance components, like other statistics, are subject to sampling variation. This topic is outside the intended scope of this paper, but pertinent issues are treated by Cronbach et al. (1972, pp. 49-56), by Searle, (1971), and to some extent by Scheffé (1959) and Winer (1971). If, however, Algorithm 2 results in a negative estimate of a variance component, then the use of either Algorithm 2 or the Venn diagram approach is questionable. Procedures for treating negative estimates are discussed by Cronbach et al. (1972, pp. 57). One such procedure involves use of expected mean squares.

Expected Mean Squares

Although a G study usually does not require expected mean squares, it is easy to obtain them for the random effects model using the notation introduced in this paper. In general, for the random effects model, the expected mean square associated with the component β is:

$$\underline{EMS}(\beta) = \sum_{\alpha} \underline{f}(\alpha)\sigma^2(\alpha) \;; \tag{10}$$

where α is any component that contains all of the indices in β, $\underline{f}(\alpha)$ is defined by Equation 4, and $\sigma^2(\alpha)$ is the random effects variance component for α.

Consider, for example, the component $\underline{p}$ in the design $\underline{p} \times (\underline{i}:\underline{s})$. From Figure 1 and Table 3, it is clear that the components that contain the index $\underline{p}$ are $\underline{p}$, $\underline{ps}$, and $\underline{pi}:\underline{s}$. Applying Equation 4 to these components given $\underline{f}(\underline{p}) = \underline{n}_i\underline{n}_s$, $\underline{f}(\underline{ps}) = \underline{n}_i$, and $\underline{f}(\underline{pi}:\underline{s}) = 1$. Therefore,

$$\underline{EMS}(\underline{p}) = \sigma^2(\underline{pi}:\underline{s}) + \underline{n}_i\sigma^2(\underline{ps}) + \underline{n}_i\underline{n}_s\sigma^2(\underline{p}) . \tag{11a}$$

Similarly,

$$\underline{EMS}(\underline{s}) = \sigma^2(\underline{pi}:\underline{s}) + \underline{n}_i\sigma^2(\underline{ps}) + \underline{n}_p\sigma^2(\underline{i}:\underline{s}) + \underline{n}_p\underline{n}_i\sigma^2(\underline{s}) \tag{11b}$$

$$\underline{EMS}(\underline{i}:\underline{s}) = \sigma^2(\underline{pi}:\underline{s}) + \underline{n}_p\sigma^2(\underline{i}:\underline{s}) \;; \tag{11c}$$

$$\underline{EMS}(\underline{ps}) = \sigma^2(\underline{pi}:\underline{s}) + \underline{n}_i\sigma^2(\underline{ps}) \;; \tag{11d}$$

$$\underline{EMS}(\underline{pi}:\underline{s}) = \sigma^2(\underline{pi}:\underline{s}) . \tag{11e}$$

Perhaps the most important use of expected mean squares in a G study is to estimate variance components when Algorithm 2 or the Venn diagram approach results in one or more negative estimates for variance components. Consider, for example, the expected mean squares provided by Equations 11a - 11e for the $\underline{p} \times (\underline{i}:\underline{s})$ design. Equation 11e can be used to estimate $\sigma^2(\underline{pi}:\underline{s})$; and then Equation 11d can be used to estimate $\sigma^2(\underline{ps})$. If the estimate of $\sigma^2(\underline{ps})$ is negative, then zero is substituted for the negative estimate, and this zero is carried forward as the estimate of $\sigma^2(\underline{ps})$ in all other expected mean square equations. This "plausible solution" to the negative estimate problem is suggested by Cronbach et al. (1972, pp. 57).

D Studies for Random, Fixed, and Mixed Models

The primary result of a typical G study is the estimated random effects variance components for the G study design. These G study variance components are for single observations based on random sampling of one condition of each facet from an infinite universe of admissible conditions (or observations) for the facet. By comparison, a decision maker will want to use these results in some D study that involves its own sample size, n', and universe size N', for each facet in the universe of generalization. If, for example, $N' \rightarrow \infty$, then the facet involves sampling from an infinite universe of generalization; and if $n' = N'$, then the facet is fixed in the universe of generalization. Here and in Cronbach et al. (1972) n refers to the size of the sample and N to the size of the universe of admissible observations from the G study. Similarly, n' refers to the sample size and N' to the size of the universe of generalization defined by some D study.

In performing a D study, then, the decision maker must specify, directly or indirectly, the sample sizes and universe sizes for each of the facets in the universe of generalization. In addition, the decision maker must specify the object of measurement. It is usually the case that the facet for persons, or some aggregate of persons (such as a class), serves as the object of measurement in a D study. However, any facet could serve as the object of measurement (see Cardinet, Tourneur, & Allal, 1976). Suppose, for example, that the design $(p:c) \times i$ were used in the G study. A D study might use persons, items, or class means as the objects of measurement. In some literature the terms "object of measurement" and "unit of analysis" are used synonymously.

However, recently the unit of analysis issue has been viewed primarily in the context of choosing an appropriate unit of analysis (see Cronbach, Deken, and Webb, 1976; Haney, 1974b). This, of course, is an important issue, but it is outside the scope of this paper. Our concern here is with issues in analyzing D study data once the object of measurement has been chosen. In order to avoid ambiguity, therefore, we use the term "object of measurement" rather than "unit of analysis."

D Study Variance Components

Suppose a G study is conducted using the design $\underline{p} \times \underline{i} \times \underline{o}$. Table 2 provides the estimated random effects variance components resulting from such a G study. A typical D study, associated with such a G study, might use $\underline{p}$ as the object of measurement. For such a D study, the observed score for person $\underline{p}$, assuming an infinite universe of generalization for the item and occasion facets, can be represented as:

$$X_{\underline{p}} = X_{\underline{pIO}} = \mu + \mu_{\underline{p}}^{\sim} + \mu_{\underline{I}}^{\sim} + \mu_{\underline{O}}^{\sim} + \mu_{\underline{pI}}^{\sim} + \mu_{\underline{pO}}^{\sim} + \mu_{\underline{IO}}^{\sim} + \mu_{\underline{pIO}}^{\sim} . \quad (12)$$

where experimental error $\underline{e}$ is completely confounded with $\mu_{\underline{pIO}}^{\sim}$. In Equation 12, an upper-case subscript indicates the mean for a D study sample of size $\underline{n}'$; i.e.,

$$\mu_{\underline{I}}^{\sim} = \frac{1}{\underline{n}_{\underline{i}}'} \sum_{\underline{i}=1}^{\underline{n}_{\underline{i}}'} \mu_{\underline{i}}^{\sim} ,$$

$$\text{and } X_{\underline{p}} = X_{\underline{pIO}} = \frac{1}{\underline{n}_{\underline{i}}'\underline{n}_{\underline{o}}'} \sum_{\underline{i}=1}^{\underline{n}_{\underline{i}}'} \sum_{\underline{o}=1}^{\underline{n}_{\underline{o}}'} X_{\underline{pio}} ,$$

$$\text{where } X_{\underline{pio}} = \mu + \mu_{\underline{p}}^{\sim} + \mu_{\underline{i}}^{\sim} + \mu_{\underline{o}}^{\sim} + \mu_{\underline{pi}}^{\sim} + \mu_{\underline{po}}^{\sim} + \mu_{\underline{io}}^{\sim} + \mu_{\underline{pio}}^{\sim}$$

Note that here we use the abbreviation X_p to mean X_{pIO}, where $\underline{p}$ is the object of measurement for the D study.

For each of the score effects in Equation 12, the estimated D study variance component is obtained by dividing the estimated G study variance component by the frequency of sampling the effect within the object of measurement. In general, the frequency of sampling the component α within the object of measurement component γ is:

$$\underline{d}(\alpha|\gamma) = \begin{cases} 1 \text{ if } \alpha \text{ contains only indices in } \gamma\text{; and, otherwise} \\ \text{the product of the D study sample sizes for all} \\ \text{indices in } \alpha \text{ that are not in } \gamma\text{.} \end{cases} \quad (13)$$

For example, for the component $\underline{p}$ in the D study design represented by the structural model in Equation 12, $\underline{d}(\alpha|\gamma) = \underline{d}(\underline{p}|\underline{p}) = 1$; and the estimated D study variance of the component $\underline{p}$ is $\hat{\sigma}^2(\underline{p})/1 = \hat{\sigma}^2(\underline{p})$. For the component $\underline{I}$, $d(\alpha|\gamma) = n_i'$, and the estimated D study variance component for $\underline{I}$ is $\hat{\sigma}^2(\underline{I}) = \hat{\sigma}^2(\underline{i})/\underline{n}_i'$. For the component $\underline{pI}$, $\underline{d}(\alpha|\gamma) = \underline{d}(\underline{pI}|\underline{p}) = \underline{n}_i'$, and the estimated D study variance component for $\underline{pI}$ is $\hat{\sigma}^2(\underline{pI}) = \hat{\sigma}^2(\underline{pi})/\underline{n}_i'$.

All D study variance components for the design $\underline{p} \times \underline{i} \times \underline{o}$ are reported in Table 8. It is important to note that these variance components are for a random effects D study; i.e., $n_i' < \underline{N}_i' \rightarrow \infty$ and $\underline{n}_o' < \underline{N}_o' \rightarrow \infty$. It is also possible to express D study variance components in terms of a model different from the random effects model. (See subsequent discussion of sampling from a finite universe.) However, even when one or more facets is fixed in the universe of generalization, it is usually more informative to use and report the random effects D study variance components. Various combinations of these components provide the summary statistics typically used in a D study. The only important exception occurs in the case of sampling from a finite universe of generalization; this possibility is considered later.

Insert Table 8 about here

By convention, here and in Cronbach et al. (1972), D study estimated variance components and summary statistics formed from them, are expressed in terms of mean scores. For example, $\hat{\sigma}^2(\underline{I}) = \hat{\sigma}^2(i)/\underline{n}_i'$ in Table 8 is the D study estimated variance component associated with the mean score for a sample of $\underline{n}_i'$ items. It is also possible to express D study variance components in terms of total scores. For example, the D study estimated variance component associated with the total score for a sample of $\underline{n}_i'$ items is $\underline{n}_i'\hat{\sigma}^2(\underline{i})$. In general, for the total score metric, D study components are obtained by multiplying (rather than dividing) G study variance components by the sampling frequency within the object of measurement (see Equation 13).

D Study Summary Statistics

D study variance components are useful in and of themselves, because they provide a direct indication of the relative magnitude in the D study, of each of the independently estimable components of score variance. In addition, D study variance components are frequently used to estimate one or more of the following:

$\sigma^2(\tau)$: the universe score variance for the object of measurement τ, which is analogous to the true score variance in classical test theory;[4]

$\mathcal{E}\sigma^2(X)$: the <u>expected</u> <u>observed</u> <u>score</u> <u>variance</u>, which is the expected value over design replications of observed <u>deviation</u> scores;

$\mathcal{E}\rho^2$: an intraclass correlation coefficient, called a coefficient of generalizability, which is analogous to a reliability coefficient in classical test theory;

$\mathcal{E}\sigma^2(\delta)$: the error variance for making <u>comparative</u> decisions among the objects of measurement (e.g., persons), which is analogous to the error of measurement in classical test theory. "The error δ is the discrepancy between the observed deviation score and the universe score expressed in deviation form" (Cronbach et al., 1972, p. 25).

$\sigma^2(\Delta)$: the average error variance <u>within</u> an object of measurement (e,g, person), where error is defined as the difference between observed and universe score; and

$\sigma^2(\varepsilon)$: the variance of errors of <u>estimate</u> from the linear regression of universe scores on observed scores; that is, $\sigma^2(\varepsilon)$ is the variance of the discrepancies between estimated and actual universe scores.

The following equations provide some useful relationships among estimates of the statistics introduced above:

$$\widehat{\mathcal{E}\sigma^2}(X) = \hat{\sigma}^2(\tau) + \widehat{\mathcal{E}\sigma^2}(\delta) \; ; \tag{14}$$

$$\widehat{\mathcal{E}\rho^2} = \frac{\hat{\sigma}^2(\tau)}{\widehat{\mathcal{E}\sigma^2}(X)} = \frac{\hat{\sigma}^2(\tau)}{\hat{\sigma}^2(\tau) + \widehat{\mathcal{E}\sigma^2}(\delta)} ; \qquad (15)$$

$$\hat{\sigma}^2(\varepsilon) \doteq \hat{\sigma}^2(\tau)(1 - \widehat{\mathcal{E}\rho^2}) \qquad (16)$$

Equation 14 is analogous to the classical test theory result that the variance of observed scores equals the variance of true scores plus the variance of error scores. Note, in particular, that the error variance in Equation 14 is $\widehat{\mathcal{E}\sigma^2}(\delta)$, not $\hat{\sigma}^2(\Delta)$. The latter has no clear analogue in classical test theory with its emphasis on parallel measurements (see Lord, 1962); however, Brennan and Kane (in press-b) show that $\sigma^2(\Delta)$ is related to a type of error variance discussed by Lord (1957) prior to the advent of generalizability theory. Also Brennan and Kane (in press-a, in press-b) show that $\sigma^2(\Delta)$ is usually an appropriate estimate of error variance for domain-referenced mastery tests, whereas $\widehat{\mathcal{E}\sigma^2}(\delta)$ is seldom appropriate.

As implied by Equation 15, a coefficient of generalizability is defined as the ratio of universe score variance to <u>expected</u> observed score variance. In terms of estimates, $\widehat{\mathcal{E}\rho^2}$ is a consistent estimator of $\sigma^2(\tau)/\mathcal{E}\sigma^2(X)$, because $\hat{\sigma}^2(\tau)$ and $\widehat{\mathcal{E}\sigma^2}(X)$ are both unbiased estimates (see Lord and Novick, 1968, pp. 201-203). Also, the notation $\widehat{\mathcal{E}\rho^2}$ is indicative of the fact that a generalizability coefficient can be interpreted as a squared correlation or intraclass correlation coefficient (Cronbach, Ikeda, & Avner, 1964), as well as an approximation to the expected value of the correlation between pairs of measurements (Cronbach, et al., 1972, Chapter 8).[5]

In Equation 16, $\hat{\sigma}^2(\varepsilon)$ is strictly appropriate only if the regression equation for universe scores on observed scores is determined from the actual conditions used in the D study. Otherwise, $\hat{\sigma}^2(\varepsilon)$ in Equation 16 is an underestimate of $\sigma^2(\varepsilon)$ (see Cronbach et al., 1972, pp. 78-84).

Combining D Study Variance Components

In order to determine which variance components enter each of the summary statistics defined above, it is necessary that the D study be clearly specified. Here, the nature of a particular D study employing a specific design will be identified in the following manner: $\underline{D}(\gamma|\underline{V}|\underline{F}|\underline{R})$, where

γ = object of measurement component (i.e., the facet that serves as the object of measurement for the D study);

$\underline{V}$ = main effect index in γ ;

$\underline{F}$ = the set of facets that are fixed in the universe of generalization (i.e., facets for which $\underline{n}' = \underline{N}'$); and

$\underline{R}$ = the set of random facets (i.e., facets for which the D study contains a random sample of $\underline{n}'$ conditions from the universe of generalization for the facet).

Here, for random facets, it is assumed that the universe of generalization is infinite (i.e., $\underline{n}' < \underline{N}' \to \infty$); later, we consider random sampling of conditions from a finite universe of generalization.

In the notation $\underline{D}(\gamma|\underline{V}|\underline{F}|\underline{R})$, $\underline{F}$ and $\underline{R}$ specify the universe of generalization, and every index in the D study design is in $\underline{V}$, $\underline{F}$, or $\underline{R}$. There are, however, two restrictions on $\underline{D}(\gamma|\underline{V}|\underline{F}|\underline{R})$. First, each index in γ must be in either $\underline{V}$ or $\underline{F}$ but not in both. For example, if $\underline{p}:\underline{c}$ is the object of measurement component γ, then $\underline{p}$ might be in $\underline{V}$ and $\underline{c}$ in $\underline{F}$, but $\underline{c}$ could not be in both $\underline{V}$ and $\underline{F}$. Second, there must be at least one index in $\underline{R}$ in order to make the D study informative; otherwise the D study would not involve generalization over any facet.

An Algebraic Procedure. Given $\underline{D}(\gamma|\underline{V}|\underline{F}|\underline{R})$, the components that enter $\sigma^2(\tau)$, $\mathcal{E}\sigma^2(X)$, $\mathcal{E}\sigma^2(\delta)$, and $\sigma^2(\Delta)$ are, respectively:

$$\tau(\gamma|\underline{V}|\underline{F}|\underline{R}) = \mathcal{E}_{\underline{R}} X_\gamma - \mathcal{E}_{\underline{V}}\mathcal{E}_{\underline{R}} X_\gamma \; ; \tag{17}$$

$$X(\gamma|\underline{V}|\underline{F}|\underline{R}) = X_\gamma - \mathcal{E}_{\underline{V}} X_\gamma \; ; \tag{18}$$

$$\delta(\gamma|\underline{V}|\underline{F}|\underline{R}) = (X_\gamma - \mathcal{E}_{\underline{V}} X_\gamma) - (\mathcal{E}_{\underline{R}} X_\gamma - \mathcal{E}_{\underline{V}}\mathcal{E}_{\underline{R}} X_\gamma) \; ; \tag{19}$$

and $$\Delta(\gamma|\underline{V}|\underline{F}|\underline{R}) = X_\gamma - \mathcal{E}_{\underline{R}} X_\gamma \; ; \tag{20}$$

where each expectation is taken over the population or universe. In Equations 17-20, $\mathcal{E}_{\underline{R}} X_\gamma$ is the universe score for the D study, and $\tau(\gamma|\underline{V}|\underline{F}|\underline{R})$ is the universe deviation score. Similarly, X_γ is the observed score for the D study and $X(\gamma|\underline{V}|\underline{F}|\underline{R})$ is the observed deviation score.

Consider, for example, the observed score $X_{\underline{p}}$ for the design $\underline{p} \times \underline{i} \times \underline{o}$ in Equation 12, and suppose that the D study is $\underline{D}(\gamma|\underline{V}|\underline{F}|\underline{R}) = \underline{D}(\underline{p}|\underline{p}|-|\underline{I},\underline{O})$, where the symbol "-" is used to indicate that there are no fixed facets in the universe of generalization.

From Equation 17, the components that enter $\sigma^2(\tau)$ are:

$$\tau(\underline{p}|\underline{p}|-|\underline{I},\underline{O}) = \mathcal{E}_{\underline{I}}\mathcal{E}_{\underline{O}} X_{\underline{p}} - \mathcal{E}_{\underline{p}}\mathcal{E}_{\underline{I}}\mathcal{E}_{\underline{O}} X_{\underline{p}}$$

$$= \mu_{\underline{p}} - \mu$$

$$= \mu_{\underline{p}}^{\sim}$$

and, therefore,

$$\sigma^2(\tau) = \sigma^2(\mu_{\underline{p}}\sim) = \sigma^2(\underline{p}) \tag{21}$$

From Equation 18, the components that enter $\mathcal{E}\sigma^2(X)$ are:

$$\begin{aligned} X(\underline{p}|\underline{p}|-|\underline{I},\underline{O}) &= X_{\underline{p}} - \mathcal{E}_{\underline{p}} X_{\underline{p}} \\ &= X_{\underline{p}} - \mu_{\underline{IO}} \\ &= \mu_{\underline{p}}\sim + \mu_{\underline{pI}}\sim + \mu_{\underline{pO}}\sim + \mu_{\underline{pIO}}\sim \ ; \end{aligned}$$

and, therefore,

$$\mathcal{E}\sigma^2(X) = \sigma^2(\underline{p}) + \sigma^2(\underline{pI}) + \sigma^2(\underline{pO}) + \sigma^2(\underline{pIO}) \ . \tag{22}$$

Note that $\mathcal{E}\sigma^2(X)$ is different from the total variance, $\sigma^2(X)$, which is the sum of all D study variance components (see Equation 12).

From Equation 19, the components that enter $\mathcal{E}\sigma^2(\delta)$ are:

$$\begin{aligned} \delta(\underline{p}|\underline{p}|-|\underline{I},\underline{O}) &= (X_{\underline{p}} - \mathcal{E}_{\underline{p}} X_{\underline{p}}) - (\mathcal{E}_{\underline{I}}\mathcal{E}_{\underline{O}} X_{\underline{p}} - \mathcal{E}_{\underline{p}}\mathcal{E}_{\underline{I}}\mathcal{E}_{\underline{O}} X_{\underline{p}}) \\ &= (X_{\underline{p}} - \mu_{\underline{IO}}) - (\mu_{\underline{p}} - \mu) \\ &= \mu_{\underline{pI}}\sim + \mu_{\underline{pO}}\sim + \mu_{\underline{pIO}}\sim \ ; \end{aligned}$$

and, therefore,

$$\mathcal{E}\sigma^2(\delta) = \sigma^2(\underline{pI}) + \sigma^2(\underline{pO}) + \sigma^2(\text{pIO}) \ . \tag{23}$$

From the above results it is clear that $\mathcal{E}\sigma^2(X)$ equals the sum of $\sigma^2(\tau)$ and $\mathcal{E}\sigma^2(\delta)$, as indicated in Equation 14.

Finally, from Equation 20, the components that enter $\sigma^2(\Delta)$ are:

$$\Delta(\underline{p}|\underline{p}|-|\underline{I},\underline{O}) = X_{\underline{p}} - \mathcal{E}_{\underline{I}}\mathcal{E}_{\underline{O}} X_{\underline{p}}$$

$$= X_{\underline{p}} - \mu_{\underline{p}}$$

$$= \mu_{\underline{I}}^{\sim} + \mu_{\underline{O}}^{\sim} + \mu_{\underline{pI}}^{\sim} + \mu_{\underline{pO}}^{\sim} + \mu_{\underline{IO}}^{\sim} + \mu_{\underline{pIO}}^{\sim} ;$$

and, therefore,

$$\sigma^2(\Delta) = \sigma^2(\underline{I}) + \sigma^2(\underline{O}) + \sigma^2(\underline{pI}) + \sigma^2(\underline{pO}) + \sigma^2(\underline{IO}) + \sigma^2(\underline{pIO}) \quad . \quad (24)$$

With the exception of $\widehat{\mathcal{E}\sigma^2}(X)$, estimates of the above results are reflected in the fourth column of Table 8. $\widehat{\mathcal{E}\sigma^2}(X)$ is most easily obtained using Equation 14; and, of course, Equations 15 and 16 can be used to obtain $\widehat{\mathcal{E}\rho^2}$ and $\hat{\sigma}^2(\varepsilon)$.

A Notational Procedure. The procedure represented by Equations 17 to 20 is a straightforward application of generalizability theory, but it does involve some degree of algebraic complexity. A simpler procedure involves a direct application of the notation for variance components used in this paper. If the D study is $\underline{D}(\gamma|\underline{V}|\underline{F}|\underline{R})$, then:

(a) variance components that enter $\mathcal{E}\sigma^2(X)$ are all variance components that contain the index in $\underline{V}$;

(b) variance components that enter $\sigma^2(\tau)$ are all variance components that contain the index in $\underline{V}$ and do not contain any index in $\underline{R}$;

(c) variance components that enter $\mathcal{E}\sigma^2(\delta)$ are all variance components that contain the index in $\underline{V}$ and one or more of the indices in $\underline{R}$; and

(d) variance components that enter $\sigma^2(\Delta)$ are all variance components that contain one or more of the indices in $\underline{R}$.

For example, for the model Equation 12 and $\underline{D}(\gamma|\underline{V}|\underline{F}|\underline{R}) = \underline{D}(\underline{p}|\underline{p}|-|\underline{I},\underline{O})$, $\mathcal{E}\sigma^2(X)$ consists of the variance components that contain the index $\underline{p}$ in $\underline{V}$. These components are $\sigma^2(\underline{p})$, $\sigma^2(\underline{pI})$, $\sigma^2(\underline{pO})$, and $\sigma^2(\underline{pIO})$; therefore, $\mathcal{E}\sigma^2(X)$ is the result provided by Equation 22. The variance components that enter $\sigma^2(\tau)$ are those which contain a $\underline{p}$ and do not contain an $\underline{I}$ or an $\underline{O}$ (the indices in $\underline{R}$). The only component that satisfies these two conditions is $\sigma^2(\underline{p})$; therefore, $\sigma^2(\tau)$ equals $\sigma^2(\underline{p})$, as specified by Equation 21. The variance components that enter $\mathcal{E}\sigma^2(\delta)$ are those which contain a $\underline{p}$, and one or both of the indices $\underline{I}$ and $\underline{O}$. These components are $\sigma^2(\underline{pI})$, $\sigma^2(\underline{pO})$, and $\sigma^2(\underline{pIO})$; therefore, $\mathcal{E}\sigma^2(\delta)$ is the result provided by Equation 23. Similarly, all variance components except $\sigma^2(\underline{p})$ contain either an $\underline{I}$ or an $\underline{O}$, or both; therefore, $\sigma^2(\Delta)$ is the result provided by Equation 24.

Illustrative D Studies

In this section, the procedures for combining variance components are discussed with reference to various D studies that might be used with each of the five illustrative designs. The results of applying either procedure are presented in tables similar to Table 8, and certain interesting and/or illustrative aspects of these results are discussed in the text. (In studying these examples it is useful to refer to the model equations in Tables 1 - 5 for the five illustrative designs.)

Insert Table 9 about here

The Design $\underline{p} \times \underline{i}$. Table 9 presents a single application of the procedures for combining variance components in the $\underline{p} \times \underline{i}$ design. For $\underline{D}(\underline{p}|\underline{p}|-|\underline{I})$,

$$\tau(\underline{p}|\underline{p}|-|\underline{I}) = \mathop{\mathcal{E}}_{\underline{I}} X_{\underline{p}} - \mathop{\mathcal{E}}_{\underline{p}} \mathop{\mathcal{E}}_{\underline{I}} X_{\underline{p}}$$

$$= \mu_{\underline{p}} - \mu$$

$$= \mu_{\underline{p}}^{\sim} ;$$

$$\delta(\underline{p}|\underline{p}|-|\underline{I}) = (X_{\underline{p}} - \mathop{\mathcal{E}}_{\underline{p}} X_{\underline{p}}) - (\mathop{\mathcal{E}}_{\underline{I}} X_{\underline{p}} - \mathop{\mathcal{E}}_{\underline{p}} \mathop{\mathcal{E}}_{\underline{I}} X_{\underline{p}})$$

$$= (X_{\underline{p}} - \mu_{\underline{I}}) - (\mu_{\underline{p}} - \mu)$$

$$= \mu_{\underline{pI}}^{\sim} ; \text{ and}$$

$$\Delta(\underline{p}|\underline{p}|-|\underline{I}) = X_{\underline{p}} - \mathop{\mathcal{E}}_{\underline{I}} X_{\underline{p}}$$

$$= X_{\underline{p}} - \mu_{\underline{p}}$$

$$= \mu_{\underline{I}}^{\sim} + \mu_{\underline{pI}}^{\sim} .$$

Therefore,

$$\sigma^2(\tau) = \sigma^2(\underline{p}) ;$$

$$\mathcal{E}\sigma^2(\delta) = \sigma^2(\underline{pI}) = \sigma^2(\underline{pi})/\underline{n}_{\underline{i}} ;$$

$$\sigma^2(\Delta) = \sigma^2(\underline{I}) + \sigma^2(\underline{pI})$$

$$= \frac{\sigma^2(\underline{i})}{\underline{n}_{\underline{i}}'} + \frac{\sigma^2(\underline{pi})}{\underline{n}_{\underline{i}}'} ;$$

$$\mathcal{E}\sigma^2(X) = \sigma^2(\tau) + \mathcal{E}\sigma^2(\delta)$$

$$= \sigma^2(\underline{p}) + \frac{\sigma^2(\underline{pi})}{\underline{n}_{\underline{i}}'} \text{; and}$$

$$\widehat{\mathcal{E}\rho^2} = \frac{\hat{\sigma}^2(\underline{p})}{\hat{\sigma}^2(\underline{p}) + \hat{\sigma}^2(\underline{pi})/\underline{n}_{\underline{i}}'} . \qquad (25)$$

Equation 25 is algebraically identical to Cronbach's (1951) Coefficient α, and, when items are scored dichotmously, Equation 25 is identical to Kuder and Richardson's (1937) Formula 20. However, the derivation of $\widehat{\mathcal{E}\rho^2}$ in Equation 25 does not require the assumption of classically parallel tests with equal means, equal variances, and equal intercorrelations. Rather, the derivation of $\widehat{\mathcal{E}\rho^2}$ requires the weaker assumption of randomly parallel tests. Two tests are randomly parallel if they both consist of a random sample of the same number of items from the same universe. Also, Equation 25 illustrates the regularity that forms the basis for the Spearman-Brown Formula for changes in test length. Increasing the number of items, $\underline{n}_{\underline{i}}'$, by a specified factor leaves $\hat{\sigma}^2(\tau)$ unchanged and decreases $\widehat{\mathcal{E}\sigma^2}(\delta)$ by the inverse of the factor. This type of regularity occurs because the universe of generalization contains only one facet--namely, the item facet. For more complicated universes of generalization, the Spearman-Brown Formula does not usually apply.

The Design p x i x o. Table 9 treats D studies for three different universes of generalization when the person $\underline{p}$ is the object of measurement. The first D study, $\underline{D}(\underline{p}|\underline{p}| - |I,O)$, has been discussed in detail. The other two involve a single fixed facet, $\underline{I}$ or $\underline{O}$.

For example, for $\underline{D}(\underline{p}|\underline{p}|\underline{I}|\underline{O})$,

$$\tau(\underline{p}|\underline{p}|\underline{I}|\underline{O}) = \mathcal{E}_{\underline{O}} X_{\underline{p}} - \mathcal{E}_{\underline{p}} \mathcal{E}_{\underline{O}} X_{\underline{p}}$$

$$= \mu_{\underline{pI}} - \mu_{\underline{I}}$$

$$= \mu_{\underline{p}}^{\sim} + \mu_{\underline{pI}}^{\sim} \text{ ; and}$$

$$\delta(\underline{p}|\underline{p}|\underline{I}|\underline{O}) = (X_{\underline{p}} - \mathcal{E}_{\underline{p}} X_{\underline{p}}) - (\mathcal{E}_{\underline{O}} X_{\underline{p}} - \mathcal{E}_{\underline{p}} \mathcal{E}_{\underline{O}} X_{\underline{p}})$$

$$= (X_{\underline{p}} - \mu_{\underline{IO}}) - (\mu_{\underline{pI}} - \mu_{\underline{I}})$$

$$= \mu_{\underline{pO}}^{\sim} + \mu_{\underline{pIO}}^{\sim} .$$

That is, $\sigma^2(\tau)$ consists of components that contain $\underline{p}$ (the index in $\underline{V}$) and do not contain $\underline{O}$ (the index in $\underline{R}$), whereas $\mathcal{E}\sigma^2(\delta)$ consists of components that contain $\underline{p}$ and $\underline{O}$. Similarly, for $\underline{D}(\underline{p}|\underline{p}|\underline{O}|\underline{I})$,

$$\tau(\underline{p}|\underline{p}|\underline{O}|\underline{I}) = \mu_{\underline{p}}^{\sim} + \mu_{\underline{pO}}^{\sim} \text{ ; and}$$

$$\delta(\underline{p}|\underline{p}|\underline{O}|\underline{I}) = \mu_{\underline{pI}}^{\sim} + \mu_{\underline{pIO}}^{\sim} .$$

That is, $\sigma^2(\tau)$ consists of components that contain $\underline{p}$ (the index in $\underline{V}$) and do not contain $\underline{I}$ (the index in $\underline{R}$), whereas $\mathcal{E}\sigma^2(\delta)$ consists of components that contain $\underline{p}$ and $\underline{I}$.

In both studies, $\mathcal{E}\sigma^2(X)$ is identical to $\mathcal{E}\sigma^2(X)$ for $\underline{D}(\underline{p}|\underline{p}|-|\underline{I},\underline{O})$. This is a particular instance of a general rule; namely, once $\underline{V}$ is specified $\mathcal{E}\sigma^2(X)$ is unaffected by changes in the universe of generalization. However, the universe of generalization does affect $\sigma^2(\tau)$ and $\mathcal{E}\sigma^2(\delta)$. Using

$\underline{D}(\underline{p}|\underline{p}|-|\underline{I},\underline{O})$ as a basis for comparison, in $\underline{D}(\underline{p}|\underline{p}|\underline{I}|\underline{O})$ the variance component $\sigma^2(\underline{pI})$ moves from $\mathcal{E}\sigma^2(\delta)$ to $\sigma^2(\tau)$; and in $\underline{D}(\underline{p}|\underline{p}|\underline{O}|\underline{I})$, $\sigma^2(\underline{pO})$ moves from $\mathcal{E}\sigma^2(\delta)$ to $\sigma^2(\tau)$.

Going one step further, if both $\underline{I}$ and $\underline{O}$ were fixed in the universe of generalization, then $\widehat{\mathcal{E}\sigma^2(X)}$ would be identical to $\hat{\sigma}^2(\tau)$, $\widehat{\mathcal{E}\hat{\sigma}^2(\delta)}$ would be unestimable, and, therefore, $\widehat{\mathcal{E}\rho^2}$ would be unity. Such a result occurs whenever there are no facets over which the decision maker generalizes. It is for this reason that $\underline{R}$ should contain at least one index for the D study to be informative.

As indicated in Table 8, $\sigma^2(\Delta)$ never includes the variance components in $\sigma^2(\tau)$, and $\sigma^2(\Delta)$ always includes the variance components in $\mathcal{E}\sigma^2(\delta)$. The remaining variance components enter $\sigma^2(\Delta)$ only if they contain an index in $\underline{R}$. For example, in $\underline{D}(\underline{p}|\underline{p}|\underline{I}|\underline{O})$, $\sigma^2(\underline{I})$ does not enter $\sigma^2(\Delta)$ because this variance component does not contain $\underline{O}$. From another point of view, $\sigma^2(\underline{I})$ does not enter $\sigma^2(\Delta)$ because $\underline{I}$ is fixed in the universe of generalization, and, therefore, $\mu_{\underline{I}}{\sim}$ is a constant for all persons.

Insert Table 10 about here

The Design $\underline{p} \times (\underline{i}:\underline{s})$. Table 10 presents illustrative D studies for a design that involves a single level of nesting in the universe of generalization. For the second D study, $\underline{D}(\gamma|\underline{V}|\underline{F}|\underline{R}) = \underline{D}(\underline{p}|\underline{p}|\underline{S}|\underline{I})$, with $\underline{S}$ fixed in the universe of generalization,

$$\begin{aligned}\tau(\underline{p}|\underline{p}|\underline{S}|\underline{I}) &= \mathcal{E}_{\underline{I}} X_{\underline{p}} - \mathcal{E}_{\underline{p}}\mathcal{E}_{\underline{I}} X_{\underline{p}} \\ &= \mu_{\underline{pS}} - \mu_{\underline{S}} \\ &= \mu_{\underline{p}}{\sim} + \mu_{\underline{pS}}{\sim} ;\end{aligned}$$

$$\delta(\underline{p}|\underline{p}|\underline{S}|\underline{I}) = (X_{\underline{p}} - \mathcal{E}_{\underline{p}} X_{\underline{p}}) - (\mathcal{E}_{\underline{I}} X_{\underline{p}} - \mathcal{E}_{\underline{p}} \mathcal{E}_{\underline{I}} X_{\underline{p}})$$

$$= (X_{\underline{p}} - \mu_{\underline{I:S}}) - (\mu_{\underline{pS}} - \mu_{\underline{S}})$$

$$= \tilde{\mu}_{\underline{pI:S}} \text{ ; and}$$

$$\Delta(\underline{p}|\underline{p}|\underline{S}|\underline{I}) = X_{\underline{p}} - \mathcal{E}_{\underline{I}} X_{\underline{p}}$$

$$= X_{\underline{p}} - \mu_{\underline{pS}}$$

$$= \tilde{\mu}_{\underline{I:S}} + \tilde{\mu}_{\underline{pI:S}} .$$

In terms of the notational procedure for combining variance components, $\sigma^2(\tau)$ consists of variance components that contain $\underline{p}$ (the index in $\underline{V}$) and do not contain $\underline{I}$ (the index in $\underline{P}$); i.e.,

$$\sigma^2(\tau) = \sigma^2(\underline{p}) + \sigma^2(\underline{pS}) .$$

Similarly, $\mathcal{E}\sigma^2(\delta)$ consists of variance components that contain $\underline{p}$ and $\underline{I}$; i.e.,

$$\mathcal{E}\sigma^2(\delta) = \sigma^2(\underline{pI:S}) ,$$

and $\sigma^2(\Delta)$ consists of variance components that contain $\underline{I}$: i.e.,

$$\sigma^2(\Delta) = \sigma^2(I:S) + \sigma^2(pI:S) .$$

If, then, $\underline{S}$ is fixed in the universe of generalization,

$$\widehat{\mathcal{E}\rho^2} = \frac{\hat{\sigma}^2(\underline{p}) + \hat{\sigma}^2(\underline{pS})}{\hat{\sigma}^2(\underline{p}) + \hat{\sigma}^2(\underline{pS}) + \hat{\sigma}^2(\underline{pI:S})} ;$$

whereas, if $\underline{S}$ is a sample from an infinite universe,

$$\widehat{\mathcal{E}\rho^2} = \frac{\hat{\sigma}^2(\underline{p})}{\hat{\sigma}^2(\underline{p}) + \hat{\sigma}^2(\underline{pS}) + \hat{\sigma}^2(\underline{pI{:}S})}$$

[see $\underline{D}(\underline{p}|\underline{p}|-|\underline{I},\underline{S})$ in Table 10].

The characteristics and utility of generalizability coefficients that take stratification of content into account, were studied by Cronbach and his colleagues (Rajaratnam, Cronbach, & Gleser, 1965; Cronbach, Schönemann, & McKie, 1965) shortly after their seminal work on generalizability theory (Cronbach, Rajaratnam, & Gleser, 1963). They concluded that if the items in a test can be divided into different content strata, then estimates of reliability should take the stratification into account; otherwise, reliability may be seriously underestimated.

Insert Table 11 about here

D Studies with Nesting in the Object of Measurement Component. Consider the design $(\underline{p}{:}\underline{c}) \times \underline{i}$ and the D study $\underline{D}(\underline{p}{:}\underline{c}|\underline{p}|\underline{c}|\underline{I})$ in Table 11. For this D study, the object of measurement component, γ, is $\underline{p}{:}\underline{c}$ and each person is nested within a particular class. Since both $\sigma^2(\underline{p}{:}\underline{c})$ and $\sigma^2(\underline{c})$ contain only indices in $\gamma = p{:}c$, the D study sampling frequency for each of these G study variance components is unity (see Equation 13). For this D study, the universe of generalization contains a single fixed class and an infinite universe of items, from which a sample of $\underline{n}_i'$ items are drawn. Consequently, $\sigma^2(\underline{c})$ does not enter $\sigma^2(\tau)$, $\sigma^2(\delta)$, $\sigma^2(\Delta)$, or $\mathcal{E}\sigma^2(X)$; for example,

$$X(\underline{p}:\underline{c}|\underline{p}|\underline{c}|\underline{I}) = X_{\underline{p}} - \mathscr{E}_{\underline{p}} X_{\underline{p}}$$

$$= X_{\underline{p}} - \mu_{\underline{cI}}$$

$$= \mu_{\underline{p}:\underline{c}}^{\sim} + \mu_{\underline{pI}:\underline{c}}^{\sim} ,$$

$$\text{and } \mathscr{E}\sigma^2(X) = \sigma^2(\underline{p}:\underline{c}) + \sigma^2(\underline{pI}:\underline{c}) ;$$

i.e., $\mathscr{E}\sigma^2(X)$ consists of variance components that contain $\underline{p}$ (the index in $\underline{V}$). It is particularly important to note that this D study is not identical to the D study for the $\underline{p} \times \underline{i}$ design in Table 9 (see Brennan, 1975).

Insert Table 12 about here

Table 12 provides illustrative D studies using $\underline{p}:\underline{c}$ as the object of measurement component in the design $(\underline{p}:\underline{c}) \times (\underline{i}:\underline{s}:\underline{t})$. Although these D studies use a considerably more complicated design, it is relatively easy to apply the notational procedure for combining variance components.

Insert Table 13 about here

D Studies with Class as the Object of Measurement. Table 13 provides illustrative D studies for the design $(\underline{p}:\underline{c}) \times \underline{i}$ when the object of measurement is the class, $\underline{c}$, or more specifically the class mean:

$$X_{\underline{c}} = X_{\underline{PI}:\underline{c}} = \mu + \mu_{\underline{p}:\underline{c}}^{\sim} + \mu_{\underline{c}}^{\sim} + \mu_{\underline{I}}^{\sim} + \mu_{\underline{cI}}^{\sim} + \mu_{\underline{PI}:\underline{c}}^{\sim} ; \qquad (26)$$

where experimental error $\underline{e}$ is completely confounded with $\mu_{\underline{PI}:\underline{c}}^{\sim}$

For the D study in Table 13 that involves generalization over both samples of persons and samples of items,

$$X(\underline{c}|\underline{c}|-|\underline{P},\underline{I}) = X_{\underline{c}} - \mathscr{E}_{\underline{c}} X_{\underline{c}}$$

$$= X_{\underline{c}} - \mu_{\underline{I}}$$

$$= \mu_{\underline{c}}^{\sim} + \mu_{\underline{P}:\underline{c}}^{\sim} + \mu_{\underline{cI}}^{\sim} + \mu_{\underline{PI}:\underline{c}}^{\sim} ,$$

$$\text{and } \mathscr{E}\sigma^2(X) = \sigma^2(\underline{c}) + \sigma^2(\underline{P}:\underline{c}) + \sigma^2(\underline{cI}) + \sigma^2(\underline{PI}:\underline{c}) \;;$$

i.e., $\mathscr{E}\sigma^2(X)$ consists of components that contain $\underline{c}$ (the index in $\underline{V}$). As noted previously, $\mathscr{E}\sigma^2(X)$ is unchanged by changes in the universe of generalization, but this is not true for $\sigma^2(\tau)$, $\mathscr{E}\sigma^2(\delta)$, or $\sigma^2(\Delta)$. In particular, Table 13 shows that when generalization is over both persons and items,

$$\sigma^2(\tau) = \sigma^2(\underline{c}) \;;$$

when generalization is over items, only,

$$\sigma^2(\tau) = \sigma^2(\underline{c}) + \sigma^2(\underline{P}:\underline{c}) \;;$$

and when generalization is over persons, only,

$$\sigma^2(\tau) = \sigma^2(\underline{c}) + \sigma^2(\underline{cI}) \;.$$

The estimate of each of these three different universe score variances [or, equivalently, the three different estimates of $\mathscr{E}\sigma^2(\delta)$] provides a different estimate of the generalizability of class means. That is, these estimates differ with respect to the facet(s) over which the decision maker generalizes.

The estimation of reliability, or generalizability, when the object of measurement is some aggregate of persons, has been a particularly troublesome problem in recent years (see Haney, 1974a, 1974b). In terms of published literature, Medley and Mitzel (1963) and Pilliner and his colleagues (Maxwell & Pilliner, 1968; Pilliner, 1965; and Pilliner, Sutherland & Taylor, 1960) appear to be among the earliest researchers to recognize that the class mean is frequently the variable of interest, rather than the score for a person. More recently, large scale evaluations, such as those undertaken for Head Start (Smith & Bissell, 1970), Follow Through (Abt Associates, 1974; Haney, 1974b), and the National Day Care Study (Stallings, Wilcox, & Travers, 1976), have frequently required estimates of reliability when class mean is the object of measurement. Similar issues arise in the study of course evaluation questionnaires (Gillmore, Kane & Naccarato, Note 1; Kane, Gillmore, & Crooks, 1976) and studies of school effectiveness and accountability (Dyer, Linn, & Patton, 1969; Marco, 1974; Page, 1975).

The literature does contain some approaches to the estimation of reliability for class means using classical test theory. For example, Shaycoft (1962), Wiley (1970), and Thrash and Porter (Note 1) developed three different coefficients, each of which assume that an observed score is the sum of a true score and an <u>undifferentiated</u> error term. However, each of these procedures makes different specific assumptions about what constitutes an appropriate estimate of error variance. As a result, each procedure gives a different estimate of the reliability of class means. Kane and Brennan (1977) show that Wiley's coefficient is equivalent to $\widehat{\mathcal{E}\rho^2}$ when items are fixed, Thrash and Porter's coefficient is equivalent to $\widehat{\mathcal{E}\rho^2}$ when persons within classes are fixed, and Shaycoft's coefficient is an overestimate of $\widehat{\mathcal{E}\rho^2}$ when persons within classes are fixed.

It is not suprising that none of these coefficients corresponds to $\widehat{\mathcal{E}\rho^2}$ when generalization is over both persons and items. Classical test theory does not specifically allow for differentiating among sources of error in a multi-faceted universe of generalization.

Insert Table 14 about here

Table 14 provides illustrative D studies using the class mean as the object of measurement in the design (p:c) x (i:s:t). The reader can easily verify the results in Table 14 using the notational procedure for combining variance components. D studies for this design are clearly more complicated than those for the (p:c) x i design; however, in large scale testing efforts involving analysis of class means it is frequently the case that data are collected according to rather complicated sampling plans. To overlook this complexity is to discard some amount of information in the data, and, therefore, to potentially restrict the utility of the results.

Sampling from Finite Universes

To this point, our discussion of generalizability theory has focused on D studies in which each of the facets in the universe of generalization is either fixed (i.e., $\underline{n}' = \underline{N}' < \infty$) or essentially infinite (i.e., $\underline{n}' < \underline{N}' \rightarrow \infty$). We have seen that such D studies can be carried out using G study random effects variance components; or, more specifically, variance components for single observations based on random sampling of one condition of each facet from an infinite universe of admissible conditions, or observations, for the facet ($\underline{N} \rightarrow \infty$). It is also possible to develop equations for calculating G study variance components for random sampling of one condition of each facet from a finite universe of admissible conditions for the facet ($\underline{N} < \infty$). These G study variance components are especially useful in D studies characterized by sampling from a finite universe of generalization.

Unfortunately, any verbal discussion of different sampling procedures in typical ANOVA terms is apt to involve considerable ambiguity. The problem is primarily evident in the term "random effect," which, in traditional ANOVA terms, actually implies "random sampling from an infinite universe," as opposed to no sampling at all (i.e., "fixed effect"), or random sampling from a finite universe. It is particularly important to note that the traditional ANOVA notion of "random effect" does not mean sampling from a finite universe, even though such sampling is "random." For this reason we will restrict our use of the term "random effect" to random sampling from an infinite universe.

G Study Considerations

In this section we develop equations for G study estimated variance components and expected mean squares for any model $\underline{M}$. That is, these equations are applicable to $\underline{n} < \underline{N} < \infty$ (sampling from a finite universe of admissible observations), $\underline{n} = \underline{N} < \infty$ (fixed effect), and $\underline{n} < \underline{N} \to \infty$ (random effect).

<u>Estimation of Variance Components</u>. If $\hat{\sigma}^2(\alpha|\underline{M})$ is the estimated variance for the component α, given a G study design using the model $\underline{M}$, then

$$\hat{\sigma}^2(\alpha|\underline{M}) = \hat{\sigma}^2(\alpha) + \sum_{\underline{j}} \left[\frac{\hat{\sigma}^2(\beta_{\underline{j}})}{\underline{F}(\beta_{\underline{j}})} \right] ; \tag{27}$$

where $\hat{\sigma}^2(\alpha)$ and $\hat{\sigma}^2(\beta_j)$ are estimated G study variance components for the random effects model calculated from Algorithm 2;

β_j = any component, except α, that contains all the indices in α and

$\underline{F}(\beta_j)$ = the product of the G study <u>universe</u> sizes ($\underline{N}$'s) for all indices in β_j <u>except</u> those indices in α. (28)

As in Algorithm 2, no distinction is made between nested and non-nested indices.

If $N \to \infty$ for all facets, then $\hat{\sigma}^2(\beta_{\underline{j}})/\underline{F}(\beta_{\underline{j}})$ is always zero, and $\hat{\sigma}^2(\alpha|\underline{M}) = \hat{\sigma}^2(\alpha)$. If, on the other hand, all effects are fixed, then all universe sizes ($\underline{N}$'s) equal the sample sizes ($\underline{n}$'s) in the G study. In the case of mixed models, some effects are random and some fixed. For other models that involve sampling from a finite universe for one or more facets, the actual universe size is used in Equation 28.

For example, in the design, ($\underline{p}$:$\underline{c}$) x $\underline{i}$, consider the component $\underline{p}$:$\underline{c}$ The only other component that contains the indices $\underline{p}$ and $\underline{c}$ is $\underline{pi}$:$\underline{c}$; therefore,

$$\hat{\sigma}^2(p{:}c|M) = \hat{\sigma}^2(\underline{p}{:}\underline{c}) + \frac{\hat{\sigma}^2(\underline{pi}{:}\underline{c})}{\underline{N}_i}$$

Now, if $\underline{i}$ is a fixed effect in the G study, then $\underline{N}_i = \underline{n}_i$ and

$$\hat{\sigma}^2(\underline{p}{:}\underline{c}|\underline{M}) = \hat{\sigma}^2(\underline{p}{:}\underline{c}) + \frac{\hat{\sigma}^2(\underline{pi}{:}\underline{c})}{\underline{n}_i}.$$

If, on the other hand, $\underline{i}$ is a random effect, then the universe size is considered infinite and $\hat{\sigma}^2(\underline{p}{:}\underline{c}|\underline{M}) = \hat{\sigma}^2(\underline{p}{:}\underline{c})$. If $\underline{n}_i$ is a sample from a finite universe of size $\underline{N}_i$, then the actual value of $\underline{N}_i$ is used in the above equation.

Also, in the design $(\underline{p}{:}\underline{c}) \times \underline{i}$ consider the component $\underline{i}$. The components that contain $\underline{i}$ are $\underline{ci}$ and $\underline{pi}{:}\underline{c}$; therefore,

$$\hat{\sigma}^2(\underline{i}|\underline{M}) = \hat{\sigma}^2(\underline{i}) + \frac{\hat{\sigma}^2(\underline{ci})}{\underline{N}_c} + \frac{\hat{\sigma}^2(\underline{pi}{:}\underline{c})}{\underline{N}_p\underline{N}_c}.$$

If, for example, $\underline{p}$ is a random effect and $\underline{c}$ is a fixed effect in the G study, then

$$\hat{\sigma}^2(\underline{i}|\underline{M}) = \hat{\sigma}^2(i) + \frac{\hat{\sigma}^2(\underline{ci})}{\underline{n}_c}.$$

Expected Mean Squares. For any model $\underline{M}$, the expected mean square for the component β is:

$$\underline{EMS}(\beta|\underline{M}) = \sum_{\alpha} h(\alpha)\underline{f}(\alpha)\hat{\sigma}^2(\alpha) \; ; \qquad (29)$$

where α is any component that contains all the indices in β; $f(\alpha)$ is defined by Equation 4; $h(\alpha)$ is the product of the terms $(1 - \underline{n}/\underline{N})$ for all main effect indices in α that are not in β; and $\hat{\sigma}^2(\alpha)$ is the estimated random effects G study variance component for α calculated from Algorithm 2.

For the component $\underline{p}$ in the design $\underline{p} \times (\underline{i}:\underline{s})$,

$$\underline{EMS}(\underline{p}|\underline{M}) = \left(1 - \frac{\underline{n}_{\underline{i}}}{\underline{N}_{\underline{i}}}\right)\hat{\sigma}^2(\underline{pi}:\underline{s}) + \left(1 - \frac{\underline{n}_{\underline{s}}}{\underline{N}_{\underline{s}}}\right)\underline{n}_{\underline{i}}\hat{\sigma}^2(\underline{ps}) + \underline{n}_{\underline{i}}\underline{n}_{\underline{s}}\hat{\sigma}^2(\underline{p}).$$

If both items and subtests are random effects, then both $(1 - \underline{n}_{\underline{i}}/\underline{N}_{\underline{i}})$ and $(1 - \underline{n}_{\underline{s}}/\underline{N}_{\underline{s}})$ are unity and $\underline{EMS}(\underline{p}|\underline{M})$ equals $\underline{EMS}(\underline{p})$ for the random effects model. If items are random and subtests are fixed, then $(1 - \underline{n}_{\underline{i}}/\underline{N}_{\underline{i}})$ is unity, $(1 - \underline{n}_{\underline{s}}/\underline{N}_{\underline{s}})$ is zero, and,

$$\underline{EMS}(\underline{p}|\underline{M}) = \hat{\sigma}^2(\underline{pi}:\underline{s}) + \underline{n}_{\underline{i}}\underline{n}_{\underline{s}}\hat{\sigma}^2(\underline{p}) .$$

If items are random and the subtests in the G study are a sample of size $\underline{n}_{\underline{s}}$ from a finite universe of size $\underline{N}_{\underline{s}}$, then

$$\underline{EMS}(\underline{p}|\underline{M}) \quad \hat{\sigma}^2(\underline{pi}.\underline{s}) + \left(1 - \frac{\underline{n}_{\underline{s}}}{\underline{N}_{\underline{s}}}\right)\underline{n}_{\underline{i}}\hat{\sigma}^2(\underline{ps}) + \underline{n}_{\underline{i}}\underline{n}_{\underline{s}}\hat{\sigma}^2(\underline{p}) .$$

D Study Considerations

The discussion thus far has focused on D studies in which each of the facets in the universe of generalization is either fixed (i.e., $\underline{n}' = \underline{N}' < \infty$) or essentially infinite (i.e., $\underline{n}' < \underline{N}' \to \infty$). It has also been assumed that G study variance components are reported for an infinite universe of admissible observations (i.e., $\underline{N} \to \infty$). For most D studies these assumptions are quite reasonable; however, a D study might involve sampling from a finite universe of generalization. More specifically, it is possible that, for one or more facets, $\underline{n}' < \underline{N}' = \underline{N} < \infty$. For each such facet, the D study uses a sample of size n' from a finite universe of generalization of size N', which is identical to the universe size, N, assumed in the G study.

For D studies characterized by sampling from a finite universe, a limiting case occurs when $n' = N' = N < \infty$. In this case, the D study actually includes all conditions of the facet in the universe of generalization; and the facet is fixed in the universe of generalization. Another limiting case occurs when $n' < N' = N \to \infty$. In this case, the D study includes a random sample from the (essentially) infinite set of conditions for the facet. (This is the definition of a random effect in the typical ANOVA sense). When $n' < N' = N < \infty$, it is also assumed that the sampling of the n' conditions is random, but the universe of generalization for the facet is finite.

Let us consider the case in which $N' = N < \infty$ for only one of the facets in the universe of generalization, and the D study involves sampling this facet $n' < N$ times. In general, the steps involved in conducting the D study are: (a) use Algorithm 3 to obtain G study variance components which reflect the fact that $N < \infty$; (b) obtain D study variance components that take into account sampling from a finite universe; and (c) employ procedures for combining D study variance components, as appropriate.

Consider, for example, the design $p \times i \times o$ with p as the D study object of measurement. Let us assume that, in the universe the item facet has a finite number of conditions, N_i, which are sampled n'_i times in the D study. Since $N_i < \infty$, the estimated G study variance components are obtained using Algorithm 3. They are reported in Table 15 for the $p \times i \times o$ design.

Insert Table 15 about here

For any D study component, α, the estimated variance of this component is:

$$\hat{\sigma}^2_{\underline{D}}(\alpha|\underline{N}_i < \infty) = \left(1 - \frac{\underline{n}'_i}{\underline{N}_i}\right)\frac{\hat{\sigma}^2_{\underline{G}}(\alpha|\underline{N}_i < \infty)}{\underline{d}(\alpha|\gamma)} \tag{30}$$

if $\underline{n}'_i$ is in $\underline{d}(\alpha|\gamma)$; otherwise,

$$\hat{\sigma}^2_{\underline{D}}(\alpha|\underline{N}_i < \infty) = \frac{\hat{\sigma}^2_{\underline{G}}(\alpha|\underline{N}_i < \infty)}{\underline{d}(\alpha|\gamma)} \; ; \tag{31}$$

where $\underline{d}(\alpha|\gamma)$ is defined as Equation 13, and $(1 - \underline{n}'_i/\underline{N}_i)$ is the finite universe correction (see Cochran, 1963, p. 23) associated with variances for the item facet. Table 16 reports the estimated D study variance components for the design p x i x o when p is the object of measurement.

It is important to note that the D study variance components defined in Equations 30 and 31 are for a random sampling model where $\underline{N}_i < \infty$ and $\underline{N}_o \to \infty$. These variance components are completely analogous to the D study variance components for a random effects model reported in the fourth column of Table 8. Indeed, for $\underline{N}_i \to \infty$ the D study variance components in Table 16 are identical to the D study variance components in Table 8. Also, Equations 17-20 and the corresponding notational procedure for combining variance components are completely applicable to D study variance components that involve sampling from a finite universe.

Consider, again, Table 16 and suppose that the D study is $\underline{D}(\underline{p}|\underline{p}|-|\underline{I},\underline{O})$ implying that occasions are randomly sampled from an infinite universe and items are randomly sampled from a finite universe of size $\underline{N}_i = \underline{N}'_i$. For this D study, the reader can verify that

$$\sigma^2(\tau) = \sigma^2(\underline{p}|\underline{N}_{\underline{i}} < \infty)$$

$$= \sigma^2(\underline{p}) + \sigma^2(\underline{pi})/\underline{N}_{\underline{i}} \;;$$

$$\mathcal{E}\sigma^2(\delta) = \sigma^2(\underline{pI}|\underline{N}_{\underline{i}} < \infty) + \sigma^2(\underline{pO}|\underline{N}_{\underline{i}} < \infty) + \sigma^2(\underline{pIO}|\underline{N}_{\underline{i}} < \infty)$$

$$= \left(1 - \frac{\underline{n}'_{\underline{i}}}{\underline{N}_{\underline{i}}}\right)\frac{\sigma^2(\underline{pi})}{\underline{n}'_{\underline{i}}} + \frac{\sigma^2(\underline{po})}{\underline{n}'_{\underline{o}}} + \frac{\sigma^2(\underline{pio})}{\underline{n}'_{\underline{i}}\underline{n}'_{\underline{o}}} \;;$$

and,
$$\mathcal{E}\sigma^2(X) = \sigma^2(\tau) + \mathcal{E}\sigma^2(\delta)$$

$$= \sigma^2(\underline{p}) + \frac{\sigma^2(\underline{pi})}{\underline{n}'_{\underline{i}}} + \frac{\sigma^2(\underline{po})}{\underline{n}'_{\underline{o}}} + \frac{\sigma^2(\underline{pio})}{\underline{n}'_{\underline{i}}\underline{n}'_{\underline{o}}} \;; \quad (32)$$

where the variance components without the conditional statement "$\underline{N}_{\underline{i}} < \infty$" are the usual random effects D study variance components.

It is both informative and instructive to note that Equation 32 is identical to Equation 22; i.e., $\mathcal{E}\sigma^2(X)$ is unchanged by whether or not the universe of generalization involves sampling from a finite universe. This is true for all of the possible D studies given a particular design and a particular object of measurement.

Consider, again, Table 16 and suppose the D study were $\underline{D}(\underline{p}|\underline{p}|\underline{O}|\underline{I})$ with occasions fixed. In this case,

$$\sigma^2(\tau) = \sigma^2(\underline{p}|\underline{N}_{\underline{i}} < \infty) + \sigma^2(\underline{pO}|\underline{N}_{\underline{i}} < \infty)$$

$$= \sigma^2(\underline{p}) + \frac{\sigma^2(\underline{pi})}{\underline{N}_{\underline{i}}} + \frac{\sigma^2(\underline{po})}{\underline{n}'_{\underline{o}}} + \frac{\sigma^2(\underline{pio})}{\underline{N}_{\underline{i}}\underline{n}'_{\underline{o}}} \;;$$

$$\mathcal{E}\sigma^2(\delta) = \sigma^2(\underline{pI}|\underline{N}_{\underline{i}} < \infty) + \sigma^2(\underline{pIO}|\underline{N}_{\underline{i}} < \infty) \;;$$

$$= \left(1 - \frac{\underline{n}'_{\underline{i}}}{\underline{N}_{\underline{i}}}\right)\left[\frac{\sigma^2(\underline{pi})}{\underline{n}'_{\underline{i}}} + \frac{\sigma^2(\underline{pio})}{\underline{n}'_{\underline{i}}\underline{n}'_{\underline{o}}}\right] \;;$$

and, $\mathcal{E}\sigma^2(X)$ is identified to Equation 32.

If the D study design were $\underline{D}(\underline{p}|\underline{p}|\underline{I}|\underline{O})$, then the item facet would be fixed in this particular D study. In this case,

$$\sigma^2(\tau) = \sigma^2(\underline{p}|\underline{N}_{\underline{i}} < \infty) + \sigma^2(\underline{pI}|\underline{N}_{\underline{i}} < \infty)$$

$$= \sigma^2(\underline{p}) + \sigma^2(\underline{pi})/\underline{n}_{\underline{i}}' \; ; \qquad (33)$$

and $$\mathcal{E}\sigma^2(\delta) = \sigma^2(\underline{pO}|\underline{N}_{\underline{i}} < \infty) + \sigma^2(\underline{pIO}|\underline{N}_{\underline{i}} < \infty)$$

$$= \sigma^2(\underline{po})/\underline{n}_{\underline{o}}' + \sigma^2(\underline{pio})/\underline{n}_{\underline{i}}'\underline{n}_{\underline{o}}' \; . \qquad (34)$$

Equations 33 and 34 are identical to those obtained using the fifth column of Table 8. This must be so, because when the item facet is fixed in the universe of generalization there is, by definition, no random sampling of the conditions of this facet; and the size of the universe has no bearing on $\sigma^2(\tau)$, $\mathcal{E}\sigma^2(\delta)$, $\sigma^2(\Delta)$, or any quantities formed from them.

The procedures discussed above can be extended to D studies that involve sampling from a finite universe for more than one facet. In such cases, estimated G study variance components are obtained using Algorithm 3, and estimated D study variance components are obtained using a more general version of Equations 30 and 31. For example, if the D study involves sampling from a finite universe for both the item facet and the occasion facet in the $\underline{p} \times \underline{i} \times \underline{o}$ design, then the finite universe correction in Equation 30 is:

(a) $(1 - \underline{n}_{\underline{i}}'/\underline{N}_{\underline{i}})(1 - \underline{n}_{\underline{o}}'/\underline{N}_{\underline{o}})$ if $\underline{d}(\alpha|\gamma)$ includes both $\underline{n}_{\underline{i}}'$ and $\underline{n}_{\underline{o}}'$;

(b) $(1 - \underline{n}_{\underline{i}}'/\underline{N}_{\underline{i}})$ if $\underline{d}(\alpha|\gamma)$ includes $\underline{n}_{\underline{i}}'$ but not $\underline{n}_{\underline{o}}'$; and

(c) $(1 - \underline{n}_{\underline{o}}'/\underline{N}_{\underline{o}})$ if $\underline{d}(\alpha|\gamma)$ includes $\underline{n}_{\underline{o}}'$ but not $\underline{n}_{\underline{i}}'$.

If $\underline{d}(\alpha|\gamma)$ includes neither $\underline{n}_{\underline{i}}'$ nor $\underline{n}_{\underline{o}}'$, then Equation 31 is applicable.

Comments and Conclusions

It is usual in both practical and theoretical contexts, to treat issues of reliability from a correlational viewpoint. The literature, for example, is filled with references to reliability coefficients that estimate "internal consistency," "equivalence," "stability," etc. While such coefficients and terms have a long and distinguished history, they can be a source of considerable confusion and ambiguity. In particular, it is frequently difficult to identify explicitly the magnitudes, types, and sources of error variance incorporated in such coefficients. The use of generalizability coefficients can avoid these problems, at least in part, if the nature of the universe of generalization is clearly specified. However, estimated variance components are even more informative and less ambiguous. Indeed, estimated variance components are the most informative outcome of a reliability study (APA, 1974). They can be used directly to obtain estimates of universe score variance and different types of error variance that are appropriate in different decision-making contexts. Variance components can be used, of course, to estimate generalizability coefficients; but such coefficients are of questionable value in the absence of the estimated variance components themselves. Note that it is the magnitude of variance components that is of primary interest--not their statistical significance. Also, variance components should not be expressed solely as a percentage or proportion of some total score variance. To do so is to obviate the more important uses of variance components.

Since the magnitudes of variance components are central to generalizability theory, it is important that the numerical estimates of variance components be as accurate as possible. Therefore, care should be taken to avoid the deleterious effects of rounding errors. For example, it is usually advisable that most, if not all, calculations involve at least three decimal places. This is particularly important when a G study involves binary data, which is the usual case for achievement tests.

The notational system used in this paper was invented in order to facilitate the statement of various "rules," procedures, and algorithms. There are only two principal ways in which this notational system differs from that used by Cronbach et al. (1972). First, this paper uses the nesting operator ":" to designate variance components that involve nesting; Cronbach and his colleagues use the "all confounded effects" procedure. Second, this paper specifies a particular D study using the notation $\underline{D}(\gamma|\underline{V}|\underline{F}|\underline{R})$. The notation $\underline{D}(\gamma|\underline{V}|\underline{F}|\underline{R})$ is very useful in specifying rules and procedures for combining D study variance components. Also, this notation clearly identifies the universe of generalization, and clearly distinguishes between the object of measurement and the universe of generalization. Cronbach et al. (1972) treat object of measurement considerations, but they do not emphasize them as much as this paper does. However, Cronbach et al. (1972) do clearly identify a fixed facet by concatenating its index with the symbol "*" or "**". In terms of certain theoretical expositions, the star notation has some distinct advantages.

This paper treats only G studies and D studies that involve orthogonal analysis of variance designs; i.e., designs that do not involve missing data and/or unequal size subgroups. The application of generalizability theory to non-orthogonal designs has received little attention in the literature. There are, however, two procedures that have been used or suggested for "converting" non-orthogonal designs to orthogonal ones. Kane et al. (1976), for example, report randomly discarding data until they had orthogonal designs for their studies of student evaluations of teaching. Also, for designs, such as $\underline{p} \times (\underline{i}:\underline{s})$, where the number of items is not a constant for all subtests, Cronbach et al. (1965) mention the possibility of using "half-sets" of items within each subtest. These procedures may not be ideal, but they are at least reasonable alternatives until research on variance components in non-orthogonal designs (see Searle, 1971) is applied to generalizability theory.

This paper provides a more detailed consideration of sampling from finite universes than is provided in Cronbach et al. (1972). Also, somewhat more consideration is given to generalizability theory in the context of different objects of measurement. However, in other respects this paper is not intended to cover, in depth or breadth, the extensive treatment of generalizability theory provided by Cronbach and his colleagues. (In particular, multivariate generalizability theory has not been treated at all here.) Rather, this paper is primarily intended to provide researchers and practitioners with a set of procedures to facilitate the application of generalizability theory to a broad range of measurement problems. It is inadvisable that these procedures be used mindlessly; the meaningful interpretation of any statistical analysis necessitates a thoughtful and informed consideration of the results.

Reference Notes

1. Gillmore, G. M., Kane, M. T., & Naccarato, R. W. The teacher and the course as units of analysis in the generalizability of student ratings of instruction (Report No. 77-9). Seattle, Washington: University of Wisconsin, Educational Assessment Center, February 1977.

2. Thrash, S. K., & Porter, A. C. Invalidity of a current method for estimating reliability. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, April 1974.

References

Abt Associates. Education as Experimentation: Evaluation of the Follow Through Planned variation model (2 vols.). Cambridge, Mass.: Abt Associates, March 1974.

Alexander, H. W. The estimation of reliability when several trials are available. Psychometrika, 1947, 12, 79-99.

American Psychological Association. Standards for educational and psycological tests (rev. ed.). Washington, D.C.: American Psychological Association, 1974.

Brennan, R. L. The calculation of reliability from a split-plot factorial design. Educational and Psychological Measurement, 1975, 35, 779-788.

Brennan, R. L., & Kane, M. T. An index of dependability for mastery tests. Journal of Educational Measurement, in press. (a)

Brennan, R. L., & Kane, M. T. Signal/noise ratios for domain-referenced tests. Psychometrika, in press. (b)

Burt C. The analysis of examination marks. In P. Hartog & E. C. Rhodes (Eds.), The marks of examiners. London: The Macmillan Company, 1936.

Burt, C. Test reliability estimated by analysis of variance. British Journal of Statistical Psychology, 1955, 8, 103-118.

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

Cochran, W. G. Sampling techniques (2nd ed.). New York: Wiley, 1963.

Cornfield, J., & Tukey, J. W. Average values of mean squares in factorials. Annals of Mathematical Statistics, 1956, 27, 907-949.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16, 292-334.

Cronbach, L. J., Deken, J. E., & Webb, N. Research on classrooms and schools: Formulation of questions, design, and analysis. Stanford, California: Stanford Evaluation Consortium, 1976.

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

Cronbach, L. J., Ikeda, M., & Avner, R. A. Intraclass correlation as an approximation to the coefficient of generalizability. Psychological Reports, 1964, 15, 727-736.

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

Cronbach, L. J., Schönemann, P., & McKie, T. D. Alpha coefficients for stratified-parallel tests. Educational and Psychological Measurement, 1965, 25, 291-312.

Dyer, H. S., Linn, R. L., & Patton, M. J. A comparison of four methods of obtaining discrepancy measures based on observed and predicted school system means on achievement tests. American Educational Research Journal, 1969, 6, 591-605.

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

Endler, N. S. Estimating variance components from mean squares for random and mixed effects analysis of variance models. Perceptual and Motor Skills, 1966, 22, 559-570.

Finlayson, D. S. The reliability of marking essays. British Journal of Educational Psychology, 1951, 35, 143-162.

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

Haney, W. The dependability of group mean scores. Unpublished special qualifying paper, Harvard Graduate School of Education, October 1974. (a)

Haney, W. Units of analysis issues in the evaluation of project Follow Through. Cambridge, Mass.: The Huron Institute, September 1974. (b)

Hoyt, C. J. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, 153-160.

Hunter, J. E. Probabilistic foundations for coefficients of generalizability. Psychometrika, 1968, 33, 1-18.

Jackson, R. W. B., & Ferguson, G. Studies on the reliability of tests. University of Toronto, 1941.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Wadsworth, 1968.

Kuder, G. F., & Richardson, M. W. The theory of the stimation of test reliability. Psychometrika, 1937, 2, 151-160.

Kane, M. T., & Brennan, R. L. The generalizability of class means. Review of Educational Research, 1977, 47, 267-292.

Kane, M. T., Gillmore, G. M., & Crooks, T. J. Student evaluations of teaching: The generalizability of class means. Journal of Educational Measurement, 1976, 13, 171-183.

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

Lord, F. M. Do tests of the same length have the same standard error of measurement? Educational and Psychological Measurement, 1957, 17, 510-521.

Lord, F. M. Test reliability: A correction. Educational and Psychological Measurement, 1962, 22, 511-512.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Loveland, E. H. Measurement of factors affecting test-retest reliability. Unpublished doctoral dissertation, University of Tennessee, 1952.

Marco, G. L. A comparison of selected school effectiveness measures based on longitudinal data. Journal of Educational Measurement, 1974, 11, 225-234.

Maxwell, A. E., & Pilliner, A. E. G. Deriving coefficients of reliability and agreement for ratings. British Journal of Mathematical and Statistical Psychology, 1968, 21, 105-116.

Medley, D. M., & Mitzel, H. E. Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), Handbook of research on teaching. Chicago, Illinois: Rand McNally, 1963.

Millman, J., & Glass, G. V. Rules of thumb for the ANOVA table. Journal of Educational Measurement, 1967, 4, 41-51.

Page, E. B. Statistically recapturing the richness within the classroom. Psychology in the Schools, 1975, 12, 339-344.

Pilliner, A. E. G. The application of analysis of variance components in psychometric experimentation. Unpublished doctoral dissertation, University of Edinburgh, 1965.

Pilliner, A. E. G., Sutherland, J., & Taylor, E. G. Zero error in moray house verbal reasoning tests. British Journal of Educational Psychology, 1960, 30, 53-62.

Rajaratnam, N. Reliability formulas for independent decision data when reliability data are matched. Psychometrika, 1960, 25, 261-271.

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

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Searle, S. R. Linear models. New York: Wiley, 1971.

Shaycoft, M. F. The statistical characteristics of school means. In J. C. Flanagan, J. T. Dailey, M. F. Shaycoft, D. B. Orr & I. Goldberg, Studies of the American high school. Pittsburgh: University of Pittsburgh, 1962.

Smith, M. S., & Bissell, J. S. Report analysis: The impact of Head Start. Harvard Educational Review, 1970, 40, 51-104.

Stallings, J., Wilcox, M., & Travers, J. Phase II instruments for the national day care cost-effects study: Instrument selection and field testing. Menlo Park, Calif.: Stanford Research Institute, January, 1976.

Vaughn, G. M., & Corballis, M. C. Beyond tests of significance: Estimating strength of effects in selected ANOVA designs. Psychological Bulletin, 1969, 72, 204-213.

Wiley, D. E. Design and analysis of evaluation studies. In D. E. Wiley & M. C. Wittrock (Eds.), The evaluation of instruction: Issues and problems. New York: Holt, Rinehart, and Winston, 1970.

Winer, B. J. Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill, 1971.

Webster, H. A generalization of Kuder-Richardson reliability formula 21. Educational and Psychological Measurement, 1960, 20, 131-138.

Footnotes

The author would like to thank Dr. Lee J. Cronbach, Dr. Michael T. Kane, Dr. Gerald M. Gillmore, and Dr. Michael B. Bunch for their many helpful comments and suggestions.

This research was partially supported by ONR Contract No. N00123-77-C-0739 between the American College Testing Program and the Navy Personnel Research and Development Center.

A previous version of this paper entitled, "'Rules of Thumb'" for Generalizability Analyses" was presented at the Annual Meeting of the American Educational Research Association, April, 1977.

[1]It may not be obvious that designs like (*p:c*) x (*i:s:t*) occur in practice. Suppose *c* is a school, *i* is an item, *s* is a content area or subtest, and *t* is a test. Given these verbal identifiers, this design means that each person is nested within a single school, each person responds to all items, each item is associated with a single content area or subtest, and each content area is associated with a single test. This kind of design very closely approximates the kind of data often collected to assess the reliability of test batteries. However, it is rarely the case that the analyses of such data distinguish among all potential sources of variance. Among other things, this paper is intended to aid researchers and practitioners in conceptualizing and performing such complex analyses.

[2]For each of the Venn diagrams in Figure 1, a circle is never nested within the intersection of two or more circles. This is a geometric indication that, for each of the five illustrative designs, no main effect is nested within an

interaction effect. Consider, however, the design (p:c) x [i:(s x o)], in which the main effect for items is nested within the interaction of subtests and occasions. This main effect would be represented i:so.

[3] The reader may omit this discussion of sums of squares without loss of continuity in the development of generalizability theory. This section is included because the notational system used here provides a convenient way to express sums of squares for a large class of ANOVA designs.

[4] Cronbach et al. (1972) usually use $\sigma^2(p)$ for universe score variance. Here, however, the general use of $\sigma^2(p)$ for universe score variance could create confusion, because objects of measurement other than the person p are treated in this paper.

[5] Generalizability coefficients have a form that is analogous to that of traditional reliability coefficients; however, the theoretical basis for generalizability coefficients is somewhat more complicated and beyond the intended scope of this paper. The interested reader can refer to Hunter (1968).

TABLE 1

Estimated Variance Components for Design $p \times i$ for Random Effects Model

Effect or Component	df	Estimated Variance Component
p	$n_p - 1$	$\hat{\sigma}^2(p) = [MS(p) - MS(pi)]/n_i$
i	$n_i - 1$	$\hat{\sigma}^2(i) = [MS(i) - MS(pi)]/n_p$
pi	$(n_p - 1)(n_i - 1)$	$\hat{\sigma}^2(pi) = MS(pi)$

Note. $X_{pi} = \mu + \mu_p^{\sim} + \mu_i^{\sim} + \mu_{pi}^{\sim} + e$.

TABLE 2

Estimated Variance Components for Design $p \times i \times o$ for Random Effects Model

Effect or Component	df	Estimated Variance Component
p	$n_p - 1$	$\hat{\sigma}^2(p) = [MS(p) - MS(pi) - MS(po) + MS(pio)]/n_i n_o$
i	$n_i - 1$	$\hat{\sigma}^2(i) = [MS(i) - MS(pi) - MS(io) + MS(pio)]/n_p n_o$
o	$n_o - 1$	$\hat{\sigma}^2(o) = [MS(o) - MS(po) - MS(io) + MS(pio)]/n_p n_i$
pi	$(n_p - 1)(n_i - 1)$	$\hat{\sigma}^2(pi) = [MS(pi) - MS(pio)]/n_o$
po	$(n_p - 1)(n_o - 1)$	$\hat{\sigma}^2(po) = [MS(po) - MS(pio)]/n_i$
io	$(n_i - 1)(n_o - 1)$	$\hat{\sigma}^2(io) = [MS(io) - MS(pio)]/n_p$
pio	$(n_p - 1)(n_i - 1)(n_o - 1)$	$\hat{\sigma}^2(pio) = MS(pio)$

Note. $X_{pio} = \mu + \mu_p^{\sim} + \mu_i^{\sim} + \mu_o^{\sim} + \mu_{pi}^{\sim} + \mu_{po}^{\sim} + \mu_{io}^{\sim} + \mu_{pio}^{\sim} + e$.

TABLE 3

Estimated Variance Components for Design $p \times (i{:}s)$ for Random Effects Model

Effect or Component	df	Estimated Variance Component
p	$n_p - 1$	$\sigma^2(p) = [MS(p) - MS(ps)]/n_i n_s$
$i{:}s$	$n_s(n_i - 1)$	$\sigma^2(i{:}s) = [MS(i{:}s) - MS(pi{:}s)]/n_p$
s	$n_s - 1$	$\sigma^2(s) = [MS(s) - MS(i{:}s) - MS(ps) + MS(pi{:}s)]/n_p n_i$
ps	$(n_p - 1)(n_s - 1)$	$\sigma^2(ps) = [MS(ps) - MS(pi{:}s)]/n_i$
$pi{:}s$	$n_s(n_p - 1)(n_i - 1)$	$\sigma^2(pi{:}s) = MS(pi{:}s)$

Note. $X_{pi:s} = \mu + \mu_p^{\sim} + \mu_{i:s}^{\sim} + \mu_s^{\sim} + \mu_{ps}^{\sim} + \mu_{pi:s}^{\sim} + e$.

TABLE 4

Estimated Variance Components for Design $(p:c) \times i$
for Random Effects Model

Effect or Component	df	Estimated Variance Component
$p:c$	$n_c(n_p - 1)$	$\hat{\sigma}^2(p:c) = [MS(p:c) - MS(pi:c)]/n_i$
c	$n_c - 1$	$\hat{\sigma}^2(c) = [MS(c) - MS(p:c) - MS(ci) + MS(pi:c)]/n_p n_i$
i	$n_i - 1$	$\hat{\sigma}^2(i) = [MS(i) - MS(ci)]/n_p n_c$
ci	$(n_c - 1)(n_i - 1)$	$\hat{\sigma}^2(ci) = [MS(ci) - MS(pi:c)]/n_p$
$pi:c$	$n_c(n_p - 1)(n_i - 1)$	$\hat{\sigma}^2(pi:c) = MS(pi:c)$

Note. $X_{pi:c} = \mu + \mu_{p:c}^{\sim} + \mu_{c}^{\sim} + \mu_{i}^{\sim} + \mu_{ci}^{\sim} + \mu_{pi:c}^{\sim} + e$.

TABLE 5

Estimated Variance Components for Design $(p{:}c) \times (i{:}s{:}t)$
for Random Effects Model

Effect or Component	df	Estimated Variance Component
$p{:}c$	$n_c(n_p - 1)$	$\hat{\sigma}^2(p{:}c) = [MS(p{:}c) - MS(pt{:}c)]/n_i n_s n_t$
c	$n_c - 1$	$\hat{\sigma}^2(c) = [MS(c) - MS(p{:}c) - MS(ct) + MS(pt{:}c)]/n_p n_i n_s n_t$
$i{:}s{:}t$	$n_s n_t(n_i - 1)$	$\hat{\sigma}^2(i{:}s{:}t) = [MS(i{:}s{:}t) - MS(ci{:}s{:}t)]/n_p n_c$
$s{:}t$	$n_t(n_s - 1)$	$\hat{\sigma}^2(s{:}t) = [MS(s{:}t) - MS(i{:}s{:}t) - MS(cs{:}t) + MS(ci{:}s{:}t)]/n_p n_c n_i$
t	$n_t - 1$	$\hat{\sigma}^2(t) = [MS(t) - MS(s{:}t) - MS(ct) + MS(sc{:}t)]/n_p n_c n_i n_s$
ct	$(n_c - 1)(n_t - 1)$	$\hat{\sigma}^2(ct) = [MS(ct) - MS(cs{:}t) - MS(pt{:}c) + MS(ps{:}ct)]/n_p n_i n_s$
$cs{:}t$	$n_t(n_c - 1)(n_s - 1)$	$\hat{\sigma}^2(cs{:}t) = [MS(cs{:}t) - MS(ci{:}s{:}t) - MS(ps{:}ct) + MS(pi{:}sc{:}t)]/n_p n_i$
$ci{:}s{:}t$	$n_s n_t(n_c - 1)(n_i - 1)$	$\hat{\sigma}^2(ci{:}s{:}t) = [MS(ci{:}s{:}t) - MS(pi{:}sc{:}t)]/n_p$
$pt{:}c$	$n_c(n_p - 1)(n_t - 1)$	$\hat{\sigma}^2(pt{:}c) = [MS(pt{:}c) - MS(ps{:}ct)]/n_i n_s$
$ps{:}ct$	$n_c n_t(n_p - 1)(n_s - 1)$	$\hat{\sigma}^2(ps{:}ct) = [MS\ ps{:}ct) - MS(pi{:}sc{:}t)]/n_i$
$pi{:}cs{:}t$	$n_c n_s n_t(n_p - 1)(n_i - 1)$	$\hat{\sigma}^2(pi{:}sc{:}t) = MS(pi{:}sc{:}t)$

Note. $X_{pi:sc:t} = \mu + \mu_{p:c}^{\sim} + \mu_{c}^{\sim} + \mu_{i:s:t}^{\sim} + \mu_{s:t}^{\sim} + \mu_{t}^{\sim} + \mu_{ct}^{\sim} + \mu_{cs:t}^{\sim} + \mu_{ci:s:t}^{\sim} + \mu_{pt:c}^{\sim} + \mu_{ps:ct}^{\sim} + \mu_{pi:sc:t}^{\sim} + e.$

TABLE 6

Components, Mean Scores, and Score Effects for Design $(p:c) \times i$

Component	Score Effects in Terms of Mean Scores	Mean Scores in Terms of Score Effects
$p:c$	$\mu_{p:c}^{\sim} = \mu_{p:c} - \mu_{c}$	$\mu_{p:c} = \mu + \mu_{p:c}^{\sim} + \mu_{c}^{\sim}$
c	$\mu_{c}^{\sim} = \mu_{c} - \mu$	$\mu_{c} = \mu + \mu_{c}^{\sim}$
i	$\mu_{i}^{\sim} = \mu_{i} - \mu$	$\mu_{i} = \mu + \mu_{i}^{\sim}$
ci	$\mu_{ci}^{\sim} = \mu_{ci} - \mu_{c} - \mu_{i} + \mu$	$\mu_{ci} = \mu + \mu_{c}^{\sim} + \mu_{i}^{\sim} + \mu_{ci}^{\sim}$
$pi:c$	$\mu_{pi:c}^{\sim} = \mu_{pi:c} - \mu_{p:c} - \mu_{ci} + \mu_{c}$	$\mu_{pi:c} = \mu + \mu_{p:c}^{\sim} + \mu_{c}^{\sim} + \mu_{i}^{\sim} + \mu_{pi:c}^{\sim}$

TABLE 7

Sums of Squares for Design $(p:c) \times i$

Component	Sums of Squares for Observed Mean Scores	Sums of Squares for Observed Score Effects: With respect to Sums of Squares for Observed Mean Scores	Sums of Squares for Observed Score Effects: With respect to Observed Score Effects
$p:c$	$[\overline{X}_{p:c}] = n_i \sum_p \sum_c \overline{X}^2_{p:c}$	$[\overline{X}_{p:c}]^\sim = [\overline{X}_{p:c}] - [\overline{X}_c]$	$= n_i \sum_p \sum_c (\overline{X}_{p:c}^\sim)^2$
c	$[\overline{X}_c] = n_p n_i \sum_c \overline{X}^2_c$	$[\overline{X}_c]^\sim = [X_c] - [\overline{X}]$	$= n_p n_i \sum_c (\overline{X}_c^\sim)^2$
i	$[\overline{X}_i] = n_p n_c \sum_i \overline{X}^2_i$	$[\overline{X}_i]^\sim = [\overline{X}_i] - [\overline{X}]$	$= n_p n_c \sum_i (\overline{X}_i^\sim)^2$
ci	$[\overline{X}_{ci}] = n_p \sum_c \sum_i \overline{X}^2_{ci}$	$[\overline{X}_{ci}]^\sim = [\overline{X}_{ci}] - [\overline{X}_c] - [\overline{X}_i] + [\overline{X}]$	$= n_p \sum_c \sum_i (\overline{X}_{ci}^\sim)^2$
$pi:c$	$[\overline{X}_{pi:c}] = \sum_p \sum_i \sum_c \overline{X}^2_{pi:c}$	$[\overline{X}_{pi:c}]^\sim = [\overline{X}_{pi:c}] - [\overline{X}_{p:c}] - [\overline{X}_{ci}] + [\overline{X}_c]$	$= \sum_p \sum_i \sum_c (\overline{X}_{pi:c}^\sim)^2$
Total	$[\overline{X}] = n_p n_i n_c \overline{X}^2$		

Note. For this design with one observation per cell, $\overline{X}_{pi:c}$ is based on only one observation, and, therefore, $\overline{X}_{pi:c} = X_{pi:c}$.

TABLE 8

D Studies for Design $p \times i \times o$
With Person (p) as the Object of Measurement

Estimated G Study Variance Components	D Study Sampling Frequency	Estimated D Study Variance Components	$(p\|p\|-\|I,O)$	$(p\|p\|I\|O)$	$(p\|p\|O\|I)$
$\hat{\sigma}^2(p)$	1	$\hat{\sigma}^2(p)$	τ	τ	τ
$\hat{\sigma}^2(i)$	n_i'	$\hat{\sigma}^2(I)$	Δ		Δ
$\hat{\sigma}^2(o)$	n_o'	$\hat{\sigma}^2(O)$	Δ	Δ	
$\hat{\sigma}^2(pi)$	n_i'	$\hat{\sigma}^2(pI)$	Δ,δ	τ	Δ,δ
$\hat{\sigma}^2(po)$	n_o'	$\hat{\sigma}^2(pO)$	Δ,δ	Δ,δ	τ
$\hat{\sigma}^2(io)$	$n_i'n_o'$	$\hat{\sigma}^2(IO)$	Δ	Δ	Δ
$\hat{\sigma}^2(pio)$	$n_i'n_o'$	$\hat{\sigma}^2(pIO)$	Δ,δ	Δ,δ	Δ,δ

Note. The entries τ, δ, and Δ indicate which estimated D study variance components enter $\hat{\sigma}^2(\tau)$, $\hat{\sigma}^2(\delta)$, and $\hat{\sigma}^2(\Delta)$, respectively.

TABLE 9

D Studies for Design $p \times i$
With Person (p) as the Object of Measurement

Estimated G Study Variance Components	D Study Sampling Frequency	Estimated D Study Variance Components	$(p\|p\|-\|I)$
$\hat{\sigma}^2(p)$	1	$\hat{\sigma}^2(p)$	τ
$\hat{\sigma}^2(i)$	n'_i	$\hat{\sigma}^2(I)$	Δ
$\hat{\sigma}^2(pi)$	n'_i	$\hat{\sigma}^2(pI)$	Δ, δ

TABLE 10

D Studies for Design $p \times (i:s)$
With Persons (p) as the Object of Measurement

Estimated G Study Variance Components	D Study Sampling Frequency	Estimated D Study Variance Components	$(p\|p\|-\|I,S)$	$(p\|p\|S\|I)$
$\hat{\sigma}^2(p)$	1	$\hat{\sigma}^2(p)$	τ	τ
$\hat{\sigma}^2(i:s)$	$n'_i n'_s$	$\hat{\sigma}^2(I:S)$	Δ	Δ
$\hat{\sigma}^2(s)$	n'_s	$\hat{\sigma}^2(S)$	Δ	
$\hat{\sigma}^2(ps)$	n'_s	$\hat{\sigma}^2(pS)$	Δ, δ	τ
$\hat{\sigma}^2(pi:s)$	$n'_i n'_s$	$\hat{\sigma}^2(pI:S)$	Δ, δ	Δ, δ

TABLE 11

D Studies for Design $(p{:}c) \times i$
With Person Nested Within Class $(p{:}c)$ as the
Object of Measurement

Estimated G Study Variance Components	D Study Sampling Frequency	Estimated D Study Variance Components	$(p{:}c\|p\|c\|I)$
$\hat{\sigma}^2(p{:}c)$	1	$\hat{\sigma}^2(p{:}c)$	τ
$\hat{\sigma}^2(c)$	1	$\hat{\sigma}^2(c)$	
$\hat{\sigma}^2(i)$	n'_i	$\hat{\sigma}^2(I)$	Δ
$\hat{\sigma}^2(ci)$	n'_i	$\hat{\sigma}^2(cI)$	Δ
$\hat{\sigma}^2(pi{:}c)$	n'_i	$\hat{\sigma}^2(pI{:}c)$	Δ,δ

TABLE 12

D Studies for Design $(p{:}c) \times (i{:}s{:}t)$
With Person Nested Within Class $(p{:}c)$ as the
Object of Measurement

Estimated G Study Variance Components	D Study Sampling Frequency	Estimated D Study Variance Components	$(p{:}c\|p\|c\|I,S,T)$	$(p{:}c\|p\|c,S,T\|I)$
$\hat{\sigma}^2(p{:}c)$	1	$\hat{\sigma}^2(p{:}c)$	τ	τ
$\hat{\sigma}^2(c)$	1	$\hat{\sigma}^2(c)$		
$\hat{\sigma}^2(i{:}s{:}t)$	$n'_i n'_s n'_t$	$\hat{\sigma}^2(I{:}S{:}T)$	Δ	Δ
$\hat{\sigma}^2(s{:}t)$	$n'_s n'_t$	$\hat{\sigma}^2(S{:}T)$	Δ	
$\hat{\sigma}^2(ct)$	n'_t	$\hat{\sigma}^2(cT)$	Δ	
$\hat{\sigma}^2(t)$	n'_t	$\hat{\sigma}^2(T)$	Δ	
$\hat{\sigma}^2(cs{:}t)$	$n'_s n'_t$	$\hat{\sigma}^2(cS{:}T)$	Δ	
$\hat{\sigma}^2(ci{:}s{:}t)$	$n'_i n'_s n'_t$	$\hat{\sigma}^2(cI{:}S{:}T)$	Δ	Δ
$\hat{\sigma}^2(pt{:}c)$	n'_t	$\hat{\sigma}^2(pT{:}c)$	Δ, δ	τ
$\hat{\sigma}^2(ps{:}ct)$	$n'_s n'_t$	$\hat{\sigma}^2(pS{:}cT)$	Δ, δ	τ
$\hat{\sigma}^2(pi{:}sc{:}t)$	$n'_i n'_s n'_t$	$\hat{\sigma}^2(pI{:}Sc{:}T)$	Δ, δ	Δ, δ

TABLE 13

D Studies for Design $(p:c) \times i$
With Class (c) as the Object of Measurement

Estimated G Study Variance Components	D Study Sampling Frequency	Estimated D Study Variance Components	$(c\|c\|-\|P,I)$	$(c\|c\|P\|I)$	$(c\|c\|I\|P)$
$\sigma^2(p:c)$	n'_p	$\sigma^2(P:c)$	Δ,δ	τ	Δ,δ
$\sigma^2(c)$	1	$\sigma^2(c)$	τ	τ	τ
$\sigma^2(i)$	n'_i	$\sigma^2(I)$	Δ	Δ	
$\sigma^2(ci)$	n'_i	$\sigma^2(cI)$	Δ,δ	Δ,δ	τ
$\sigma^2(pi:c)$	$n'_p n'_i$	$\sigma^2(PI:c)$	Δ,δ	Δ,δ	Δ,δ

TABLE 14

D Studies for Design $(p{:}c) \times (i\ s{:}t)$
With Class (c) as the Object of Measurement

Estimated G Study Variance Components	D Study Sampling Frequency	Estimated D Study Variance Components	$(c\mid c\mid \cdot \mid P,I,S,T)$	$(c\mid c\mid S,T\mid P,I)$
$\hat\sigma^2(p{:}c)$	n'_p	$\hat\sigma^2(P{:}c)$	Δ,δ	Δ,δ
$\hat\sigma^2(c)$	1	$\hat\sigma^2(c)$	τ	τ
$\hat\sigma^2(i{:}s{:}t)$	$n'_i n'_s n'_t$	$\hat\sigma^2(I{:}S{:}T)$	Δ	Δ
$\hat\sigma^2(s{:}t)$	$n'_s n'_t$	$\hat\sigma^2(S{:}T)$	Δ	
$\hat\sigma^2(t)$	n'_t	$\hat\sigma^2(T)$	Δ	
$\hat\sigma^2(ct)$	n'_t	$\hat\sigma^2(cT)$	Δ,δ	τ
$\hat\sigma^2(cs{:}t)$	$n'_s n'_t$	$\hat\sigma^2(cS{:}T)$	Δ,δ	τ
$\hat\sigma^2(ci{:}s{:}t)$	$n'_i n'_s n'_t$	$\hat\sigma^2(cI{:}S{:}T)$	Δ,δ	Δ,δ
$\hat\sigma^2(pt{:}c)$	$n'_p n'_t$	$\hat\sigma^2(PT{:}c)$	Δ,δ	Δ,δ
$\hat\sigma^2(ps{:}ct)$	$n'_p n'_s n'_t$	$\hat\sigma^2(PS{:}cT)$	Δ,δ	Δ,δ
$\hat\sigma^2(pi{:}sc{:}t)$	$n'_p n'_i n'_s n'_t$	$\hat\sigma^2(PI{:}Sc{:}T)$	Δ,δ	Δ,δ

TABLE 15

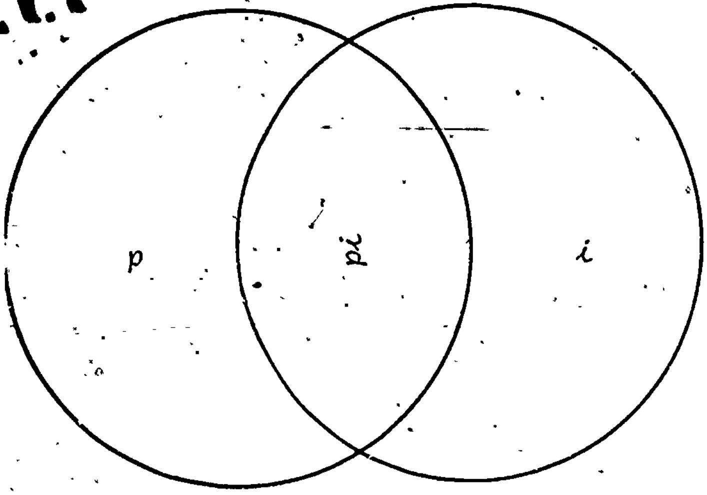
G Study and D Study Variance Components
for the Design $p \times i \times o$, with Person (p) as the Object of Measurement
when Items are Sampled from a Finite Universe ($N_i = N'_i < \infty$)

Estimated G Study Variance Components for Random Sampling	D Study Sampling Frequency	Finite Universe Correction	Estimated D Study Varaince Components for Random Sampling
$\hat{\sigma}^2(p \mid N_i < \infty) = \hat{\sigma}^2(p) + \hat{\sigma}^2(pi)/N_i$	1	1	$\hat{\sigma}^2(p \mid N_i < \infty) = \hat{\sigma}^2(p) + \hat{\sigma}^2(pi)/N_i$
$\hat{\sigma}^2(i \mid N_i < \infty) = \hat{\sigma}^2(i)$	n'_i	$(1 - n'_i/N_i)$	$\hat{\sigma}^2(I \mid N_i < \infty) = (1 - n'_i/N_i)\hat{\sigma}^2(i)/n'_i$
$\hat{\sigma}^2(o \mid N_i < \infty) = \hat{\sigma}^2(o) + \hat{\sigma}^2(io)/N_i$	n'_o	1	$\hat{\sigma}^2(O \mid N_i < \infty) = \hat{\sigma}^2(o)/n'_o + \hat{\sigma}^2(io)/N_i n'_o$
$\hat{\sigma}^2(pi \mid N_i < \infty) = \hat{\sigma}^2(pi)$	n'_i	$(1 - n'_i/N_i)$	$\hat{\sigma}^2(pI \mid N_i < \infty) = (1 - n'_i/N_i)\hat{\sigma}^2(pi)/n'_i$
$\hat{\sigma}^2(po \mid N_i < \infty) = \hat{\sigma}^2(po) + \hat{\sigma}^2(pio)/N_i$	n'_o	1	$\hat{\sigma}^2(pO \mid N_i < \infty) = \hat{\sigma}^2(po)/n'_o + \hat{\sigma}^2(pio)/N_i n'_o$
$\hat{\sigma}^2(io \mid N_i < \infty) = \hat{\sigma}^2(io)$	$n'_i n'_o$	$(1 - n'_i/N_i)$	$\hat{\sigma}^2(IO \mid N_i < \infty) = (1 - n'_i/N_i)\hat{\sigma}^2(io)/n'_i n'_o$
$\hat{\sigma}^2(pio \mid N_i < \infty) = \hat{\sigma}^2(pio)$	$n'_i n'_o$	$(1 - n'_i/N_i)$	$\hat{\sigma}^2(pIO \mid N_i < \infty) = (1 - n'_i/N_i)\hat{\sigma}^2(pio)/n'_i n'_o$

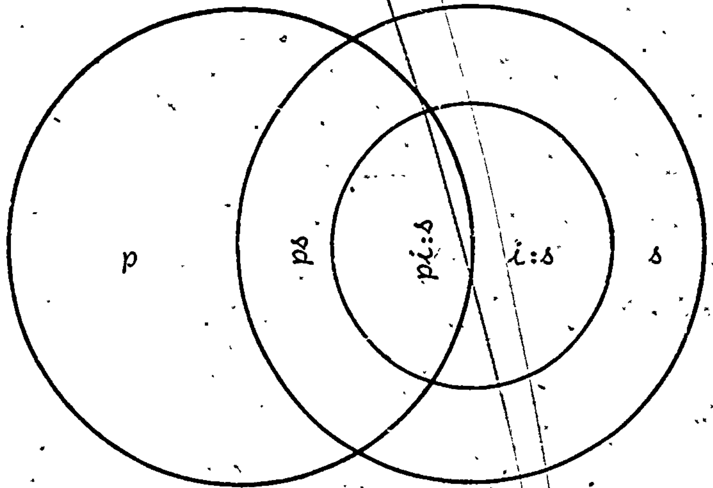
Figure Captions

Figure 1. Venn diagrams for five illustrative designs.

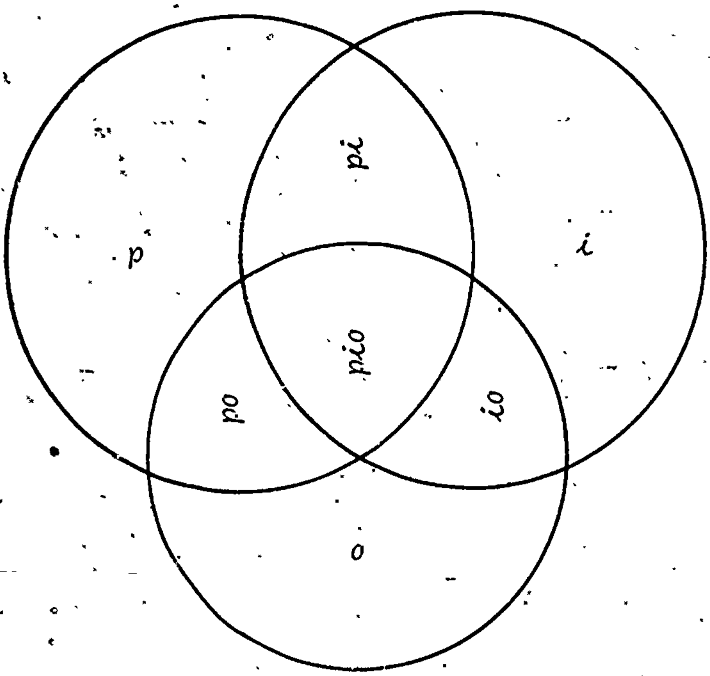
Figure 2. Decomposition of three variance components, for the random effects model, in terms of mean squares for the design (p:c) x i.



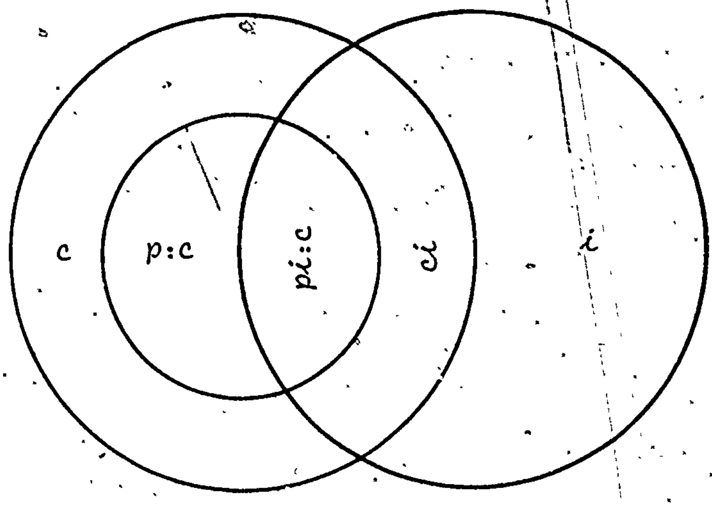
Design $p \times i$



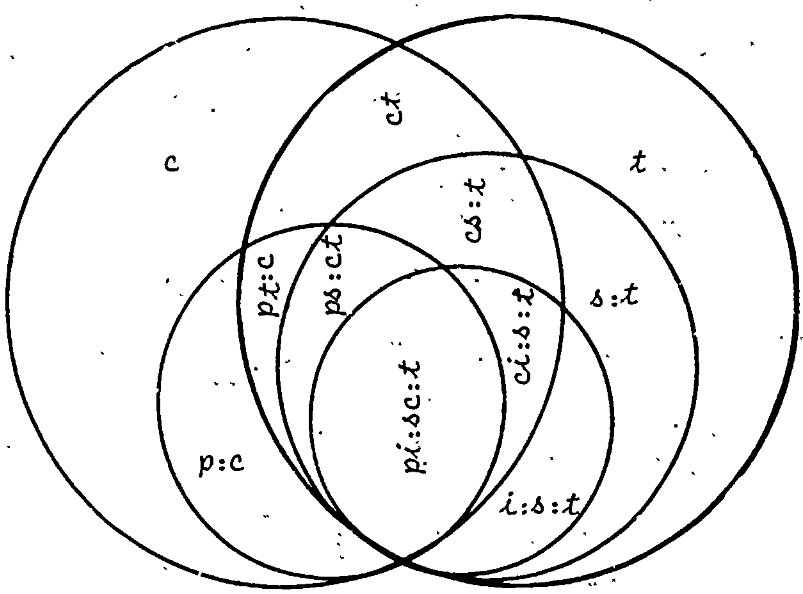
Design $p \times (i{:}s)$



Design $p \times i \times o$



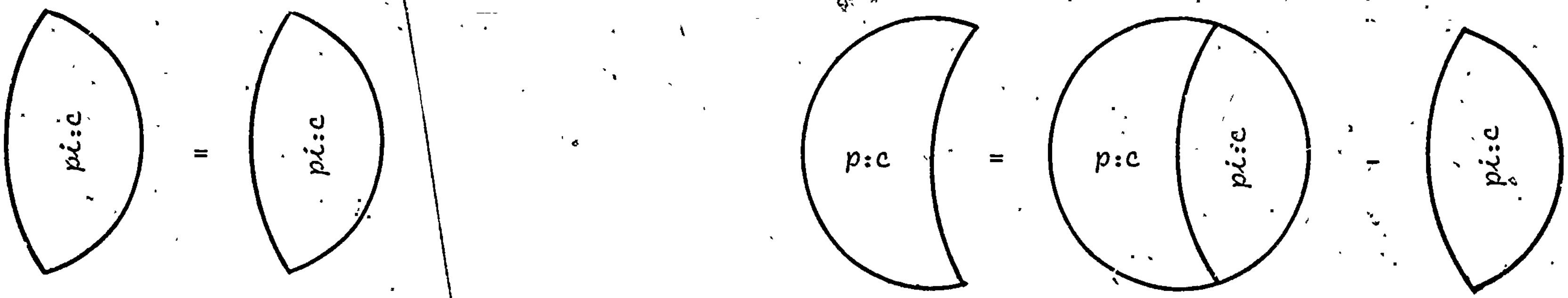
Design $(p{:}c) \times i$



Design $(p{:}c) \times (i{:}s{:}t)$

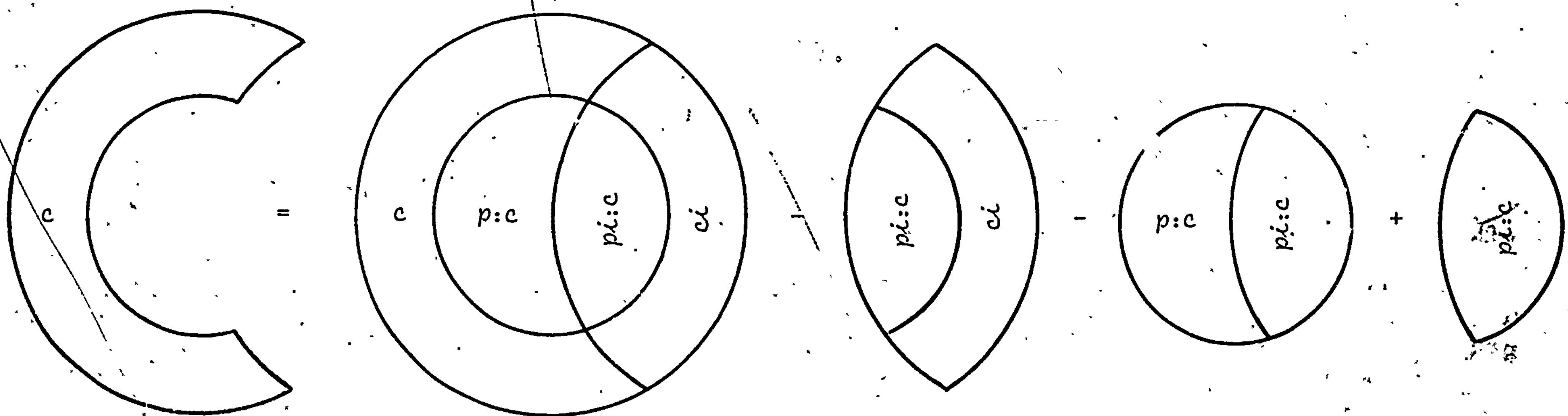
FIGURE i

FIGURE 2



$$\sigma^2(pi:c) = MS(pi:c)$$

$$n_i\sigma^2(p:c) = MS(p:c) - MS(pi:c)$$



$$n_p n_i \sigma^2(c) = MS(c) - MS(ci) - MS(p:c) + MS(pi:c)$$