Fall 2010

Estimating standard errors of estimated variance components in generalizability theory using bootstrap procedures

Joann Lynn Moore
University of Iowa

ESTIMATING STANDARD ERRORS OF ESTIMATED VARIANCE COMPONENTS IN GENERALIZABILITY THEORY USING BOOTSTRAP PROCEDURES

by

Joann Lynn Moore

An Abstract

Of a thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychological and Quantitative Foundations
(Educational Measurement and Statistics)
in the Graduate College of
The University of Iowa

December 2010

Thesis Supervisors: Associate Professor Robert D. Ankenmann
Professor Robert L. Brennan

ABSTRACT

This study investigated the extent to which rules proposed by Tong and Brennan (2007) for estimating standard errors of estimated variance components held up across a variety of G theory designs, variance component structures, sample size patterns, and data types. Simulated data was generated for all combinations of conditions, and point estimates, standard error estimates, and coverage for three types of confidence intervals were calculated for each estimated variance component and relative and absolute error variance across a variety of bootstrap procedures for each combination of conditions. It was found that, with some exceptions, Tong and Brennan's (2007) rules produced adequate standard error estimates for normal and polytomous data, while some of the results differed for dichotomous data. Additionally, some refinements to the rules were suggested with respect to nested designs. This study provides support for the use of bootstrap procedures for estimating standard errors of estimated variance components when data are not normally distributed.

Abstract Approved: ____
Thesis Supervisor

____
Title and Department

____
Date

____
Thesis Supervisor

____
Title and Department

____
Date

ESTIMATING STANDARD ERRORS OF ESTIMATED VARIANCE COMPONENTS IN GENERALIZABILITY THEORY USING BOOTSTRAP PROCEDURES

by

Joann Lynn Moore

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychological and Quantitative Foundations
(Educational Measurement and Statistics)
in the Graduate College of
The University of Iowa

December 2010

Thesis Supervisors: Associate Professor Robert D. Ankenmann
Professor Robert L. Brennan

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Joann Lynn Moore

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Psychological and Quantitative Foundations (Educational Measurement and Statistics) at the December 2010 graduation.

Thesis Committee: ___________________________________
Robert D. Ankenmann, Thesis Supervisor

____
Robert L. Brennan, Thesis Supervisor

____
Deborah J. Harris

____
Michael J. Kolen

____
Mary Kathryn Cowles

ACKNOWLEDGMENTS

I would like to thank Drs. Deborah Harris and Robert Brennan for their help in formulating my dissertation topic, and Dr. Robert Ankenmann for helping me clarify my ideas and the flow of the text to make this an interesting and comprehensible dissertation. I would also like to thank Dr. Ye Tong, whose research provided me with an interesting dissertation topic, and whose C code was an immense help in writing all of the programs required for this dissertation. Last but not least, I would like to thank my family and friends for their love, advice, and support; I could not have done this without you.

ABSTRACT

This study investigated the extent to which rules proposed by Tong and Brennan (2007) for estimating standard errors of estimated variance components held up across a variety of G theory designs, variance component structures, sample size patterns, and data types. Simulated data was generated for all combinations of conditions, and point estimates, standard error estimates, and coverage for three types of confidence intervals were calculated for each estimated variance component and relative and absolute error variance across a variety of bootstrap procedures for each combination of conditions. It was found that, with some exceptions, Tong and Brennan's (2007) rules produced adequate standard error estimates for normal and polytomous data, while some of the results differed for dichotomous data. Additionally, some refinements to the rules were suggested with respect to nested designs. This study provides support for the use of bootstrap procedures for estimating standard errors of estimated variance components when data are not normally distributed.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Generalizability theory (G theory) provides a framework for differentiating among various sources of observed score variance, including variance due to the objects of measurement and variance due to the conditions of measurement. For example, in a testing context there can be variance in observed scores attributable to the population of examinees, and there can also be multiple sources of error variance (e.g., test forms, items, raters, occasions, etc.). The decomposition of total observed score variance into these differentiated sources of variance, or variance components, is the backbone of G theory.

Generalizability theory is an extension of classical test theory (CTT). Both G theory and CTT can be used to make inferences, for example, about the optimal number of items to be included in a test to achieve a given level of score reliability. However, a unique contribution of G theory is that it allows for multiple sources of error variance to be investigated simultaneously, as well as interactions between the variance components, whereas CTT only allows for a single, undifferentiated error term.

A full explanation of G theory requires a consideration of both: (a) a universe of admissible observations (UAO) and generalizability studies (G studies); and (b) a universe of generalization (UG) and decision studies (D studies). A UAO consists of those facets (e.g., items, raters, occasions, etc.) that an investigator considers relevant for a given measurement procedure. The purpose of a G study is to estimate the variance components associated with the facets in the UAO as well as the objects of measurement "facet," typically persons. It is important to understand that the estimated variance components are for single conditions of facets. These estimated variance components, which are the focus of most of this dissertation, are the statistics that are considered in statistical books such as *Variance Components* by Searle, Casella, and McCulloch

(1992). More specifically, this dissertation focuses on the so-called ANOVA estimates of variance components, as discussed in Chapter 2.

By contrast, a UG and a D study are associated with mean scores over investigator-specified numbers of conditions of facets. Conceptually, a UG consists of randomly parallel instances of a measurement procedure. For example, suppose that: (a) examinees (the objects of measurement) were each administered the same two essay prompts; (b) each prompt was scored by the same two raters; and (c) each examinee's score was his or her mean score over the four ratings. Each randomly parallel (usually hypothetical) instance of such a measurement procedure would consist of a different set of two prompts and a different set of two raters, provided both prompts and raters are defined as random by the investigator (so-called fixed facets are possible, but outside the scope of this dissertation). Furthermore, for this example, the D study would be $p \times I \times H$, where "*p*" stands for examinees or persons, "*I*" stands for mean scores over essay prompts, and "*H*" stands for mean scores over raters. The variance components for the D study are almost always estimated using the G study estimated variance components. Certain combinations of the resulting D study estimated variance components give estimates of one or more types of error variances (i.e., relative and absolute error variance) that are associated with the mean scores for examinees. These error variances are a secondary focus of this dissertation.

It is particularly important to note that in this dissertation, the word "error" has multiple meanings. The primary focus of this dissertation is on estimated standard *errors* of estimated variance components in the G study sense. Additionally, this dissertation is concerned with the estimated standard *errors* of certain types of estimated *error* variances in a D study. It should be clear from the context how the word "error" is to be interpreted.

Statement of the Problem

Several approaches have been used to estimate standard errors of estimated variance components. A common approach makes use of a theoretical model based on assumptions that the score effects (components of the observed scores that can be attributed to the various objects and conditions of measurement, analogous to treatment effects in an analysis of variance (ANOVA) context) are normally distributed, random and independent (Searle, 1971; Searle et al., 1992). However, it is often unrealistic to assume normality of score effects, particularly in a multiple choice testing environment, where the data are scored dichotomously (i.e., correct or incorrect), or in writing or performance assessments, where the data are often scored polytomously (i.e., finite integers such as a 6-point scale from 0 to 5) (Brennan, 2001). Therefore, other approaches have been proposed for estimating standard errors of estimated variance components, such as bootstrap and jackknife resampling methods, which do not require assumptions about the distributions of the score effects.

Bootstrap resampling involves taking repeated independent random samples with replacement from a dataset and estimating the variance components for each bootstrap sample. The standard deviations of the estimated variance components across bootstrap samples are estimates of the standard errors of the estimated variance components.

Jackknife resampling involves sampling without replacement. Briefly, the usual jackknife steps for estimating the standard error of some estimate of a parameter θ based on s data points are: (a) obtain the statistic of interest $\hat{\theta}$ for all s data points; (b) obtain the s estimates of θ that result from deleting each one of the data points, designating these estimates as $\hat{\theta}_{-j}$; (c) using $\hat{\theta}$ and $\hat{\theta}_{-j}$, obtain statistics called pseudovalues; and (d) the jackknife estimate of the standard error of $\hat{\theta}$ is the standard error of the mean of the pseudovalues. As discussed by Brennan (2001, pp. 182-185), the jackknife steps are considerably more complicated when estimating standard errors of estimated variance components for the types of designs used in G theory. In particular, the pseudovalues are much more complicated. Furthermore, even with today's computers, the jackknife can be extraordinarily computationally intensive for estimating standard errors of estimated variance components in G theory. For these reasons, among others, the focus of this dissertation is on bootstrap methods.

Use of the bootstrap for estimating standard errors of estimated variance components is complicated by the fact that G theory designs usually involve multiple facets. For a given G theory design, several bootstrap procedures could be used to estimate each variance component and its standard error, thus decisions need to be made regarding which procedure or procedures to use. These procedures differ with respect to which facet or facets are resampled. Different bootstrap procedures yield better estimates of different variance components and their standard errors, depending on which facets are resampled, and no single bootstrap procedure yields accurate estimates of standard errors for all estimated variance components in a given design (Brennan, Harris, & Hanson, 1987; Tong & Brennan, 2007). For example, in the simplest G theory design, a group of students take a test consisting of several items. The resulting data are arranged in a matrix of scores, one score for each student's response to each item. Various bootstrap procedures could be used to estimate the standard errors of the estimated variance components. In this case, three of the possible procedures involve resampling the students but not the items, resampling the items but not the students, or resampling both students and items. Each of these bootstrap procedures will produce slightly different estimates, and the bootstrap procedure that produces the optimal estimates for one variance component and its standard error may not be the same bootstrap procedure that produces the optimal estimates for another variance component and its standard error. Because of this, rules for deciding which bootstrap procedure to use to estimate each of the variance components and its standard error have been proposed (Tong & Brennan, 2007). However, the rules are somewhat ad hoc at this point and have only been tested in a limited number of studies. Further research is needed to investigate the robustness of these rules across a variety of G theory designs, with varied sample sizes and varied parameter values for the variance components.

In summary, there are multiple ways of obtaining estimates of standard errors of estimated variance components, and there is no single best approach or correct answer to the question of which approach to use. There are reasons for preferring one method over another; for example, if there is reason to suspect that the data violate the assumptions of a particular theoretical approach, then the bootstrap may be preferred. However, even if one decides to use the bootstrap to estimate the standard errors, further decisions are required about which bootstrap procedures to use. Previous studies have provided some guidance, but more research is needed to ensure that the proposed rules apply in a broader context than what has been studied thus far. This dissertation contributes to the existing research by evaluating Tong and Brennan's (2007) proposed rules in a wider range of G theory designs, with several combinations of variance component structures, sample sizes, and data types.

Significance of the Study

The error associated with estimates of variance components is an important topic because the results of G studies and D studies are used to make inferences, for example, about examinees or test instruments. The accuracy of these estimates is particularly important in high-stakes situations such as selection, placement, certification, diagnosis, or monitoring for accountability purposes. The accuracy of estimated variance components must be considered whenever results of studies based on G theory are interpreted and decisions are made based on those interpretations. Examinees could be negatively impacted if high-stakes decisions are made about them based on inaccurate estimates of variance components. Additionally, testing programs could be impacted by inaccurate estimates of variance components, for example, by wasting resources if the number of raters chosen to score essay responses is more than necessary, or by having less reliable test scores than anticipated if the number of raters chosen is too few.

As early as Cronbach, Gleser, Nanda, and Rajaratnam's (1972) seminal book on G theory, concerns were raised that the accuracy of estimates of variance components is important for drawing sound conclusions from research in educational and social settings. Additionally, the *Standards for Educational and Psychological Testing* (American Psychological Association, 1985; American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) state that reporting standard errors of estimates is important. The previous revision of the Standards (American Psychological Association, 1985) explicitly addressed issues of precision of variance components, saying "Not all sources of error are expected to be relevant for a given test. Thus the estimation of clearly labeled components of observed and error score variance is a particularly useful outcome of a reliability study, both for the test developer who wishes to improve the reliability of an instrument and for the user who wants to interpret test scores in particular circumstances with maximum understanding. Reporting standard errors, confidence intervals, or other measures of imprecision of estimates is also helpful" (p. 19). In the current revision of the Standards (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), Standard 2.1 states that "For each total score, subscore, or combination of scores that is interpreted, estimates of relevant reliabilities and standard errors of measurement or test information functions should be reported" (p. 31). This particular reference is to test scores rather than variance components, but it is clear from the tone of both revisions that all estimated statistics should be presented in conjunction with information regarding their precision if we are to make sound judgments based on those estimates.

Standard 7.9 addresses the consequences of using estimates to inform policy such as the *No Child Left Behind* legislation (NCLB, 2002), stating: "When tests or assessments are proposed for use as instruments of social, educational, or public policy, the test developers or users proposing the test should fully and accurately inform policymakers of the characteristics of the tests as well as any relevant and credible information that may be available concerning the likely consequences of test use" (p. 83). This underscores the importance of taking into account the errors associated with estimates of variance components, because how the results of G theory studies are used can have important consequences. Unfortunately, it appears that some researchers utilizing G theory methods overlook the fact that the estimated variance components are estimates and thus contain error. This could result in erroneous conclusions if the researchers have too much confidence in the estimated variance components and reliability estimates resulting from their studies.

A July 2009 search of the Educational Resources Information Center (ERIC) database for recent articles containing the keywords "generalizability theory" yielded several articles that used a G theory framework for disentangling various sources of error variance. The articles examined various types of assessments, such as high-stakes medical certification exams (Burch, Norman, Schmidt, & van der Vleuten, 2008; Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams, 2009; Harik, Clauser, Grabovsky, Nungester, Swanson, & Nandakumar, 2009; Murphy, Bruce, Mercer, & Eva, 2009), psychological testing (Follesdal & Hagtvet, 2009), which could have important consequences for subjects diagnosed based on test results, and test development and validation (Chafouleas, Christ, & Riley-Tillman, 2009; Solano-Flores & Li, 2009), which could have important consequences for test developers as well as test users. Most of the articles lacked any acknowledgement that the estimated variance components might contain error, and that this error should be taken into account when interpreting the results. Only one of the reviewed articles contained an explicit caution regarding the lack of available information on the standard errors of variance components estimates in their study (Solano-Flores & Li, 2009). It appears that further work needs to be done in terms of bringing this issue to the forefront so that researchers using a G theory framework are aware of it and take the accuracy of estimated variance components into account in the conclusions drawn from their research.

A further complication of the issue of the accuracy of estimated variance components involves the assumptions upon which the most common method of calculating standard errors of estimated variance components rests. Of the surveyed studies that did report standard errors of variance component estimates, the standard errors they presented tended to be based on Searle's (1971) equations that assume multivariate normality, while the data analyzed in the studies tended to be polytomously scored, making the normality assumption questionable (Chafouleas, Christ, & Riley-Tillman, 2009; Gagnon, Charlin, Lambert, Carrière, & Van der Vleuten, 2009; Huang, 2008; Lee & Kantor, 2007). It has been shown that the estimated standard errors obtained using these equations are inaccurate when the data are not normally distributed (Brennan et al., 1987). Additionally, although the authors of these studies presented estimates of standard errors of estimated variance components in their tables of results, they did not make any mention of them in the text of their papers. The failure of the authors of these studies to discuss the impact of the error associated with the estimated variance components is troublesome, because it is unclear whether they took the error into account when interpreting their results, regardless of the adequacy of those standard error estimates. Researchers should have a sense of the variability of estimated variance components when drawing conclusions about the relative contributions of various facets to the total error variance; for example, concluding that a larger proportion of variance was due to the raters facet than the items facet. When the error associated with the estimated variance components is not taken into account, one can have little confidence that the differences between variance components estimates are due to actual differences in the amount of variance contributed by the various facets or merely due to sampling error.

In summary, the precision of estimated variance components is an important topic. It is important for researchers to be aware that estimated variance components contain error, and be able to accurately quantify the amount of error associated with the estimated variance components. Evaluating the precision of estimated variance components is important in terms of validity and score interpretation. Messick (1989) defined validity as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *inferences* and *actions* based on test scores or other modes of assessment" (p. 13). If researchers are overly confident in the accuracy of their estimates of variance components (i.e., the "evidence"), they may draw inaccurate conclusions regarding the reliability of scores obtained from their instrument, thus putting the validity of the inferences made from the resulting test scores at risk.

Research Questions

The main research question of this study was whether the rules proposed by Tong and Brennan (2007) for choosing which bootstrap procedures to use to estimate each of the variance components and its standard error hold up across various G theory designs, variance component structures, sample size patterns, and data types. The main research question can be further broken down into the following questions:

1. How well do the rules work across various G theory designs, including the $p \times i$, $p \times i \times h$, $i : p$, $i : h : p$, and $p \times (i : h)$ designs?
2. How well do the rules work when the variance components for the various effects are relatively similar to one another (i.e., each variance component contributes a similar amount of variance to the total variance)? When they are relatively dissimilar (i.e., most of the total variance is contained within a smaller subset of variance components)?

3. How well do the rules work when the sample sizes for the various effects are similar to one another (i.e., the number of persons is similar to the number of items and/or raters)? When the sample sizes are dissimilar to one another?
4. How well does the workaround for the bootstrap estimate of the standard error of the absolute error variance [$SE(\Delta)$] proposed by Tong and Brennan (2007) hold up across various G theory designs, variance component structures, sample size patterns, and data types?
5. To what extent do standard normal, percentile, and bias corrected percentile confidence intervals adequately capture the variability of estimated variance components and relative and absolute error variances? Is one type of confidence interval preferred to the others?

The research questions were answered by simulating data using various G theory designs, variance component structures, sample size patterns, and data types, and comparing the obtained estimates of variance components and their associated standard errors to the parameter values used to simulate the data. If the rules hold up, then the estimates should be close to the parameter values. Additionally, measures of confidence interval coverage were obtained to further assess the variability of the estimates of the variance components in relation to the parameter values.

Hypotheses

It was expected that different bootstrap procedures would produce optimal estimates of different variance components and their standard errors, as found in Tong and Brennan (2007). Specifically, it was expected that bootstrap procedures that resample along a given facet would produce better estimates for the main effect of that facet (e.g., resampling persons would produce better estimates for persons and resampling items will produce better estimates for items), and resampling along the facet with a larger sample size would produce better estimates for the interactions (e.g., resampling persons would produce better estimates for the person by item interaction if there are more persons than items). For nested designs, it was expected that bootstrap procedures that resample along the primary, or nested facet would produce better estimates for nested main effects (e.g., resampling items will produce better estimates for item nested within persons), and bootstrap procedures that resample along the nested facet with a larger sample size would produce better estimates for nested interactions (e.g., resampling persons would produce better estimates for the persons and items nested within raters interaction if there were more persons than items). Additionally, estimates were expected to improve as sample size increased (Othman, 1995; Tong & Brennan, 2007).

It was expected that bias corrected percentile confidence intervals would produce confidence interval coverage that was closer to the expected coverage (i.e., 90% of a 90% confidence interval) than percentile confidence intervals, and percentile confidence intervals were expected to outperform standard normal confidence intervals. It was also expected that confidence interval coverage would be closer to the expected coverage for those bootstrap procedures that produced better estimates of standard errors of variance components than those bootstrap procedures that produced poor estimates of standard errors of variance components.

CHAPTER 2: LITERATURE REVIEW

Overview of Generalizability Theory

The historical roots of measurement theory lie in *classical test theory* (CTT), which is a "body of assumptions and derived results" that aid in quantifying properties of test scores such as reliability and measurement error (Feldt & Brennan, 1989, p. 108). Classical test theory defines an observed test score X as the sum of an unobservable true score T and error E:

$$X = T + E \qquad (2.1)$$

Classical test theory acknowledges the existence of various sources of error, but does not quantify the relative contributions of various sources of error, and therefore only makes use of a single undifferentiated error term. *Generalizability theory* (G theory) can be viewed as an extension of CTT in that G theory involves separating out various sources of error. Therefore, the G theory definition of an observed test score is the sum of an unobservable true score T and multiple error components, each denoted E_i:

$$X = T + E_1 + E_2 + \ldots + E_k \qquad (2.2)$$

In G theory, *objects of measurement* are the entities that we are interested in measuring. In most testing contexts, the objects of measurement are students, but they could also be other entities such as classrooms or schools. *Universes of admissible observations* are collections of *facets* associated with measurement conditions (e.g., items, test forms, raters, etc.). Objects of measurement, universes of admissible observations, and facets must be defined at the beginning of any G study.

In general, the objects of measurement are not technically considered to be facets, but they are not different mathematically from the other facets in terms of analysis, and are sometimes referred to as facets. It should also be noted that we generally have samples of students and raters from larger populations of students and raters, and samples of items from larger universes of items, and we wish to generalize beyond our specific samples to the populations and universes of interest. Therefore, the objects of measurement and facets are considered *random*, because upon replication, we would likely use different samples of students, items, and raters. If we were not interested in generalizing beyond our sample for a given facet, we would consider that facet to be *fixed*. For example, if we have a test made up of content areas such as math, English, and science, and those content areas will always remain the same across test forms, occasions, students, etc., then the content area facet is fixed. This study dealt with random facets only.

The most conceptually simple G theory design is a single-facet crossed design, which can be written symbolically as $p \times i$, in which the objects of measurement (e.g., persons, or *p*) are *crossed* with a single facet (e.g., items, or *i*). This means that each person in a sample of persons is administered each item in a test, with both persons and items being randomly drawn from theoretically infinite populations of persons and items. With this design, there are three separate variance components due to persons, items, and the interaction between persons and items (which includes all residual error).

Another common design is the $p \times i \times h$ design, which is a two-facet crossed design; for example, every person (*p*) is administered every item (*i*), and every item response is scored by every rater (*h*). This design is similar to the $p \times i$, except that the addition of another facet introduces four additional variance components due to raters, the interaction between persons and raters, the interaction between items and raters, and a three-way interaction among persons, items, and raters. In this case, the three-way interaction contains the residual error.

Designs can also be *nested*, such that not every object of measurement is exposed to every measurement condition. A simple single-facet nested design is the $i : p$ design; for example, items (*i*) are nested within persons (*p*). This means that each person is administered a different set of items. In this design, there are two variance components due to persons and items nested within persons. The items nested within persons variance component is confounded with the residual error.

Nested designs can have more than one level of nesting. One example of a two-facet nested design is the $i : h : p$ design, where items (*i*) are nested within raters (*h*) which are nested within persons (*p*). This means that each person takes a different subset of items, each person's responses are rated by a different subset of raters, and each rater rates a different subset of item responses within the subset of items taken by a given person. This might be the case in a medical licensure test in which each examinee is presented with a different series of patient scenarios, and responses to each scenario are evaluated by a different examiner. This design can be decomposed into three variance components due to persons, raters nested within persons, and items nested within raters nested within persons, which also contains the residual error.

Another common design is the $p \times (i : h)$ design, such as when persons (*p*) are crossed with items (*i*) nested within raters (*h*). A common example of this design is when items are nested within raters such that all students respond to each item but each rater only rates a subset of the item responses, meaning that across a given student's responses, some ratings from every rater are represented. This design illustrates some of the complexities that can arise in G theory, as this design involves crossed as well as nested effects. For the $p \times (i : h)$ design, there are five separate variance components due to persons, raters, the interaction between persons and raters, items nested within raters, and the interaction between persons and items nested within raters, which also contains the residual error.

Generalizability Theory Notation and Equations

Using the notation of Brennan (2001), an observed score in G theory can be defined in terms of a linear model,

$$X_{pih} = \mu + \sum \nu_{\alpha} \tag{2.3}$$

where μ is the grand mean and v_α corresponds to the various *score effects* (α is used generically throughout this manuscript to denote any score effect). For example, for the $p \times i \times h$ design, the linear model is

$$X_{pih} = \mu + v_p + v_i + v_h + v_{pi} + v_{ph} + v_{ih} + v_{pih} \quad (2.4)$$

The linear models for each of the designs of interest in this study are presented in Appendix A, Table A1.

The $p \times i \times h$ design has seven score effects. There are three main effects: v_p, the person effect; v_i, the item effect; and v_h, the rater effect. There are also four interaction effects: v_{pi}, the interaction between persons and items; v_{ph}, the interaction between persons and raters; v_{ih}, the interaction between items and raters; and v_{pih}, the three-way interaction among persons, items, and raters, which also contains the residual error. Score effects are linear combinations of mean scores. For example, the seven score effects for the $p \times i \times h$ design are displayed in Table 1, where μ_α is the expected value (over the universe of admissible observations) of a person's observed score across the facets *not* in α. The score effects for each of the designs of interest in this study are presented in Appendix A, Tables A2 – A6.

Table 1. Score Effects for a $p \times i \times h$ Design

Effect (α)	Score effect v_α	
p	$\mu_p - \mu$	(2.5)
i	$\mu_i - \mu$	(2.6)
h	$\mu_h - \mu$	(2.7)
pi	$\mu_{pi} - \mu_p - \mu_i + \mu$	(2.8)
ph	$\mu_{ph} - \mu_p - \mu_h + \mu$	(2.9)
ih	$\mu_{ih} - \mu_i - \mu_h + \mu$	(2.10)
pih	$X_{pih} - \mu_{pi} - \mu_{ph} - \mu_{ih} + \mu_p + \mu_i + \mu_h - \mu$	(2.11)

Because score effects are assumed to be uncorrelated, the variance of observed scores can be decomposed into separate variances attributable to the various effects. The decomposition of observed score variance into variance components for each of the designs of interest is presented in Appendix A, Table A7. For example, for the $p \times i \times h$ design:

$$\sigma^2(X_{pih}) = \sigma^2(p) + \sigma^2(i) + \sigma^2(h) + \sigma^2(pi) + \sigma^2(ph) + \sigma^2(ih) + \sigma^2(pih) \tag{2.12}$$

These variances corresponding to the various effects are called *variance components*. Each variance component can be defined in terms of expected values of squared deviations or score effects. For example, the definitions of the seven variance components in the $p \times i \times h$ design are given in Table 2. The definitions of the variance components for the other designs of interest are presented in Appendix A, Tables A8 – A12. Additionally, the overall observed score variance is defined as

$$\sigma^2(X_{pih}) = \underset{p}{\mathrm{E}}\,\underset{i}{\mathrm{E}}\,\underset{h}{\mathrm{E}}\left(X_{pih} - \mu\right)^2 \tag{2.13}$$

In practice, the parameter values and expected values are unknown; therefore, G theory typically uses analysis of variance (ANOVA) procedures to compute terms that can be used to obtain estimates of variance components. Unlike ANOVA, however, G theory is typically not concerned with tests of statistical significance, but employs ANOVA sums of squares and mean squares to obtain estimates of variance components. The variance components can be estimated from data by calculating the ANOVA mean squares [$MS(\alpha)$] and sample sizes (n_α), and then attaining certain linear combinations of mean squares. For example, for a $p \times i \times h$ design, the equations in Table 3 can be used to estimate each of the variance components. Tables A13 – A17 in Appendix A contain equations for calculating each of the variance components estimates for all of the designs of interest in this study.

Table 2. Definitions of Variance Components for a $p \times i \times h$ Design

Effect (α)	Variance component definitions $\sigma^2(\alpha)$	
p	$\underset{p}{\mathrm{E}}(\mu_p - \mu)^2 = \underset{p}{\mathrm{E}}(\nu_p^2)$	(2.14)
i	$\underset{i}{\mathrm{E}}(\mu_i - \mu)^2 = \underset{i}{\mathrm{E}}(\nu_i^2)$	(2.15)
h	$\underset{h}{\mathrm{E}}(\mu_h - \mu)^2 = \underset{h}{\mathrm{E}}(\nu_h^2)$	(2.16)
pi	$\underset{p}{\mathrm{E}}\,\underset{i}{\mathrm{E}}(\mu_{pi} - \mu_p - \mu_i + \mu)^2 = \underset{p}{\mathrm{E}}\,\underset{i}{\mathrm{E}}(\nu_{pi}^2)$	(2.17)
ph	$\underset{p}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(\mu_{ph} - \mu_p - \mu_h + \mu)^2 = \underset{p}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(\nu_{ph}^2)$	(2.18)
ih	$\underset{i}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(\mu_{ih} - \mu_i - \mu_h + \mu)^2 = \underset{i}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(\nu_{ih}^2)$	(2.19)
pih	$\underset{p}{\mathrm{E}}\,\underset{i}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(X_{pih} - \mu_{pi} - \mu_{ph} - \mu_{ih} + \mu_p + \mu_i + \mu_h - \mu)^2 = \underset{p}{\mathrm{E}}\,\underset{i}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(\nu_{pih}^2)$	(2.20)

Table 3. Estimators of Variance Components for a $p \times i \times h$ Design

Effect (α)	Estimator of $\sigma^2(\alpha)$	
p	$[MS(p) - MS(pi) - MS(ph) + MS(pih)] / n_i n_h$	(2.21)
i	$[MS(i) - MS(pi) - MS(ih) + MS(pih)] / n_p n_h$	(2.22)
h	$[MS(h) - MS(ph) - MS(ih) + MS(pih)] / n_p n_i$	(2.23)
pi	$[MS(pi) - MS(pih)] / n_h$	(2.24)
ph	$[MS(ph) - MS(pih)] / n_i$	(2.25)
ih	$[MS(ih) - MS(pih)] / n_p$	(2.26)
pih	$MS(pih)$	(2.27)

The mean squares can be estimated by calculating the sums of squares (*SS*) for each score effect, and then dividing by its corresponding degrees of freedom (*df*). The sums of squares can be calculated from *T* terms, which are sums of squared mean scores (not to be confused with sums of squares, which are sums of squared deviation scores). For example for single-facet designs, the *T* term for the observed score mean is

$$T(\mu) = n_p n_i \bar{X}^2 \qquad (2.28)$$

and for two-facet designs, the *T* term for the observed score mean is

$$T(\mu) = n_p n_i n_h \bar{X}^2 \qquad (2.29)$$

The *T* terms, sums of squares, and degrees of freedom for all of the score effects in the $p \times i \times h$ design are presented in Table 4. The *T* terms, sums of squares, and degrees of freedom for all of the score effects for each of the designs of interest are presented in Appendix A, Tables A18 – A22.

Table 4. *T* Terms, Sums of Squares, and Degrees of Freedom for a $p \times i \times h$ Design

Effect α	*T* terms $T(\alpha)$	Sums of squares $SS(\alpha)$	Degrees of freedom $df(\alpha)$	
p	$n_i n_h \sum \bar{X}_p^2$	$T(p) - T(\mu)$	$n_p - 1$	(2.30-32)
i	$n_p n_h \sum \bar{X}_i^2$	$T(i) - T(\mu)$	$n_i - 1$	(2.33-35)
h	$n_p n_i \sum \bar{X}_h^2$	$T(h) - T(\mu)$	$n_h - 1$	(2.36-38)
pi	$n_h \sum\sum \bar{X}_{pi}^2$	$T(pi) - T(p) - T(i) + T(\mu)$	$(n_p - 1)(n_i - 1)$	(2.39-41)
ph	$n_i \sum\sum \bar{X}_{ph}^2$	$T(ph) - T(p) - T(h) + T(\mu)$	$(n_p - 1)(n_h - 1)$	(2.42-44)
ih	$n_p \sum\sum \bar{X}_{ih}^2$	$T(ih) - T(i) - T(h) + T(\mu)$	$(n_i - 1)(n_h - 1)$	(2.45-47)
pih	$\sum\sum\sum X_{pih}^2$	$T(pih) - T(pi) - T(ph) - T(ih) + T(p) + T(i) + T(h) - T(\mu)$	$(n_p - 1)(n_i - 1)(n_h - 1)$	(2.48-50)

Methods for Estimating Standard Errors

Standard errors of estimated variance components can be estimated via three general methods: empirically, theoretically, and by resampling. A brief overview of these three methods follows.

Empirical Estimation

The most conceptually straightforward method of estimating standard errors of estimated variance components is to replicate a study several times and calculate the standard deviation of the estimated variance components across replications, which is referred to as the empirical method (see Brennan, 2001, pp. 180-181 for an example). However, this method requires replication of a study several times, which may not be feasible in a practical setting.

Theoretical Estimation

Another method, the theoretical or traditional method, is to assume multivariate normality and random and independent score effects; then, closed-form equations are used to estimate the theoretical standard errors of the estimated variance components (Searle, 1971). These equations are discussed more fully in Ch. 3 of this dissertation, with respect to calculating the standard errors of normally distributed simulated data, and can be found in Appendix A, Tables A23 – A27. A limitation of this method is that the assumptions are not always realistic, particularly in a multiple choice testing environment where the data are scored dichotomously, or in writing or performance assessments where the data are often scored polytomously (Brennan, 2001).

Normal data consist of scores for which the score effects are distributed as normal. This may be a reasonable assumption for some data, but often is not a realistic assumption for data consisting of test scores. Test scores are more often dichotomous or polytomous. *Dichotomous* data, also known as binary data, occur when the item responses are scored dichotomously (1 for correct or success; 0 for incorrect or failure), as multiple choice questions are often scored. *Polytomous* data occur when item responses are scored as finite ordered integers, such as constructed response items rated on a 6-point scale from 0 to 5.

The nature of the data is important. When the data are normally distributed, we have a pretty good idea of how the data will behave; for example, there are equations that work well for estimating standard errors of estimated variance components. Therefore, the normally distributed data in this dissertation were included as more of a validation check that the programs were working as they should, because we know what to expect for the results. The main focus is on the dichotomous and polytomous data, about which less is known. These non-normal distributions are the ones for which bootstrap or other procedures that do not assume normality are needed to get accurate estimates of the standard errors.

Bootstrap Estimation

Resampling methods can be used to estimate the variability of statistics when theoretical models or distributions of the statistics are not available, either because they do not exist, or because they are extremely complicated (Shao & Tu, 1995). As noted earlier, there are two main resampling methods that have been used to estimate standard errors of estimated variance components: the bootstrap, and the jackknife. These two methods are called resampling methods because they involve taking multiple samples from a dataset to create additional sample datasets, which can then be used to estimate the standard errors of the variance component estimates obtained across the sample datasets. For reasons cited in Chapter 1, the bootstrap was the focus of this dissertation. A brief overview of the bootstrap methodology follows.

The bootstrap is a resampling method that involves drawing a number of independent random samples (with replacement) from a dataset of interest. Each sample from the dataset is called a *bootstrap sample*, not to be confused with the dataset from which the bootstrap sample is taken. A Monte Carlo algorithm is used to draw bootstrap samples. Monte Carlo algorithms are computationally intensive procedures that use random number generators to draw independent samples from a dataset of interest. An estimate of the statistic of interest is calculated for each bootstrap sample. The standard deviation of these estimates across bootstrap samples is taken as an estimate of the standard error of the statistic. In the case of G theory, variance components would be estimated for each of a given number of bootstrap samples, and the standard deviation of each variance component estimate would be calculated across the bootstrap samples as an estimate of the standard error of the variance component. An advantage of the bootstrap is that it does not require assumptions about the distribution of the statistic (i.e., the data are not required to be normally distributed).

Using the bootstrap to estimate standard errors of estimated variance components is more complicated than it may appear at first glance. This is because G theory designs can involve multiple facets, and decisions must be made about which facet or facets to resample when constructing the bootstrap samples (e.g., persons, items, raters, etc.). There are several sampling schemes that could be utilized to construct bootstrap samples for a given G theory design. By convention, the sampling schemes are named for the facet(s) that are sampled for that particular bootstrapping procedure. For example, the $p \times i$ design consists of the objects of measurement (*p*) and a single facet (*i*), thus three sampling schemes that could be utilized are boot-*p*, boot-*i*, and boot-*p,i*. The *boot-p* sampling scheme involves sampling n_p persons with replacement, but not sampling items (i.e., each sample contains the same n_i items). The *boot-i* sampling scheme involves sampling n_i items with replacement, but not sampling persons. The *boot-p,i* sampling scheme involves sampling both persons and items with replacement (Brennan Harris, & Hanson, 1987; Brennan, 2001). There are several other sampling schemes that are further elaborated upon in Chapter 3 of this dissertation.

Another issue regarding the bootstrap is that of bias. One consequence of the bootstrap methodology when applied with designs of the type in G theory is that the resulting estimates of variance components can be biased (see Brennan, 2001, p. 186 for an explanation of how bias arises). Therefore, bias correction factors were proposed by Wiley (2001), and extended by Brennan (2007).

A further consequence of some of the bootstrap sampling schemes is that when a facet is *not* resampled, that facet is in effect treated as fixed. For example, as stated previously, in the boot-*p* sampling scheme, persons are resampled but items are not. Items then are fixed which violates an assumption of the random effects model (Wiley, 2001). Therefore, the resulting estimates may be inaccurate.

Additionally, previous research has shown that some bootstrap procedures yield more accurate estimates of particular variance components and their standard errors, and other bootstrap procedures yield more accurate estimates of other variance components and their standard errors (Brennan, Harris, & Hanson, 1987; Luecht & Smith, 1989; Othman, 1995; Wiley, 2001; Tong & Brennan, 2007). Therefore, when a decision is made to use the bootstrap to estimate standard errors of estimated variance components, further decisions must be made regarding which specific bootstrap procedures to use. A few ad hoc rules for which bootstrap procedures to use for estimating each of the variance components and their standard errors were suggested by Brennan et al. (1987), and further elaborated upon by Wiley (2001). Tong and Brennan (2007) provided a more extensive set of rules based on more extensive analyses, but their rules are still ad hoc. Therefore, further research is needed employing various G theory designs, with various variance component structures and sample size patterns, to see how well the rules lead to accurate estimates of variance components and their corresponding standard errors.

Another concern regarding the bootstrap approach of estimating standard errors of estimated variance components is that complications arise with respect to estimating relative error variances and especially absolute error variances and their standard errors, as explained in the following paragraphs. Relative and absolute error variances are used to calculate reliability estimates; therefore, their accuracy is very important. These estimates are used to make inferences, for example, regarding the optimal numbers of items and raters to include in a testing situation. Thus, if there is not information available about the accuracy of these estimates, poor decisions leading to less than ideal measurement conditions may result.

Relative error (δ) is the difference between a person's observed deviation score (the observed score minus the mean score across all facets but not across the objects of measurement) and their universe deviation score (the person mean minus the grand mean). *Relative error variance* $\sigma^2(\delta)$ is the sum of all of the variance components that involve the objects of measurement except for the objects of measurement variance component (typically persons). *Absolute error* (Δ) is the difference between a person's observed score and their universe score (the *universe score* is the expected value of a person's mean score across every combination of measurement procedures). *Absolute error variance* $\sigma^2(\Delta)$ is the sum of all of the variance components except for the objects of measurement variance component. For example, in a simple $p \times i$ design, relative error is

$$\sigma^2(\delta) = \frac{\sigma^2(pi)}{n_i} \tag{2.50}$$

and absolute error is

$$\sigma^2(\Delta) = \sigma^2(\delta) + \frac{\sigma^2(i)}{n_i} \tag{2.51}$$

The relative and absolute error variances for all of the designs of interest are presented in Appendix A, Tables A28 and A29.

If the bootstrap procedures used to estimate the standard errors of $\hat{\sigma}^2(i)$ and $\hat{\sigma}^2(pi)$ in the $p \times i$ design are different from one another, then it is unclear which bootstrap procedure(s) should be used to estimate $\sigma^2(\Delta)$. Additionally, the use of a single bootstrap procedure to calculate the standard error of $\hat{\sigma}^2(\Delta)$ will likely not produce an accurate estimate. This problem applies to both relative and absolute error variances and their standard errors in more complicated G theory designs, where several variance components may play a role in defining the relative and absolute error variances. The use of different optimal bootstrap procedures to estimate the variance components that make up the relative or absolute error variances could affect the resulting estimates. Tong and Brennan (2007) found a bootstrap procedure that produced reasonable results for estimating the standard error of the relative error variance, but not for estimating the standard error of the absolute error variance. They proposed a work-around for estimating the standard error of the absolute error variance involving the use of multiple bootstrap procedures, which is described in more detail in a later section of Chapter 2 of this dissertation.

Previous Studies of Standard Errors of Estimated Variance Components

Several studies have investigated the use of the bootstrap to estimate standard errors of estimated variance components. Early studies focused on one-facet, crossed designs, resulting in recommendations for the particular bootstrap procedures to use to estimate standard errors of estimated variance components for the $p \times i$ design (Brennan, Harris, & Hanson, 1987; Luecht & Smith, 1989; Othman, 1995). However, it was noted that the resulting estimates tended to be biased (Brennan, Harris, & Hanson, 1987; Othman, 1995). Brennan et al. (1987) suggested correction factors that were subsequently derived mathematically by Wiley (2001) for the $p \times i$ design. Later, Brennan (2007) provided correction factors and their mathematical proofs for any design. Finally, a recent study examined bias-corrected bootstrap estimates of variance components in more complex two-facet designs, resulting in a set of rules suggesting which bootstrap procedures to use to estimate standard errors in any G theory design (Tong & Brennan, 2007). Details regarding the conditions of these studies are in Appendix B, Table B1. A summary of these studies follows.

Single-Facet Designs

Brennan, Harris, and Hanson (1987) looked at traditional, bootstrap, and jackknife procedures for estimating the standard errors of estimated variance components for normal and dichotomous data for $p \times i$ designs. They found that for simulated data with normally distributed score effects, the traditional procedures worked well for estimating variance components and their standard errors, which makes sense because the traditional procedures assume that the score effects are normally distributed. Boot-*p,i* and boot-*p,i,r* bootstrap approaches (where the *r* refers to random sampling of the residuals in addition to sampling persons and items) did not work very well for estimating any of the variance components. Boot-*p* worked well for estimating $\sigma^2(p)$ and $\sigma^2(pi)$ and their standard errors, but not very well for estimating $\sigma^2(i)$ or its standard error. Boot-*i* worked well for estimating $\sigma^2(i)$ and its standard error, but not for estimating $\sigma^2(p)$ or $\sigma^2(pi)$ or their standard errors. The jackknife worked well for estimating all three variance components and their standard errors.

Brennan et al. (1987) also looked at confidence intervals of estimated variance components for normally distributed data. They found that 80% Satterthwaite intervals, which require an assumption of multivariate normality, were nearly the same as the results obtained from simulating 2,000 random samples of size n = 200 persons and k = 20 items and taking the 10th and 90th percentiles. They found that bootstrap intervals based on the 10th and 90th percentiles of the estimated variance components across the bootstrap samples, and jackknife intervals based on the use of Student's *t* distribution, were "reasonably close."

Brennan et al. (1987) also looked at dichotomous data which came from a large-scale licensure testing program. Their "population" consisted of N = 2,000 persons and K = 200 items, from which they took 2,000 samples of size n = 200 persons and k = 20 items to calculate "parameter values" and approximate standard errors to compare to the bootstrap estimates. They found that boot-*p* worked well for estimating $\sigma^2(p)$, boot-*i* worked well for estimating $\sigma^2(i)$, and boot-*i* and boot-*p,i* both worked well for the estimating $\sigma^2(pi)$, which is slightly different from the results for normal data, suggesting that the nature of the data may play a role in which procedures work best for estimating variance components and their standard errors. However, they found that the traditional and bootstrap estimates were too low for $SE(p)$ and $SE(pi)$, and the jackknife estimates for $SE(p)$ and $SE(pi)$ were more variable than the traditional or bootstrap estimates.

Brennan et al. (1987) also looked at confidence intervals of estimated variance components for the dichotomous data using the traditional, bootstrap, and jackknife approaches. Satterthwaite intervals were much too narrow for $\sigma^2(pi)$, and jackknife intervals were somewhat narrow, but considered adequate for practical use. Bootstrap intervals were comparable to the jackknife intervals.

Brennan et al. (1987) came up with some preliminary rules for deciding which bootstrap procedure to use to estimate each of the variance components and their standard errors in a $p \times i$ design, depending on the data type. For normal data, they suggested using boot-*p* to estimate variance components and standard errors for persons, boot-*i* for items, and boot-*p* for the person by item interaction if $n \geq k$ or boot-*i* for the person by item interaction if $n \leq k$. For dichotomous data, they suggested using boot-*p* to estimate variance components and standard errors for persons, boot-*i* for items, and boot-*p,i* for the person by item interaction They also noted that the estimates improved after applying ad hoc bias correction factors. It is important to note that these ad hoc dichotomous-data recommendations were basically based on a single set of data, which makes generalization tenuous at best.

Luecht and Smith (1989) simulated normally distributed data and compared two different bootstrap procedures: resampling across a single facet and resampling both facets in a $p \times i$ design using several sample size patterns. They also generated 476 datasets containing 100 persons and 50 items, and using 200 boot-*p* replications for each simulated dataset, calculated empirical percentile bootstrap confidence intervals of estimated variance components to determine whether the boot-*p* procedure resulted in accurate estimates. They found that resampling both facets resulted in overestimation of the variance components and standard errors for the main effects and underestimation of the variance component and standard error for the interaction effect. Overestimation was particularly evident in the facet for which there were more levels (i.e., greater overestimation of the persons facet when there were disproportionately more persons (n_p = 150) than items (n_i = 20) in the design). Their suggestions were that single-factor resampling (i.e., boot-*p* and boot-*i* rather than boot-*p,i*) is preferable, particularly when the sample size of one facet is disproportionately larger than the sample size of the other facet.

Othman (1995) examined the issue of disproportionality between sample sizes of persons and items such that in most testing applications there are many more persons than there are items. Using a $p \times i$ design and several sample size patterns with simulated normal and dichotomous data, he compared traditional and bootstrap methods for obtaining point estimates, standard errors, and confidence intervals for estimated variance components. Like Brennan et al. (1987), he found that traditional methods worked well for normal data. Additionally, he compared ANOVA and bootstrap procedures for estimating variance components and found that the ANOVA procedures worked adequately for estimating variance components from dichotomous data, but neither traditional nor bootstrap procedures worked well for producing confidence intervals with dichotomous data. Disproportionality of sample size patterns was found to result in inaccurate estimates when the number of items was very small (five), but improved as the number of items increased (to 20). Othman (1995) concluded that for dichotomous data, boot-*p* should be used for estimating $\sigma^2(p)$ and $\sigma^2(pi)$, and boot-*i* should be used for estimating $\sigma^2(i)$, but that the number of items should be large in order to get adequate estimates for items.

Bias Corrections

Wiley (2001) focused on the theory behind bootstrap sampling and investigated the bias associated with bootstrap estimates of variance components and standard errors. Using several $p \times i$ simulation studies of various sample size patterns, he concluded that bias correction does indeed bring bootstrap estimates of variance components closer to their theoretical values for normal data. However, different bootstrap procedures should be used for estimating different variance components and their standard errors. Specifically, Wiley recommended resampling across the facet(s) of interest for that variance component, mirroring the recommendations of Brennan et al. (1987) for dichotomous data. Therefore, one should use boot-*p* for estimating $\sigma^2(p)$ and $SE(p)$, boot-*i* for estimating $\sigma^2(i)$ and $SE(i)$, and boot-*p,i* for estimating $\sigma^2(pi)$ and $SE(pi)$. He also presented ad hoc suggestions for extending these recommendations to multifaceted designs such as the $p \times i \times h$.

Brennan (2007) extended Wiley's (2001) corrections for bias in bootstrap estimates of variance components. Wiley's bias corrections were only applied to the $p \times i$ design, with suggestions for a $p \times i \times h$ design, whereas Brennan (2007) provided procedures for extending the bias corrections to any random balanced design. Brennan's (2007) corrections utilized modified expected *T* terms, which are sums of squared mean scores (Brennan, 2001); thus, the unbiased estimates could be calculated directly during the bootstrapping, whereas Wiley's (2001) corrections were applied to the estimates resulting from the bootstrapping procedures. Brennan's (2007) bias corrections were used in this study, and are presented in Appendix C. Additional details regarding Brennan's (2007) bias corrections can be found in Brennan (2006).

Two-Facet Designs

Tong and Brennan (2007) looked at bootstrap procedures for estimating standard errors of estimated variance components for normal, dichotomous, and polytomous data in the $p \times i \times h$ and $p \times (i : h)$ designs. Seven different bootstrap procedures were utilized for each design: boot-*p*, boot-*i*, boot-*h*, boot-*pi*, boot-*ph*, boot-*ih*, and boot-*pih*. For each design, 1,000 datasets of size 100 × 20 × 2 were generated, and within each dataset 1,000 samples were drawn for each bootstrap procedure. The variance components were estimated within each bootstrap sample, and the mean and variance of each estimated variance component was calculated across the 1,000 bootstrap samples within each dataset, yielding a mean and variance for each variance component estimate. Across the 1,000 datasets, the averages across the means were calculated as estimates of the variance components, and the square root of the averages across the variances were calculated as estimates of the standard errors of the estimated variance components. Relative and absolute error variances and their standard errors were estimated in the same fashion as the variance components. Tong and Brennan (2007) looked at point estimates and standard errors, but did not include confidence intervals. Additional details regarding Tong and Brennan's (2007) study can be found in Tong and Brennan (2006).

For the $p \times i \times h$ design, they found that the raw unadjusted bootstrap estimates of variance components tended to be biased, but were very close to the parameter values after using Brennan's (2007) bias corrections. The raw standard error estimates also appeared to be biased, but results were mixed in terms of how well the bias corrections worked in bringing the values closer to the parameter values. In general, boot-*p* worked the best for *SE*(*p*), *SE*(*pi*), *SE*(*ph*), and *SE*(*pih*), boot-*i* worked best for SE(*i*) and SE(*ih*), and boot-*h* worked best for SE(*h*). None of the bootstrap procedures worked well in estimating $SE(\Delta)$ when $n_h = 2$; however, Tong and Brennan (2007) mentioned that boot-*p* performed well in estimating $\sigma^2(\Delta)$ when $n_h = 4$, 6, or 8, with estimates of $SE(\Delta)$ near zero.

For the $p \times (i : h)$ design, the bias corrections did not work as well for estimating variance components with boot-*h* and boot-*ph*, but worked better for boot-*p*, boot-*i*, and boot-*p,i,h*. In terms of the standard errors, results were similar to those for the $p \times i \times h$ design. Boot-*p* worked best for estimating $SE(p)$, $SE(ph)$, $SE(pi{:}h)$ and $SE(\delta)$. Boot-*h* worked best for $SE(h)$ and boot-*i,h* worked best for $SE(i{:}h)$, but no procedure worked well for estimating $SE(\Delta)$ with $n_h = 2$.

Tong and Brennan's (2007) recommendations echoed those of Brennan et al. (1987) for normal data; that is, sampling along the facet of interest for each main effect, and sampling along the facet that has the largest sample size for each interaction. Tong and Brennan (2007) also expanded the recommendations to include rules for nested designs and relative and absolute error variances. Their rules in full can be found in Appendix D, but a summary of their rules follows. For nested main effects they recommended resampling along the primary index; for example, using boot-*i* to estimate $SE(i{:}h)$. For nested interactions they recommended resampling along the primary index with the larger sample size; for example, using boot-*p* to estimate $SE(pi{:}h)$ if $n_p \geq n_i$ or using boot-*i* if $n_p < n_i$. For estimating the standard error of the estimated relative error variance $S\hat{E}(\delta)$, they recommended resampling along the objects of measurement dimension, which is usually persons, thus boot-*p* would be used. For estimating the standard error of the estimated absolute error variance $S\hat{E}(\Delta)$, they recommended using a combination of bootstrap procedures such that the optimal standard error estimate of each of the relevant variance components is substituted into the appropriate equation. For the $p \times i \times h$ design, the standard error of the estimated absolute error variance would be estimated as follows, assuming that $n_i \geq n_h$:

$$S\hat{E}(\Delta)=\sqrt{S\hat{E}^2(\delta \mid p)+\frac{S\hat{E}^2(i \mid i)}{n_i^2}+\frac{S\hat{E}^2(h \mid h)}{n_h^2}+\frac{S\hat{E}^2(ih \mid i)}{n_i^2 n_h^2}} \tag{2.52}$$

where $S\hat{E}^2(\alpha \mid \lambda)$ refers to the squared estimated standard error of α given boot-λ. For the $p \times (i{:}h)$ design, the standard error of the absolute error variance would be estimated as follows:

$$S\hat{E}(\Delta)=\sqrt{S\hat{E}^2(\delta \mid p)+\frac{S\hat{E}^2(h \mid h)}{n_h^2}+\frac{S\hat{E}^2(i:h \mid i)}{n_i^2 n_h^2}} \tag{2.53}$$

Summary of Previous Research

Although several studies have investigated bootstrap estimated standard errors of estimated variance components, and rules have been proposed and tested to some extent, there is still not a comprehensive body of evidence supporting these rules. Most of the studies involved single-facet $p \times i$ designs; only one of the studies investigated more complicated two-facet and nested designs (Tong & Brennan, 2007).

There has also been some inconsistency in the rules suggested by various studies. For example, Tong and Brennan (2007) suggested that *SE*(*pi*) should be estimated with boot-*p* if $n_p \geq n_i$ or boot-*i* if $n_p \leq n_i$, mirroring the suggestions of Brennan et al. (1987) for normal data; whereas Wiley (2001) suggested that *SE*(*pi*) should be estimated with boot-*p*,*i*, mirroring the suggestions of Brennan et al. (1987) for dichotomous data. However, Wiley's (2001) results do provide some support for Tong and Brennan's (2007) rules, such that the estimates of $\sigma^2(pi)$ and *SE*(*pi*) were at least as accurate under boot-*p* resampling as they were under boot-*p*,*i* resampling.

Additionally, more than one study made use of the same parameter values for the variance component structures and the same sample size patterns. Various variance component structures and sample size patterns should be investigated to ensure that these rules apply beyond the situations studied.

Therefore, this study investigated bootstrap procedures for estimating standard errors of estimated variance components and relative and absolute error variances in several designs. The same simulated datasets were used across the various bootstrap procedures so that they could be directly compared. This study contributes to the literature by extending Tong and Brennan's (2007) study of bootstrap procedures in two-facet and nested designs to include various G theory designs, variance component structures, sample size patterns, and data types, as well as by providing confidence interval coverage of the variance components and relative and absolute error variances.

CHAPTER 3: METHODOLOGY

The purpose of this study was to assess how well the rules suggested by Tong and Brennan (2007) for estimating standard errors of estimated variance components using bootstrap procedures, apply in situations beyond those considered in previous studies (Brennan et al., 1987; Luecht & Smith, 1989; Othman, 1995; Wiley, 2001; Tong & Brennan, 2007). Therefore, this study examined various bootstrap procedures with respect to four factors. The factors (and number of levels of each) that were studied are: G theory design (5), variance component structure within G theory design (2), sample size pattern within G theory design (2), and data type (3). Within each G theory design the variance component structures and sample size patterns were fully crossed, yielding 20 unique combinations of G theory design, variance component structure, and sample size pattern. Crossing these 20 combinations with three data types yielded 60 unique combinations of conditions altogether. Simulated data was generated for each of the 60 combinations of conditions, and several bootstrap procedures were applied to each to obtain unbiased point estimates, standard errors, and confidence interval coverage of variance components, relative error variance, and absolute error variance.

Simulated data was used rather than actual test data, for several reasons. Simulation allows for greater control over the studied conditions. With simulated data, the parameter values are specified a priori, thus establishing the “truth;” whereas with empirical data, the true parameter values are never actually known. Knowing the true parameter values is important in terms of being able to evaluate the estimates of the variance components and standard errors with respect to how close they are to the parameter values. Additionally, having greater control over the data allows for the construction of complete datasets with no missing values, thus removing the complexity of having to deal with missing data. Simulation also allows for the manipulation of conditions such that a large number of conditions can be studied without the cost in time and money required to collect real empirical data. By making slight changes to the conditions of the study, insight can be gained regarding the extent to which the rules hold up under varying conditions, and if they do not hold up, perhaps provide support for adjustments to the rules or new rules. Additionally, simulation is useful when an analytical solution does not exist. As discussed previously, test data are often scored dichotomously or polytomously, rendering commonly used standard error equations inappropriate due to violations of the normality assumption. For all of these reasons, simulation was determined to be a preferable approach for this study.

Factors, Levels, and Conditions of the Study

G Theory Designs

Five G theory designs were employed in this study: $p \times i$, $p \times i \times h$, $i : p$, $i : h : p$, and $p \times (i : h)$. Several studies have investigated bootstrap standard errors in the $p \times i$ design (Brennan, Harris, & Hanson, 1987; Luecht & Smith, 1989; Othman, 1995, Wiley, 2001), and Tong and Brennan (2007) examined two two-facet designs, the $p \times i \times h$ and the $p \times (i : h)$. This study also included the $i : p$ and $i : h : p$ designs to determine whether the proposed rules hold for these single-facet and nested designs as well. Including more designs allows for the rules to be tested to a greater extent, thus determining the extent of their applicability more broadly.

Variance Component Structures and Sample Size Patterns

Two variance component structures were studied for each of the five G theory designs. These variance component structures varied from one G theory design to another. Two sample size patterns were studied for each of the five G theory designs, and these patterns varied across designs. The variance component structures and sample size patterns associated with each G theory design are shown in Table 5. For each G theory design, the variance component structures were fully crossed with the sample size patterns. Therefore, there were 20 unique combinations of G theory design, variance component structure, and sample size pattern.

Table 5. Variance Component Structures and Sample Size Patterns Considered in this Study

		$\sigma^2(\alpha)$ Structures		Sample size patterns	
Design	α	A	B	1	2
$p \times i$	*p*	0.35	0.30	200	100
	i	0.05	0.20	10	50
	pi	0.60	0.50		
$p \times i \times h$	*p*	0.19	0.20	200	100
	i	0.01	0.15	5	10
	h	0.04	0.05	2	3
	pi	0.01	0.10		
	ph	0.24	0.10		
	ih	0.01	0.05		
	pih	0.50	0.35		
$i : p$	*p*	0.10	0.30	50	50
	i:*p*	0.90	0.70	5	50
$i : h : p$	*p*	0.28	0.15	100	50
	h:*p*	0.02	0.15	5	2
	i:*h*:*p*	0.70	0.70	20	5
$p \times (i : h)$	*p*	0.20	0.10	200	100
	h	0.01	0.05	5	10
	ph	0.05	0.15		
	i:*h*	0.04	0.15	15	3
	pi:*h*	0.70	0.55		

The variance component structures used in Tong and Brennan (2007) for normal data were similar to those used in other studies (Brennan et al., 1987) except extended to include a second facet. For dichotomous data, the same variance component parameter values that were used to create the normal data were used as a starting point, but the resulting score effects were dichotomized based on the resulting simulated scores. For polytomous data, the variance component parameter values used by Tong and Brennan (2007) were the same as those used in Feng (2002), which were based on the results of empirical studies which used test data (Gao & Brennan, 2001; Lane, Liu Ankenmann, & Stone, 1996). Brennan et al. (1987) described the rationale behind how they chose variance component structures as follows. Large magnitudes of variance components were chosen to highlight small differences between statistics. Large person by task interactions, and large differences between person variance and item variance were chosen because these patterns reflect what is often seen in practice.

The rationale for choosing the parameter values for the particular variance component structures and sample size patterns for this study was based on three main principles. First, one of the goals of this study was to expand upon current research on the topic of bootstrap standard errors of estimated variance components, so values were chosen to be different from those already studied in this context (Brennan, Harris, & Hanson, 1987; Luecht & Smith, 1989; Othman, 1995; Wiley, 2001; Tong & Brennan, 2007).

Second, like Brennan et al. (1987), values were chosen that were considered to be reasonable based on previous G studies in various contexts. It is important that the values chosen are realistic, so that the results are applicable to practical use. One factor that was taken into consideration was the differences in variance component structures that are reasonable given different facets; for example, for a persons crossed with items crossed with raters study versus a persons crossed with items crossed with occasions study. Additionally, based on the literature, it is known that the highest order interaction term tends to have the largest variance, and the persons by items interaction also tends to be large, followed by the variance component for the objects of measurement, which is typically persons. Sample size patterns were similarly chosen to be reasonable given the context. For example, it is expected that the number of raters or occasions will typically be small relative to the numbers of persons and items, but the number of items could also be small if the context is a performance assessment.

Lastly, within the bounds of these reasonable variance component structures and sample size patterns based on previous research, values were chosen to reflect a wide range of possible structures, from more homogeneous structures in which the variance components of a given G theory design contribute similar amounts to the total variance, to more extreme structures in which the majority of the variance is contained within a smaller subset of variance components. To study the robustness of Tong and Brennan's (2007) bootstrap rules, it is important to test them under conditions that stretch the boundaries of what is possible while still being realistic.

Several empirical G studies were reviewed with respect to their G theory designs, sample sizes, and resulting variance components estimates in order to determine what types of variance component structures were reasonable and realistic. The values chosen for this study were guided by using typical values in the empirical literature as a starting point, with adjustments made where it seemed necessary. A summary of these studies follows, and the proportions of variance explained by each estimated variance component for these and several other studies reviewed can be found in Appendix E.

For each G theory design, several published studies were examined to the extent that they could be found employing that design. For the $p \times i$ design, the two variance component structures that guided the selection of the parameter values were based on a persons crossed with items study in a math performance assessment context (Lane, Liu, Ankenmann, & Stone, 1996). For the $p \times i \times h$ design, the two variance component structures were based on a persons by occasions by tasks study in a science concept maps assessment context (Yin & Shavelson, 2008) and a persons by raters by occasions study in a behavioral rating scale context (Chafouleas, Christ, & Riley-Tillman, 2009). For the

$i : p$ design, the two variance components structures were based on an items within persons study in a medical certification context (Burch, Norman, Schmidt, & van der Vleuten, 2008). For the $i : h : p$ design, the two variance components structures were based on a persons within groups within forms study in a standard setting context (Lee & Lewis, 2008). Lastly, for the $p \times (i : h)$ design, the two variance component structures were based on a persons crossed with items within passages study in a standardized testing context (Lee & Frisbie, 2009) and a persons crossed with items within content areas in a psychological inventory context (Christophersen, Helseth, & Lund, 2008).

In terms of sample size patterns, Tong and Brennan (2007) used n_p = 100 persons, n_i = 20 items, and n_h = 2 or 4 raters. The ratio of persons to items was 5, the ratio of persons to raters was 5 and 10, and the ratio of items to raters was 25 and 50. Other studies examining single-facet designs used person to item ratios ranging from 0.13 (n_p = 20, n_i = 150, Leucht & Smith, 1989) to 120 (n_p = 600, n_i = 5, Wiley, 2001). Smaller sample sizes are of interest because if the procedures yield accurate results at smaller sample sizes, they likely will yield results at least that good or better for larger sample sizes (Brennan et al., 1987). Brennan et al. (1987) and Tong and Brennan (2007) chose large numbers of persons and smaller numbers of items because this is often what is seen in practice, as is a small number of raters relative to persons and items.

The variance component structures and sample size patterns that were considered in this study are presented in Table 5. Within each of the five G theory designs, two variance component structures and two sample size patterns were crossed, yielding four conditions for each of the five designs. These 20 combinations of conditions were crossed with three data types, yielding a total of 60 conditions. For the normally distributed data, the total variance was set equal to one. Normally distributed score effects were dichotomized to create the dichotomous data, thus the parameter values for the variance component structures and standard errors were not known a priori. However, by the properties of the Bernoulli distribution, it is known that the total observed score variance across variance components for the dichotomous data will sum to no larger than 0.25. The total variance of the polytomous data was set approximately equal to one. The decision to set the total variance for the normal and polytomous data equal to one was based on wanting to make interpretation of the relative magnitudes of variance from different sources more straightforward. Additionally, the values seem reasonable, given the empirical literature. The decision to dichotomize the normal data to create the dichotomous data was based on convenience and familiarity, as this was the method used by Tong and Brennan (2007).

Data Types

Three data types were studied: normal, dichotomous, and polytomous. Normal data was simulated because they provide a good baseline with which to compare the non-normal data. This is because some of the properties of data that are normally distributed are known, unlike the properties of dichotomous and polytomous data. Equations for calculating standard errors of estimated variance components have been derived and shown to be useful for normal data; therefore, the bootstrap estimates obtained from normal data could be compared to theoretical estimates as a validation check that the simulation and bootstrap procedures were working correctly. Additionally, if the procedures do not work well for normal data, it is doubtful that they will work for dichotomous or polytomous data (Brennan, Harris, & Hanson, 1987). Dichotomous and polytomous data were simulated because of their practical value. These are the types of data more commonly found in a testing context, and therefore many of the applications of G theory involve these types of data. Crossing these three data types with the 20 unique combinations of G theory design, variance component structure, and sample size pattern yielded 60 unique combinations of conditions altogether.

In summary, five G theory designs were utilized in this study. Within each G theory design, two variance component structures were crossed with two sample size patterns, to yield 20 combinations of conditions. Each of these 20 combinations was crossed with three data types, to yield a total of 60 conditions. Tong and Brennan (2007) examined point estimates and standard errors of variance components, relative error variance, and absolute error variance. This study investigated these estimates, as well as confidence interval coverage. To obtain the estimated variance components and their standard errors, Tong and Brennan (2007) used several bootstrap procedures to see if their rules produced accurate estimates. This study did the same.

Bootstrap Procedures

The specific bootstrap procedures applied to each condition were dependent on the G theory design of that condition. For each G theory design, the bootstrap procedures that were used correspond to the facets of that G theory design, and can be expressed as boot-λ, where the λ in boot-λ refers to the facet(s) that were resampled in that bootstrap procedure. The facets that make up λ can be seen in Table 5. For example, for the $p \times i$ and $i : p$ designs, there are three bootstrap procedures that were employed: boot-*p*, boot-*i*, and boot-*p*,*i*. For the $p \times i \times h$, $i : h : p$, and $p \times (i : h)$ designs, seven bootstrap procedures were employed: boot-*p*, boot-*i*, boot-*h*, boot-*p*,*i*, boot-*p*,*h*, boot-*i*,*h*, and boot-*p*,*i*,*h*.

When two or more facets were resampled for a given bootstrap procedure, those facets were sampled independently. For example, boot-*p*,*i* involved resampling *p* and independently resampling *i*. For nested designs, the order of bootstrapping is important. Specifically, the nested facet should be resampled first for each level of the non-nested facet, then the non-nested facet should be resampled. For example, for boot-*i*,*h* in the $p \times (i : h)$ design, first *i* should be resampled for each level of *h*, then *h* should be independently resampled. This results in the same set of items being resampled when *h* is the same, which should be the case when *i* is nested within *h*.

Bootstrap estimates were corrected for bias using the equations found in Brennan (2007), and the equations for these bias corrections can be found in Appendix C. Programs written in ANSI C were used to simulate data and conduct all of the bootstrap resampling and analyses.

Simulated Data Generation

This section describes how data was generated for the various conditions of this study. For each combination of conditions (i.e., G theory design, variance component structure, sample size pattern, and data type), 1,000 simulated datasets were generated. Each dataset constitutes a replication. For the $p \times i$ and $i : p$ designs, each dataset consisted of $n_p n_i$ observations. For the $p \times i \times h$, $i : h : p$, and $p \times (i : h)$ designs, each dataset consisted of $n_p n_i n_h$ observations.

Normal Data

The formulas that were used to simulate a single dataset of normal data for each G theory design are presented in Table 6, where $\sigma(\alpha)$ corresponds to the square root of a pre-specified variance component parameter value and z_α denotes a random and independent draw from a unit normal distribution (Brennan et al, 1987). The specific variance component parameter values that were employed for each G theory design are given in Table 5. The μ parameter values were also set a priori, but they are not of any consequence because the focus is on the variance components. For the normal data, μ was set equal to 0 for convenience. Table 6 also contains the matrix dimensions of the generated datasets. The specific values (i.e., sample size patterns) of the matrix dimensions that were employed for each G theory design are given in Table 5.

For normally distributed score effects, the general form of the equation used to calculate the parameter values for the standard errors of the estimated variance components is

$$SE(\alpha) = \sigma\left[\hat{\sigma}^2(\alpha)\right] = \frac{1}{\pi(\dot{\alpha})}\sqrt{\sum_{\beta}\frac{2\left[\mathrm{E}MS(\beta)\right]^2}{df(\beta)}} \tag{3.6}$$

where $\dot{\alpha}$ refers to the indices *not* in α, $\pi(\dot{\alpha})$ is the product of the sample sizes for all indices *not* in α, and β refers to the indices that make up the expected mean squares associated with $\hat{\sigma}^2(\alpha)$ (Brennan, 2001; Tong & Brennan, 2007). For example, consider the $p \times i$ design in Table 5 with normal data, Variance Component Structure A ($\sigma^2(p)$ = 0.35 and $\sigma^2(pi)$ = 0.60), and Sample Size Pattern 1 (n_p = 200 and n_i = 10). In this case,

$$\mathrm{E}MS(p) = \sigma^2(pi) + n_i\sigma^2(p) = 0.60 + 10(0.35) = 4.10$$

$$\mathrm{E}MS(pi) = \sigma^2(pi) = 0.60$$

and the standard error of the estimated variance component for p is

$$SE(p) = \sigma\left[\hat{\sigma}(p)\right] = \frac{1}{10}\sqrt{\frac{2(4.10)^2}{199} + \frac{2(0.60)^2}{(199)(9)}} = 0.04115$$

Table 6. Formulas for Generating Normal Data

Design	Formula for generating normal data	Matrix of observed scores	
$p \times i$	$X_{pi} = \mu + \sigma(p)z_p + \sigma(i)z_i + \sigma(pi)z_{pi}$	$n_p \times n_i$	(3.1)
$p \times i \times h$	$X_{pih} = \mu + \sigma(p)z_p + \sigma(i)z_i + \sigma(h)z_h + \sigma(pi)z_{pi} + \sigma(ph)z_{ph} + \sigma(ih)z_{ih} + \sigma(pih)z_{pih}$	$n_p \times n_i \times n_h$	(3.2)
$i : p$	$X_{pi} = \mu + \sigma(p)z_p + \sigma(i{:}p)z_{i:p}$	$n_p \times n_i$	(3.3)
$i : h : p$	$X_{pih} = \mu + \sigma(p)z_p + \sigma(h{:}p)z_{h:p} + \sigma(i{:}h{:}p)z_{i:h:p}$	$n_p \times n_i \times n_h$	(3.4)
$p \times (i : h)$	$X_{pih} = \mu + \sigma(p)z_p + \sigma(h)z_h + \sigma(ph)z_{ph} + \sigma(i{:}h)z_{i:h} + \sigma(pi{:}h)z_{pi:h}$	$n_p \times n_i \times n_h$	(3.5)

The estimation version of Equation 3.6 is

$$S\hat{E}(\alpha) = \hat{\sigma}\left[\hat{\sigma}^2(\alpha)\right] = \frac{1}{\pi(\dot{\alpha})}\sqrt{\sum_{\beta}\frac{2[MS(\beta)]^2}{df(\beta)+2}} \tag{3.7}$$

where $\left[S\hat{E}(\alpha)\right]^2$ is an unbiased estimate of $\sigma^2[\hat{\sigma}^2(\alpha)]$ (see Searle, 1971; Searle et al., 1992). The specific estimation equations associated with each G theory design are given in Appendix A, Tables A23 – A27.

Dichotomous Data

The formulas that were used to generate simulated normal data for the five G theory designs were also used to generate simulated dichotomous data by assigning a value of one if the simulated observed score was greater than one; otherwise assigning a value of zero (the cut point of one is arbitrary). With this approach, parameter values for the variance components and standard errors were not known a priori. Therefore, they were obtained based on 5,000 independent replications of simulated data of size $n_p \times n_i$ or $n_p \times n_i \times n_h$, depending on the G theory design. Variance components were estimated for each replication, and the means and standard deviations of the estimated variance components across the 5,000 replications were taken as the variance component and standard error parameter values, respectively (Tong & Brennan, 2007). Five thousand replications were used because it was the number of replications used by Tong and Brennan (2007) to produce the unknown parameter values for the dichotomous and polytomous data in their study. This number seemed to be a reasonable balance between the increased computation time that is required as the number of replications increase, and the diminishing returns in terms of increased precision of estimates that result from increasing the number of replications.

Polytomous Data

To simulate polytomous data for each design, the equations in Table 7 were used, where *BIN* denotes a random and independent binomial value with *n* trials and probability of success π (Feng, 2002; Tong & Brennan (2007). Trial and error was used to obtain the binomial values for the number of trials and probability of success yielding mean values within 0.0001 of the parameter value. For example, for the $p \times i$ design, random draws from a binomial distribution with $n = 4$ trials and probability of success $\pi = 0.0969$ yields parameter values centered at approximately 0.35, which is the parameter value for $\sigma^2(p)$ in variance component structure A in Table 5. The binomial values that were used to simulate the polytomous data are presented in Table 8.

Table 7. Formulas for Generating Polytomous Data

Design	Formula for generating polytomous data	
$p \times i$	$X_{pi} = BIN_p + BIN_i + BIN_{pi}$	(3.8)
$p \times i \times h$	$X_{pih} = BIN_p + BIN_i + BIN_h + BIN_{pi} + BIN_{ph} + BIN_{ih} + BIN_{pih}$	(3.9)
$i : p$	$X_{pi} = BIN_p + BIN_{i:p}$	(3.10)
$i : h : p$	$X_{pih} = BIN_p + BIN_{h:p} + BIN_{i:h:p}$	(3.11)
$p \times (i : h)$	$X_{pih} = BIN_p + BIN_h + BIN_{ph} + BIN_{i:h} + BIN_{pi:h}$	(3.12)

Because the score effects of polytomous data are not assumed to be normally distributed, the parameter values of the standard errors for the polytomous data could not be calculated using normality-based equations derived from Equation 3.6, and thus were not known a priori. The parameter values of the standard errors of the variance components were obtained in the same fashion as those for the dichotomous data, based on the standard deviations of estimated variance components across 5,000 replications.

Table 8. Binomial Values for Generating Polytomous Data

Design	α	Structure A		Structure B	
		n	π	n	π
$p \times i$	*p*	4	0.0969	3	0.1127
	i	5	0.0101	2	0.1127
	pi	3	0.2764	2	0.5000
$p \times i \times h$	*p*	4	0.0500	2	0.1127
	i	1	0.0101	1	0.1838
	h	4	0.0101	5	0.0101
	pi	1	0.0101	1	0.1127
	ph	1	0.4000	1	0.1127
	ih	1	0.0101	5	0.0101
	pih	2	0.5000	4	0.0969
$i : p$	*p*	1	0.1127	3	0.1127
	i:*p*	4	0.3419	3	0.3709
$i : h : p$	*p*	3	0.1042	1	0.1838
	h:*p*	2	0.0101	1	0.1838
	i:*h*:*p*	3	0.3709	3	0.3709
$p \times (i : h)$	*p*	2	0.1127	1	0.1127
	h	1	0.0101	5	0.0101
	ph	5	0.0101	1	0.1838
	i:*h*	4	0.0101	1	0.1838
	pi:*h*	3	0.3709	3	0.2418

Relative and Absolute Error Variance Parameter Values

The parameter values for the relative and absolute error variances and their standard errors were obtained as follows. For each of the 60 combinations of conditions (i.e., G theory design, variance component structure, sample size pattern, and data type), parameter values for the relative and absolute error variances – $\sigma^2(\delta)$ and $\sigma^2(\Delta)$, respectively – were obtained by applying the parameter values of the variance components and sample sizes associated with a given combination of conditions to the equations relevant to that G theory design. The equations for calculating the parameter values of the relative and absolute error variances are presented in Appendix A, Tables A28 and A29.

The parameter values for the standard errors of the relative and absolute error variances – $SE(\delta)$ and $SE(\Delta)$, respectively – were obtained as follows. For the normally distributed data, the parameter values were obtained by applying the expected mean squares associated with the parameter values of the variance components and sample sizes for a given combination of conditions to the equations relevant to that G theory design. The equations for calculating the parameter values for the standard errors of the relative and absolute error variances for normal data are presented in Appendix A, Tables A30 and A31. These equations were derived using Equation 3.6. For dichotomous and polytomous data, the parameter values were obtained in the same fashion as the standard errors of the variance components for the dichotomous and polytomous data, based on the standard deviations of the estimated relative and absolute error variances across 5,000 replications

Analyses

Estimating all of the statistics of interest in this study (i.e., point estimates, standard errors, and confidence interval coverage of the variance components, relative error variance, and absolute error variance for each combination of conditions) involved a complicated series of steps, including estimation within bootstrap samples, aggregation of estimates across bootstrap samples within each dataset, and aggregation of estimates across datasets. To present these procedures as clearly as possible, this section is broken down into four sub-sections. The first three sub-sections deal with estimation of variance components and their standard errors, estimation of relative and absolute error variances and their standard errors, and calculation of confidence interval coverage. The final sub-section provides an example to help clarify and integrate the steps of the analyses.

Estimation of Variance Components and Standard Errors

As described in Chapter 2, the bootstrap involves taking repeated independent random samples with replacement from a dataset and calculating the statistics of interest within each bootstrap sample. The standard deviations of the statistics across bootstrap samples are taken as estimates of the standard errors of the statistics (in this case, estimated variance components).

In this study, there were 1,000 datasets for each combination of conditions (i.e., G theory design, variance component structure, sample size pattern, and data type) instead of a single dataset. Each G theory design has associated with it a set of bootstrap procedures as previously specified. The number of bootstrap procedures varies by G theory design. For each dataset within a particular combination of conditions, 999 bootstrap samples were drawn for each applicable bootstrap procedure. The sample size of each bootstrap sample was the same as the sample size of the dataset being sampled. For each bootstrap sample, the relevant variance components were estimated using the equations in Appendix A, Tables A13 – A17.

Nine hundred ninety-nine bootstrap samples were used to obtain variance component, relative and absolute error variance, and standard error estimates primarily because larger numbers are needed to get adequate confidence intervals. Additionally, larger numbers of bootstrap samples should yield better estimates because the bootstrap is based on asymptotic theory. This number also allows for simple calculation of percentile confidence interval bounds. The equations for calculating the lower and upper bounds of a percentile confidence interval are

$$\frac{\alpha}{2}(R+1) \text{ and } \left(1-\frac{\alpha}{2}\right)(R+1), \quad (3.13\text{-}14)$$

respectively, where α is the type two error rate and R is the number of bootstrap samples. When R = 999, the lower and upper bounds of a 90% percentile confidence interval are the 50th and 950th ordered observations, respectively.

The bootstrap estimates of variance components within each bootstrap sample were corrected for bias using the equations and procedures found in Brennan (2006; 2007). The bias corrections applicable to the estimated variance components within each combination of conditions are dependent on the G theory design and the bootstrap procedure under consideration. The bias corrections that were used are presented in Appendix C. The results of this study focus on the bias-corrected estimates.

The bootstrap estimates of variance components were aggregated across bootstrap samples as follows. Across the 999 bootstrap samples within a given dataset, the mean and variance of each of the estimated variance components was calculated.

The means and variances of the estimated variance components were aggregated across datasets as follows. Across the 1,000 datasets, the means of the mean estimates were taken as the final estimates of the variance components, and the square roots of the means of the variance of the estimates were taken as the final estimates of the standard errors of the estimated variance components. Therefore, aggregated across 999 bootstrap samples and 1,000 datasets, ultimately each estimated variance component and its standard error were based on 999,000 estimates.

Estimation of Relative and Absolute Error Variances and Standard Errors

Relative and absolute error variances and their standard errors were also of interest in this study. Relative error variance and its standard error were estimated in much the same fashion as the variance components. The relative error variance was estimated within each bootstrap sample by substituting the relevant variance component estimates $\hat{\sigma}^2(\alpha)$ for the parameters $\sigma^2(\alpha)$ in the equation appropriate to that G theory design. The equations for calculating the relative error variance parameters for each G theory design are presented in Appendix A, Table A28. The mean and variance of the 999 bootstrap estimates of the relative error variance were calculated within each dataset. The mean of the mean relative error variances across the 1,000 datasets was taken as the final estimate of the relative error variance, and the square root of the mean of the variance of the relative error variances across the 1,000 datasets was taken as the final estimate of the standard error of the relative error variance.

Absolute error variance was estimated in the same fashion as the relative error variance, but the standard error of the absolute error variance was estimated using two different methods because Tong and Brennan's (2007) workaround for the standard error of the absolute error variance involves a combination of bootstrap procedures. In addition to estimating the standard error of the absolute error variance in the same fashion as the other standard error estimates, it was also estimated within each dataset by substituting the optimal standard error estimate of each of the relevant estimated variance components into the equation appropriate to that G theory design. This workaround involves estimating the standard error of the absolute error variance within each dataset but not within each bootstrap sample. The equations for calculating the absolute error variance for each G theory design are presented in Appendix A, Table A29; estimates were obtained by substituting the relevant variance component estimates $\hat{\sigma}^2(\alpha)$ for the parameters $\sigma^2(\alpha)$ in the equation appropriate to that G theory design. The equations for estimating the standard error of the absolute error variance for each G theory design using the workaround proposed by Tong and Brennan (2007) are presented in Table A32.

Confidence Interval Coverage

For each bootstrap procedure within each combination of conditions, three different types of ninety percent confidence intervals were calculated within each dataset for the variance components, relative error variance, and absolute error variance: standard normal confidence intervals, percentile confidence intervals, and bias-corrected percentile confidence intervals. Because the parameter values were known, it could be determined whether a given parameter value lied within the bounds of the confidence interval, and the number of times the parameter value was within the confidence interval could be counted across the 1,000 datasets. Confidence interval coverage, or the number of times that the interval contained the parameter, was calculated for each bootstrap procedure within each combination of conditions, and the coverage of the three types of confidence intervals was compared. Using a ninety percent confidence interval, it was expected that approximately 900 of the 1,000 intervals would contain the parameter value.

A commonly employed confidence interval is the standard normal confidence interval. To calculate a 90% standard normal confidence interval, the following equation was used.

$$\bar{X} \pm z(SE), \tag{3.15}$$

where $\bar{X}$ is the mean estimate of the variance component, relative error variance, or absolute error variance across the 999 bootstrap replications; z = 1.645 for a 90% confidence interval; and *SE* is the square root of the variance of the estimated variance component, relative error variance, or absolute error variance across the 999 bootstrap replications. Standard normal confidence intervals are based on the assumption of normality, whereas variance components are asymmetric; therefore, the standard normal confidence intervals may not be accurate. However, they are commonly used and are simple to calculate; therefore, they were of interest in this study.

Percentile confidence intervals are commonly used to calculate confidence intervals for bootstrap procedures, and do not require a normality assumption. Percentile confidence intervals were calculated as follows. Each dataset contained 999 bootstrap estimates of the variance components and relative and absolute error variances for each bootstrap procedure; therefore, the 5th and 95th quantiles were taken as the bounds of the confidence intervals within each dataset for a given bootstrap procedure. Specifically, the 999 bootstrap estimates of each variance component, relative error variance, and absolute error variance were sorted in ascending order, and the 50th and 950th observations were taken as the bounds of the confidence interval for that estimate. Percentile confidence intervals were expected to be more accurate than standard normal confidence intervals for non-normal data; however, if the distribution of the statistic of interest is not symmetric, percentile confidence intervals may be inaccurate (Efron & Tibshirani, 1993).

Because the distributions of estimated variance components tend to be asymmetric, bias-corrected percentile confidence intervals were also calculated for each bootstrap procedure within each of the combinations of conditions. Bias corrected percentile confidence intervals take into account the asymmetry of the distribution of the statistic of interest. It was expected that these confidence intervals would be more accurate than the percentile confidence intervals.

To obtain bias-corrected confidence intervals, first the statistic of interest $\hat{\theta}$ is calculated from the original dataset. In this study, $\hat{\theta}$ is an estimate of a variance component, relative error variance, or absolute error variance calculated from each of the 1,000 datasets. Next, $\hat{\theta}*$ is calculated for each of the 999 bootstrap samples for a given bootstrap procedure, where $\hat{\theta}*$ is a bootstrap estimate of a variance component, relative error variance, or absolute error variance. Then the number of bootstrap estimates of $\hat{\theta}*$ are counted that are less than or equal to $\hat{\theta}$, and this number is designated a. Next, b is calculated using the following equation

$$b = \Phi^{-1}(a/R), \tag{3.16}$$

where Φ^{-1} is the inverse cumulative distribution function of the normal distribution, *a* is the number of bootstrap estimates less than or equal to the estimate from the dataset as previously defined, and *R* is the number of bootstrap samples (999).

Then, the percentiles of the lower and upper lower bounds are calculated from the following equations,

$$Q_L = (R+1)\Phi(2b + z_{.05}) \text{ and } Q_U = (R+1)\Phi(2b - z_{.05}) \qquad (3.17\text{-}18)$$

where $z_{.05}$ = -1.645. Q_L is the percentile of the bootstrap distribution required for the lower endpoint of the bias corrected confidence interval and Q_U is the percentile of the bootstrap distribution required for the upper endpoint of the bias corrected confidence interval. The bias corrected confidence interval, then is ($\hat{\theta}*_{Q_L}, \hat{\theta}*_{Q_U}$). A more detailed discussion of bias corrected percentile confidence intervals can be found in Carpenter and Bithell (2000).

Illustrative Example

To illustrate the process of simulating data and calculating bootstrap standard errors, the $p \times i$ design was used because of its simplicity, but the procedures are virtually the same for the other designs. For the $p \times i$ design, this procedure was the same across the 12 combinations of two variance component structures, two sample size patterns, and three data types.

For example, for the $p \times i$ design in Table 5 with Variance Component Structure A, Sample Size Pattern 1, and normal data, 1,000 datasets were simulated. Equation 3.1 in Table 6 was used to generate the datasets, and the specific parameter values for Variance Component Structure A and Sample Size Pattern 1 can be found in Table 5. Within each of the 1,000 datasets, three bootstrap procedures were applied, boot-*p*, boot-*i*, and boot-*p,i*; the three bootstrap procedures are virtually the same except for the facet(s) that are resampled. Therefore, 999 bootstrap samples were taken from each dataset using boot-*p*; 999 bootstrap samples were taken from each dataset using boot-*i*; and 999 bootstrap samples were taken from each dataset using boot-*p*,*i*.

For the 999 bootstrap samples obtained using boot-*p*, estimates of the variance components $\hat{\sigma}^2(p)$ (Equation A.51), $\hat{\sigma}^2(i)$ (Equation A.52), and $\hat{\sigma}^2(pi)$ (Equation A.53) were calculated, as well as estimates of the relative error variance $\hat{\sigma}^2(\delta)$ (Equation A.151) and the absolute error variance $\hat{\sigma}^2(\Delta)$ (Equation A.156). The estimates of the variance components were corrected for bias using Equations C.3-C.5; therefore, any subsequent reference to estimated variance components in the analyses are technically referring to bias-corrected variance component estimates. Across the 999 bootstrap samples from a given dataset, the means of the 999 estimates of the variance components and relative and absolute error variances were calculated, and the variances of the 999 estimates of the variance components and relative and absolute error variances were calculated. The same procedure was repeated for boot-*i* and boot-*p*,*i*.

Within each dataset, an estimate of the squared standard error of the absolute error variance $SE(\Delta)$ (Equation A.171) was also calculated, based on the variances of the relevant statistics calculated across the 999 bootstrap samples from the appropriate bootstrap procedures as suggested by Tong and Brennan (2007). For the $p \times i$ design, these statistics were the variance of $\hat{\sigma}^2(pi)$ obtained from the boot-*p* procedure and the variance of $\hat{\sigma}^2(i)$ obtained using the boot-*i* procedure.

Additionally, for the estimates of the variance components and relative and absolute error variances within each bootstrap procedure, the lower and upper bounds for three types of 90% confidence intervals were calculated within each dataset. A value of one was assigned if the parameter value fell within the resulting confidence interval, or a value of zero was assigned if the parameter value fell outside the interval.

Across the 1,000 datasets and for each bootstrap procedure, the square roots of the means of the variances of the three estimated variance components and relative error variance were taken as the final bootstrap estimates of the standard errors of the variance components and relative and absolute error variances. The square root of the mean of the 1,000 estimates of the squared standard error of the absolute error variance was taken as the final workaround estimate of the standard error of the absolute error variance. The number of confidence intervals containing the parameter value was calculated across the 1,000 datasets for the variance components and relative and absolute error variances for the three types of confidence intervals. These counts were taken as the confidence interval coverage for the standard normal, percentile, and bias-corrected percentile confidence intervals.

Although bias-corrected means for the estimated variance components for each dataset were calculated, these bias-corrected estimates were only of interest in terms of ensuring that they were reasonably close to the parameter values. Brennan (2006; 2007) has proven that the bias-corrected means for the estimated variance components must equal the parameters; provided, of course, that the random number generator works well, there are no programming errors, and the number of bootstrap samples approaches infinity.

This same basic procedure was followed for each condition of the study. The procedure was repeated for the other three combinations of conditions in Table 5, and for each of the four combinations of conditions for dichotomous data and polytomous data, yielding a total of 12 conditions for the $p \times i$ design. The procedure was slightly different for the other four G theory designs because different bootstrap procedures were used depending on the design, but otherwise the procedure was largely the same for the remaining 48 conditions.

<u>Criteria for Evaluating the Bootstrap Estimates</u>

Each of the bootstrap estimation procedures in each of the 60 conditions yielded estimates of all possible variance components for a given design, as well as the standard errors and confidence interval coverage corresponding to each estimated variance component. Relative and absolute error variances and their standard errors and confidence interval coverage were produced for each bootstrap procedure. A single estimate of the standard error of the absolute error variance using Tong and Brennan's (2007) proposed workaround was also produced for each of the 60 conditions. The criteria for determining the adequacy of the standard error estimates were as follows.

For the normally distributed data, the estimated standard errors were compared to the exact standard errors calculated using Equation 3.6. For the dichotomous data, the estimated standard errors were compared to the "parameter values" calculated based on averages across 5,000 replications. For the polytomous data, the estimated standard errors were compared to the "parameter values" calculated based on averages across 5,000 replications. For all three data types, the estimated standard errors for the estimated relative and absolute error variances were compared to the parameter values.

Tong and Brennan (2007) used the following criteria for judging whether the estimates were close to the parameter values: "If an estimate deviated more than 5% in absolute magnitude from its corresponding parameter, the difference is considered large" (p. 809). This study started with the 5% criterion, but found it to be very stringent, such that nearly 40% of the selected standard error estimates were considered to be large. Furthermore, many of the estimates that were flagged as deviant based on the 5% criterion (i.e., in relative terms) were very close to the parameter values in absolute terms (i.e., differing by no more than 0.001). Therefore, the criterion was modified such that if an estimate deviated 10% or more from its respective parameter, the difference was considered to be large. This modified criterion resulted in 25% of the selected standard error estimates being designated as large differences from the parameter values, but the estimates that were not considered large were reasonably close to their respective parameter values. It should be noted, however, that common sense should be used when drawing conclusions regarding differences between parameter values and estimates. In particular, when the standard errors are very small, the magnitude of difference might be large, while the estimates themselves may still be good enough for practical use.

The confidence interval coverage for the three types of confidence intervals calculated for each statistic of interest across G theory designs, variance component structures, sample size patterns, data types, and bootstrap procedures provided additional evidence of the adequacy of the estimates (i.e., whether approximately 90% of the 90% confidence intervals contain the parameter value). If too few of the intervals contained the parameter value, then that provided evidence that the bootstrap procedures may have produced inaccurate variance component estimates or underestimated the standard error estimates.

CHAPTER 4: RESULTS AND DISCUSSION

The main purpose of this study was to test the rules suggested by Tong and Brennan (2007) for choosing optimal bootstrap procedures to estimate standard errors of estimated variance components under a broader set of conditions than those studied by Tong and Brennan (2007). The current study utilized five G theory designs, crossed with two variance component structures and two sample size patterns within each design, crossed with three data types, yielding 60 combinations of conditions. Point estimates, standard error estimates, and three types of confidence interval coverage were calculated for each of the variance components and relative and absolute error variances across several different bootstrap procedures for each of the combinations of conditions. Estimates were evaluated with respect to how close they were to the corresponding parameter values in terms of the magnitude of difference between the estimate and parameter value. Equation 4.1 was used to calculate the magnitude of difference between an estimate and its corresponding parameter value, expressed as a percentage, where θ refers to the parameter value of interest and $\hat{\theta}$ refers to the estimate of interest.

$$100 \times \left(\frac{\hat{\theta} - \theta}{\theta} \right) \tag{4.1}$$

Estimated variance components and relative and absolute error variances, as well as estimated standard errors of estimated variance components and relative and absolute error variances were based on 999 bootstrap samples taken from 1,000 generated datasets. Therefore, each estimate was based on 999,000 replications.

Estimated Variance Components and Relative and Absolute Error Variances

The estimated variance components and relative and absolute error variances were not the main focus of this dissertation, but it is important to determine the adequacy of the estimated variance components and relative and absolute error variances before examination of the estimated standard errors of the estimated variance components and relative and absolute error variances. If the estimates of the variance components and relative and absolute error variances themselves are not accurate, there can be little confidence in the adequacy of the estimated standard errors.

The estimated variance components and relative and absolute error variances for all of the combinations of conditions in this study can be found in Appendix F, Tables F1 – F15. Empirical estimates of the variance components and relative and absolute error variances are also presented; that is, the estimates of variance components and relative and absolute error variances obtained without bootstrapping from each of the 1,000 sets of generated data for each combination of conditions, averaged across the 1,000 datasets. These estimates were calculated to verify that the generated datasets produced estimates that were close to the parameter values.

Overall, across the 60 combinations of conditions of this study, the estimated variance components and relative and absolute error variances were reasonably close to the parameter values. The estimated variance components and relative and absolute error variances were typically not identical to the parameter values, but the estimates typically deviated by less than 1% from their corresponding parameter values.

The largest discrepancies between variance components estimates and parameter values were typically estimates of variance components involving the *h* facet, which tended to have the smallest sample size, ranging from $n_h = 2$ to $n_h = 10$. Discrepancies between estimates and parameter values are likely to be due in part to the instability resulting from small sample sizes. Additionally, because the bootstrap methodology is based on asymptotic theory, and because a finite number of bootstrap samples were drawn, sampling error could have played a role in the discrepancies between estimates and parameter values. It should also be noted that the bootstrap estimates tended to be close to the empirical parameter estimates, which were averaged across estimates calculated directly from each dataset without bootstrapping.

The discrepancies between the parameter values and estimates of the variance components and relative and absolute error variances were similar to those found in Tong and Brennan (2007). Their results also showed the greatest discrepancies in the facets with the smallest sample sizes (in their case, $n_h = 2$). For example, for the $p \times i \times h$ design using polytomous data, their estimates of $\sigma^2(h)$ deviated as much as 30% from the parameter value, which is comparable to the findings of the current study.

The relative and absolute error variance estimates tended to be closer to the parameter values. Only a few of the estimates deviated by more than 1% from their corresponding parameter values, and all were within 5% of the parameter values. The maximum discrepancies for the relative and absolute error variances were found for the $p \times i \times h$ design, Sample Size pattern 1, with polytomous data, for which the estimated absolute error variance deviated by about 3% across all seven bootstrap procedures.

Recall from Chapter 1 of this dissertation the distinction between generalizability studies (G studies), in which variance components are estimated, and decision studies (D studies), in which decisions are made based on the D study results. The relative and absolute error variances are the estimates that are used to make decisions; therefore, it is particularly important that these estimates are accurate, and it is reassuring to note that these estimates tended to be close to the parameter values.

Estimated Standard Errors of Estimated Variance Components and Relative and Absolute Error Variances

Estimated standard errors of estimated variance components and relative and absolute error variances are presented in Appendix G, Tables G1 – G15. The differences between the standard error estimates and their respective parameter values are presented in Appendix H, Tables H1 – H15. These differences are expressed as a percentage of the parameter value, and were calculated using Equation 4.1. Positive values indicate over-estimation of the parameter value, and negative values indicate under-estimation of the parameter value.

Within each of the tables in Appendices G and H, the estimates that would have been selected based on Tong and Brennan's (2007) rules (Appendix D) are bolded, and the estimates that would have been selected based on the results of this study are boxed. Of the estimates that would have been selected based on the results of this study, those which differed from the parameter values by 10% or more are shaded, indicating that while these estimates were determined to be the optimal estimates, they still differed substantially from the parameter values.

Due to the large scope of this study, the results needed to be synthesized in such a way as to allow for comparisons between conditions that would enable the researcher to identify patterns and determine the extent to which Tong and Brennan's (2007) rules held up across the various combinations of conditions. This was accomplished in the following manner. First, the magnitudes of the differences between estimates and parameter values that are presented in Appendix H were calculated using Equation 4.1. Then, for each standard error within each combination of G theory design and data type, the differences were aggregated across variance component structures and sample size patterns for each bootstrap procedure to determine which bootstrap procedure produced estimates that were, on average, closest to the corresponding parameter values. For example, for the $p \times i$ design with normal data, the differences between estimates and parameter values for *SE*(*p*) obtained using boot-*p* were aggregated across the four combinations of variance component structure and sample size pattern. These aggregated differences were compared to the aggregated differences obtained using boot-*i* and boot-*p,i* to determine which of the three bootstrap procedures, on average, produced the most accurate estimates of *SE*(*p*). This same procedure was followed for the rest of the standard errors for all of the combinations of conditions of this study.

The results of these analyses are presented below. As noted previously, the estimates and percent differences in Appendices G and H that are in boxes are the estimates and percent differences corresponding to the bootstrap procedures determined to be optimal based the results of this study, not to be confused with the most accurate estimates that were actually obtained within each combination of conditions. These are the estimates based on the bootstrap procedures that, on average, produced the most accurate estimates of each standard error.

It should be noted that the tables in Appendices G and H also contain standard error estimates calculated using normality-based equations (Appendix A, Tables A23 – A27, A30, and A31). These estimates are discussed in a later section of this chapter.

The $p \times i$ Design

Estimated standard errors of estimated variance components and relative and absolute error variances for the $p \times i$ design are presented in Appendix G, Tables G1 – G3. The signed differences between the standard error estimates and corresponding parameter values, expressed as a percentage of the parameter value, are presented in Appendix H, Tables H1 – H3. Within each of these tables, the estimates that would have been selected based on Tong and Brennan's (2007) rules are bolded, and the estimates that would have been selected based on the results of the current study are boxed. Of the estimates that would have been selected based on the results of this study, those which differed from the parameter values by 10% or more are shaded.

Normal Data

For normal data (Tables G1 and H1), boot-*p* provided the best estimates of $SE(p)$, $SE(pi)$, and $SE(\delta)$ across the four combinations of variance component structure and sample size pattern. All of these estimates differed by less than 1% from their respective parameter values. Boot-*i* provided the best estimates of $SE(i)$ across the four conditions, differing by less than 2% from the parameter values. Boot-*i* also provided the best estimates of *SE*(Δ) across the four combinations of variance component structure and sample size pattern. The boot-*i* estimates of *SE*(Δ) were within 5% of the parameter values for three of the four combinations; for Variance Component Structure A, Sample Size Pattern 1, the boot-*i* estimate of *SE*(Δ) deviated 40% from the parameter value, but the boot-*p* and boot-*p,i* estimates were also inaccurate, deviating 32% and 68% from the parameter value, respectively.

Dichotomous Data

For dichotomous data (Tables G2 and H2), boot-*p* generally produced the best estimates of *SE*(*p*) across the four combinations of variance component structure and sample size pattern. However, the estimates deviated from their parameter values by between 6% (Variance Component Structure A, Sample Size Pattern 2) and 40% (Variance Component Structure B, Sample Size Pattern 1). Boot-*i* tended to produce the best estimates of *SE*(*i*), with estimates deviating by 4% to 10% from the parameter values. Unlike the results for the normal data, boot-*p,i* tended to produce the best estimates of *SE*(*pi*), *SE*(*δ*), and *SE*(Δ) across the four combinations of variance component structure and sample size pattern, with deviations from the parameter values ranging 1% to 17% for *SE*(*pi*) and *SE*(*δ*), and 2% to 16% for *SE*(Δ).

Polytomous Data

For polytomous data (Tables G3 and H3), boot-*p* produced the best estimates of *SE*(*p*), *SE*(*pi*), and *SE*(*δ*) across the four combinations of variance component structure and sample size pattern, deviating no more than 2% from the parameter value for each of these estimates. Boot-*p,i* produced slightly better estimates of *SE*(*i*) than did boot-*i*, but the difference was small, and the estimates deviated from the parameter values by less than 10% for both boot-*i* and boot-*p,i* for three of the four combinations of variance component structure and sample size pattern. For Variance Component Structure A, Sample Size Pattern 1, the estimate of *SE*(*i*) deviated 14% from the parameter value using boot-*p,i* and 16% from the parameter value using boot-*i*. Boot-*i* produced the best estimates of *SE*(Δ) for Sample Size Pattern 1 (where n_i = 10), and boot-*p,i* produced the best estimates of *SE*(Δ) for Sample Size Pattern 2 (where n_i = 50). Both boot-*i* and boot-*p,i* produced estimates that deviated from the parameter value less than 10%.

Overall, for the $p \times i$ design, the results of this study are consistent with the findings of Tong and Brennan (2007), with the exceptions of the estimates of *SE*(*pi*) and *SE*(δ) for dichotomous data. It should be noted that the variance component corresponding to the interaction between persons and items is confounded with all undefined sources of residual error, which may be why the optimal estimates of *SE*(*pi*) and *SE*(δ) found in this study for dichotomous data were not the same as those expected to be optimal.

The $p \times i \times h$ Design

Estimated standard errors of estimated variance components and relative and absolute error variances for the $p \times i \times h$ design are presented in Appendix G, Tables G4 – G6. The signed differences between the standard error estimates and corresponding parameter values, expressed as a percentage of the parameter value, are presented in Appendix H, Tables H4 – H6. Within each of these tables, the estimates that would have been selected based on Tong and Brennan’s (2007) rules are bolded, and the estimates that would have been selected based on the results of the current study are boxed. Of the estimates that would have been selected based on the results of this study, those which differed from the parameter values by 10% or more are shaded.

Normal Data

For normal data (Tables G4 and H4), boot-*p* produced the best estimates of *SE*(*p*), *SE*(*pi*), *SE*(*ph*), *SE*(*pih*), and *SE*(δ) across the four combinations of variance component structure and sample size pattern. All of these estimates were within 1% of the parameter values. Boot-*i* produced reasonable estimates of *SE*(*i*) across the four combinations of variance component structure and sample size pattern, with all four estimates deviating by less than 10% from the parameter values. Boot-*p,i* also produced reasonable estimates of *SE*(*i*) for Variance Component Structure B. Boot-*h* tended to produce the best estimates of *SE*(*h*), although the estimates were not particularly close to the parameter values for Sample Size Pattern 1, where $n_h = 2$. These estimates deviated from the parameter values by 26%, which was better than any of the other bootstrap procedures. Boot-*p,h* produced estimates of *SE*(*h*) which were close to the estimates produced by boot-*h*, but tended to be slightly further from the parameter values. Boot-*i* produced the best estimates of *SE*(*ih*) across the four combinations of variance component structure and sample size pattern, all of which deviated by less than 10% from the parameter value. None of the bootstrap procedures or the workaround consistently produced estimates of *SE*(Δ) that were close to the parameter values, although boot-*p,i* did produce the best estimate for two of the conditions, and the second best estimate for another of the conditions. The boot-*p,i* estimates of *SE*(Δ) deviated between 8% and 31% from the parameter values.

Dichotomous Data

For the $p \times i \times h$ design with dichotomous data (Tables G5 and H5), none of the bootstrap procedures produced standard error estimates that were uniformly the best across all four combinations of variance component structure and sample size pattern. For example, boot-*p* produced the best estimate of *SE*(*p*) for Variance Component Structure A, Sample Size Pattern 1, boot-*p,i* produced the best estimate of *SE*(*p*) for Variance Component Structure A, Sample Size Pattern 2, boot-*i* produced the best estimate of *SE*(*p*) for Variance Component Structure B, Sample Size Pattern 1, and boot-*h* produced the best estimate of *SE*(*p*) for Variance Component Structure B, Sample Size Pattern 2. However, boot-*p,i* produced the second best estimates for three of the four combinations of variance component structure and sample size pattern. The extent to which the boot-*p,i* estimates of *SE*(*p*) deviated from the parameter values ranged from 3% to 26%, and two of the estimates were within 10% of the parameter values.

Similar interpretive patterns persisted across the standard error estimates of the other variance components. Boot-*i* tended to produce more reasonable estimates of *SE*(*i*) than the other bootstrap procedures, with estimates deviating from the parameter values by between 9% and 18%. Boot-*h* tended to produce the most reasonable estimates of *SE*(*h*), with estimates deviating from the parameter values by 14% on average. Boot-*p,h* also produced reasonable estimates of *SE*(*h*) compared to the rest of the bootstrap procedures, but on average did not perform as well as boot-*h*, deviating on average by 15% from the parameter values. Boot-*p* and boot-*i* both produced reasonable estimates of *SE*(*pi*) compared to the other bootstrap procedures. However, boot-*i* performed slightly better than boot-*p*, with estimates deviating by less than 20% from the parameter values across all four combinations of variance component structures and sample size patterns. Boot-*p* produced the most reasonable estimates of *SE*(*ph*) compared to the rest of the bootstrap procedures. All of the boot-*p* estimates of *SE*(*ph*) deviated from the parameter values by 25% to 34%, whereas the other bootstrap procedures produced estimates that deviated by more than 200% from the parameter values. Boot-*i* tended to produce the most reasonable estimates of *SE*(*ih*), with estimates that differed by 13% on average from the parameter values, and all of the estimates deviated by less than 20% from the parameter values. Boot-*p,i* tended to produce the most reasonable estimates of *SE*(*pih*). All of the estimates deviated less than 30% from the parameter values across the four combinations of variance component structures and sample size patterns, whereas all of the other bootstrap procedures produced estimates that deviated more than 50% from the parameter values for at least one of the combinations of variance component structures and sample size patterns.

Boot-*h* tended to produce better estimates of *SE*(δ) compared to estimates produced by the other bootstrap procedures. All of the estimates deviated by less than

30% from the parameter values across the four combinations of variance component structures and sample size patterns. Boot-*p,h* also produced estimates of *SE*(δ) that were reasonable, and were closer to the parameter values than the estimates produced by boot-*h* for Variance Component Structure A. Boot-*p,h* tended to produce the best estimates of *SE*(Δ) compared to the other bootstrap procedures. All of the boot-*p,h* estimates deviated no more than 14% from the parameter values, and deviated by 10% on average from the parameter values.

Polytomous Data

For polytomous data (Tables G6 and H6), boot-*p* produced the best estimates of *SE*(*p*), *SE*(*ph*), *SE*(*pih*), and *SE*(δ) across the four combinations of variance component structures and sample size patterns, and produced the best estimates of *SE*(*pi*) for three of the four combinations of variance component structures and sample size patterns.

Boot-*i* and boot-*p,i* both produced the most reasonable estimates of *SE*(*i*) across the four combinations of variance component structures and sample size patterns, with boot-*p,i* slightly outperforming boot-*i*. All of the estimates for both bootstrap procedures varied from the parameter values by less than 25%, with the boot-*i* estimates differing by 14% on average from the parameter values, and the boot-*p,i* estimates differing by 12% on average.

Boot-*p,i,h* produced the best estimates of *SE*(*h*) across the four combinations of variance component structures and sample size patterns, with estimates deviating by no more than 25% from the parameter values. Boot-*h*, boot-*p,h*, and boot-*i,h* also performed reasonably well relative to the rest of the bootstrap procedures, with estimates deviating by no more than 30% from the parameter values. The other bootstrap procedures produced estimates that differed from the parameter values by 60% to 90%.

None of the bootstrap procedures consistently produced the best estimates of *SE*(*ih*) across the four combinations of variance component structure and sample size pattern. However, when taking into account the relative performance of each bootstrap procedure across the four combinations of conditions, boot-*p,h* produced the best estimates of *SE*(*ih*), closely followed by boot-*p,i*, boot-*h*, and boot-*i*. On average, estimates of *SE*(*ih*) deviated from the parameter values by 14%, 15%, 16%, and 18% for the boot-*p,h*, boot-*p,i*, boot-*h*, and boot-*i* procedures, respectively. The maximum deviation for these bootstrap procedures was less than 30% across the four combinations of variance component structures and sample size patterns.

Like the estimates of *SE*(*ih*), there was not a clear cut best procedure for estimating *SE*(Δ). Boot-*h* and boot-*p,h*, produced estimates that were on average closer to the parameter value than the other bootstrap procedures, with differences of 19% and 18%, for boot-*h* and boot-*p,h*, respectively, and all of the estimates for these two bootstrap procedures deviated from the parameter values by less than 40%.

Overall, for the $p \times i \times h$ design, the results of this study are consistent with the findings of Tong and Brennan (2007), with the exceptions of the estimates of *SE*(*pih*) and *SE*(δ) for dichotomous data. As noted for the $p \times i$ design, the variance component corresponding to the highest order interaction is confounded with residual error, which may explain why the optimal estimates of *SE*(*pih*) and *SE*(δ) found in this study for dichotomous data were not the same as those expected to be optimal.

The $i : p$ Design

Estimated standard errors of estimated variance components and relative and absolute error variances for the $i : p$ design are presented in Appendix G, Tables G7 – G9. The signed differences between the standard error estimates and corresponding parameter values, expressed as a percentage of the parameter value, are presented in Appendix H, Tables H7 – H9. Within each of these tables, the estimates that would have been selected based on Tong and Brennan's (2007) rules are bolded, and the estimates that would have been selected based on the results of the current study are boxed. Of the estimates that would have been selected based on the results of this study, those which differed from the parameter values by 10% or more are shaded. It should be noted that for the $i : p$ design, $\sigma^2(\delta) = \sigma^2(\Delta)$, and therefore $SE(\delta) = SE(\Delta)$.

Normal Data

For normal data (Tables G7 and H7), boot-*p* produced the best estimates of $SE(p)$ across the four combinations of variance component structure and sample size pattern, with estimates deviating from the parameter value by less than 2%. Boot-*p* produced the best estimates of $SE(i : p)$, $SE(\delta)$, and $SE(\Delta)$ for Sample Size Pattern 1, where $n_p = 50$ and $n_i = 5$; all of the estimates were within 2% of the parameter values. Boot-*p* and boot-*i* both produced estimates of $SE(i : p)$, $SE(\delta)$, and $SE(\Delta)$ that were within 2% of the parameter values for Sample Size Pattern 2, where $n_p = n_i = 50$.

Dichotomous Data

For dichotomous data (Tables G8 and H8), boot-*p* produced the best estimates of $SE(p)$, $SE(i : p)$, $SE(\delta)$, and $SE(\Delta)$ across the four combinations of variance component structure and sample size pattern. The estimates for Sample Size Pattern 1 were somewhat closer to the parameter values than the estimates for Sample Size Pattern 2. Estimates for Sample Size Pattern 1 were all within 2% of the parameter values, whereas estimates for Sample Size Pattern 1 differed by 1.5% to 7%.

Polytomous Data

For polytomous data (Tables G9 and H9), boot-*p* produced the best estimates of $SE(p)$ across the four combinations of variance component structure and sample size pattern, with estimates deviating from the parameter value by no more than 6%. Boot-*p* produced the best estimates of $SE(i : p)$, $SE(\delta)$, and $SE(\Delta)$; the estimates deviated no more than 2%. Boot-*i* also produced reasonable estimates of $SE(i : p)$, $SE(\delta)$, and $SE(\Delta)$

for Sample Size Pattern 2, where $n_p = n_i = 50$; the estimates were within 2% of the parameter values.

Overall, for the $i : p$ design, the results of this study are consistent with the findings of Tong and Brennan (2007) for estimating $SE(p)$ and $SE(\delta)$; however, the current study found a different optimal bootstrap procedure for estimating $SE(i : p)$.

The $i : h : p$ Design

Estimated standard errors of estimated variance components and relative and absolute error variances for the $i : h : p$ design are presented in Appendix G, Tables G10 – G12. The signed differences between the standard error estimates and corresponding parameter values, expressed as a percentage of the parameter value, are presented in Appendix H, Tables H10 – H12. Within each of these tables, the estimates that would have been selected based on Tong and Brennan's (2007) rules are bolded, and the estimates that would have been selected based on the results of the current study are boxed. Of the estimates that would have been selected based on the results of this study, those which differed from the parameter values by 10% or more are shaded. It should be noted that as with the $i : p$ design, $\sigma^2(\delta) = \sigma^2(\Delta)$, and therefore $SE(\delta) = SE(\Delta)$, in the $i : h : p$ design.

Normal Data

For normal data (Tables G10 and H10), boot-p produced the best estimates of $SE(p)$, $SE(h : p)$, $SE(i : h : p)$, $SE(\delta)$, and $SE(\Delta)$ for three of the four combinations of variance component structure and sample size pattern, with estimates deviating by no more than 2% from the parameter values. For Variance Component Structure B, Sample Size Pattern 2, boot-p,i produced the best estimates of $SE(p)$, $SE(h : p)$, $SE(\delta)$, and $SE(\Delta)$, and boot-p,h produced the best estimate of $SE(i : h : p)$. However, for all but $SE(p)$, these estimates deviated from the parameter values by about 25%. The boot-p estimates of $SE(p)$, $SE(h : p)$, $SE(i : h : p)$, $SE(\delta)$, and $SE(\Delta)$ for Variance Component

Structure B, Sample Size Pattern 2 all deviated from the parameter values by about 30%; therefore, the boot-*p* estimates were not unreasonable relative to the rest of the bootstrap procedures.

Dichotomous Data

For dichotomous data (Tables G11 and H11), boot-*p* produced the best estimates of $SE(p)$, $SE(h:p)$, $SE(i:h:p)$, $SE(\delta)$, and $SE(\Delta)$ across the four combinations of variance component structure and sample size pattern. The boot-*p* estimates deviated from the parameter values by no more than 3%.

Polytomous Data

For polytomous data (Tables G12 and H12), boot-*p* tended to produce the best estimates of $SE(p)$, $SE(h:p)$, $SE(i:h:p)$, $SE(\delta)$, and $SE(\Delta)$ across the four combinations of variance component structure and sample size pattern. The boot-*p* estimates deviated from the parameter values by no more than 5%.

Overall, for the $i:h:p$ design, the results of this study are consistent with the findings of Tong and Brennan (2007) for estimating $SE(p)$ and $SE(\delta)$; however, the current study found different optimal bootstrap procedures for estimating $SE(h:p)$ and $SE(i:h:p)$.

The $p \times (i:h)$ Design

Estimated standard errors of estimated variance components and relative and absolute error variances for the $i:h:p$ design are presented in Appendix G, Tables G13 – G15. The signed differences between the standard error estimates and corresponding parameter values, expressed as a percentage of the parameter value, are presented in Appendix H, Tables H13 – H15. Within each of these tables, the estimates that would have been selected based on Tong and Brennan's (2007) rules are bolded, and the estimates that would have been selected based on the results of the current study are boxed. Of the estimates that would have been selected based on the results of this study, those which differed from the parameter values by 10% or more are shaded.

Normal Data

For normal data (Tables G13 and H13), boot-*p* produced the best estimates of $SE(p)$, $SE(ph)$, $SE(pi:h)$, and $SE(\delta)$ across the four combinations of variance component structure and sample size pattern. All of these estimates deviated by less than 1% from the parameter values. Boot-*h* produced the best estimates of $SE(h)$ for Sample Size Pattern 2, where $n_h = 10$, and boot-*p,h* produced the best estimates of $SE(h)$ for Sample Size Pattern 1, where $n_h = 5$. However, the boot-*h* estimates of $SE(h)$ for Sample Size Pattern 1 were also reasonable, and all of the boot-*h* estimates of $SE(h)$ deviated less than 7% from the parameter values. Boot-*i* produced the best estimates of $SE(i:h)$ for Sample Size Pattern 1, where $n_i > n_h$; and boot-*h* produced the best estimates of $SE(i:h)$ for Sample Size Pattern 2, where $n_i < n_h$. However, boot-*h* produced reasonable estimates of $SE(i:h)$ across the four combinations of variance component structure and sample size pattern, with estimates deviating from the parameter values by less than 10%.

None of the bootstrap procedures produced reasonable estimates of $SE(\Delta)$ across the four combinations of variance component structure and sample size pattern. Boot-*p* produced the best estimates of $SE(\Delta)$ for Variance Component Structure A, and Boot-*h* produced estimates of $SE(\Delta)$ for Variance Component Structure B that were reasonable relative to estimates produced by the other bootstrap procedures. Boot-*h* estimates deviated the least from the parameter values on average across the four combinations of variance component structure and sample size pattern; estimates deviated from the parameter values by less than 30% for three of the four combinations of conditions. However, for Variance Component Structure A, Sample Size Pattern 1, the boot-*h* estimate of $SE(\Delta)$ deviated from the parameter value by 90%. Boot-*p* estimates tended to deviate further from the parameter values on average, with estimates deviating from the parameter values by 21% to 73%.

Dichotomous Data

For dichotomous data (Tables G14 and H14), boot-*p,i* tended to produce the best estimates of *SE*(*p*); all of the estimates deviated from the parameter values by less than 20%. Boot-*h* tended to produce the best estimates of *SE*(*h*); all of the estimates deviated from the parameter values by less than 10%. Boot-*p* tended to produce reasonable estimates of *SE*(*ph*), with estimates deviating by less than 20% from the parameter values for three of the four combinations of variance component structure and sample size pattern. Boot-*h* produced better estimates of *SE*(*ph*) than those produced by boot-*p* for Variance Component Structure B; however, the boot-*h* estimates of *SE*(*ph*) were not substantially more accurate than the boot-*p* estimates, differing by less than six percentage points. Boot-*h* produced the best estimates of $SE(i:h)$ for Variance Component Structure A, deviating by about 11% from the parameter values; and boot-*p,h* produced the best estimates of $SE(i:h)$ for Variance Component Structure B, deviating by about 5% from the parameter values. However, the boot-*h* estimates for Variance Component Structure B were not unreasonable, deviating by about 11% from the parameter values. Boot-*p,h* produced the best estimates of $SE(pi:h)$ for three of the four combinations of variance component structure and sample size pattern. Three of the four estimates deviated from the parameter values by less than 5%, and the maximum deviation was 16%.

None of the bootstrap procedures consistently produced reasonable estimates of $SE(\delta)$ across the four combinations of variance component structure and sample size pattern. Boot-*h* produced estimates that deviated by 30% on average from the parameter values, and the deviations ranged from 9% to 80% from the parameter values. Boot-*p* produced estimates that differed from the parameter values by 39% on average, and the deviations ranged from 20% to 59%.

Boot-*h* produced the best estimates of $SE(\Delta)$ across three of the four combinations of variance component structure and sample size pattern, with estimates deviating from the parameter values by less than 10%. For Variance Component Structure A, Sample Size Pattern 1, the boot-*h* estimate of $SE(\Delta)$ deviated from the parameter value by 61%.

Polytomous Data

For polytomous data (Tables G15 and H15), boot-*p* produced reasonable estimates of $SE(p)$, $SE(ph)$, $SE(pi:h)$, and $SE(\delta)$, with all estimates deviating from the parameter values by no more than 2%. Boot-*h* and boot-*p,h* both produced reasonable estimates of $SE(h)$, with boot-*p,h* producing estimates that were slightly closer to the parameter values. The boot-*h* and boot-*p,h* estimates of $SE(h)$ for Sample Size Pattern 2, where $n_h = 10$, differed from the parameter values by 4% and 7% on average, respectively. The estimates for Sample Size Pattern 1, where $n_h = 5$, were larger, averaging 22% from the parameter values. Boot-*p,i* produced the best estimates of $SE(i:h)$ for Sample Size Pattern 1, where $n_i > n_h$; and boot-*p,h* produced the best estimates of $SE(i:h)$ for Sample Size Pattern 2, where $n_i < n_h$. These estimates all differed from the parameter values by less than 10%.

None of the bootstrap procedures produced estimates of $SE(\Delta)$ that were uniformly the best across the four combinations of variance component structure and sample size pattern. Boot-*h* and boot-*p,h* both produced reasonable estimates; boot-*h* produced estimates that deviated by 8% on average from the parameter values, and by no more than 15%; and boot-*p,h* produced estimates that deviated by 10% on average from the parameter values, and by no more than 16%.

Overall, for the $p \times (i:h)$ design, the results of this study are consistent with the findings of Tong and Brennan (2007) for all but $SE(i:h)$ for normal and polytomous data. For dichotomous data, the results of the current study are consistent with Tong and Brennan's (2007) findings for the estimates of $SE(p)$, $SE(h)$, and $SE(ph)$, but found different optimal estimates of $SE(i:h)$, $SE(pi:h)$, and $SE(\delta)$. As was the case for the $i:p$ and $i:h:p$ designs, the current study found different optimal bootstrap procedures for estimating standard errors of estimated variance components that involved nesting. Additionally, as was the case for the $p \times i$ and $p \times i \times h$ designs, the current study found different optimal bootstrap procedures for estimating standard errors of estimated variance components for the highest order interactions, which contain residual error and may explain why the optimal estimates of $SE(pi:h)$ and $SE(\delta)$ found in this study for dichotomous data were not the same as those expected to be optimal.

Overall, across the five G theory designs, the rules proposed by Tong and Brennan (2007) were largely supported for normal and polytomous data, with some exceptions, and the rules were supported to a lesser extent for dichotomous data. Across all three data types, the rules held up for estimating standard errors of variance components for non-nested main effects and simple interactions, but this study found different optimal bootstrap procedures for estimating effects involving nesting. For dichotomous data, this study also found different optimal bootstrap procedures for estimating the standard errors of variance components corresponding to the highest order interactions and relative error variance. It should be noted that the variance components corresponding to the highest order interactions are residual terms, containing all sources of variance that are unaccounted for in a particular G theory design. Additionally, these residual terms are components of the relative error variance estimates. It is possible that these residual effects may explain the differences found for dichotomous data versus normal and polytomous data.

Workaround for Estimating the Standard Error of the Estimated Absolute Error Variance

The estimated standard errors of the estimated absolute error variances using the workaround proposed by Tong and Brennan (2007) are presented in Appendix I, Tables I1 – I5, along with the percent difference between the estimates and their respective parameter values. These differences are expressed as a percentage of the parameter value, and were calculated using Equation 4.1. Positive values indicate over-estimation of the parameter value, and negative values indicate under-estimation of the parameter value.

In general, the workaround for estimating the standard error of the estimated absolute error variance tended to produce reasonable estimates of *SE*(Δ), particularly for normal and polytomous data. Estimates of *SE*(Δ) were particularly close to the parameter values for the $i:p$ and $i:h:p$ designs; only one of the estimates of *SE*(Δ) for these designs deviated from the parameter values by more than 3%. However, the estimated standard errors of the estimated absolute error variances for these nested designs are based only on the estimated standard errors of the estimated relative error variance, as can be seen in Appendix A, Table A32; thus the workaround estimates for these two G theory designs are redundant to the optimal bootstrap estimates of *SE*(δ), which were generally found to be reasonable.

Estimates of *SE*(Δ) tended to be larger for dichotomous data for the $p \times i$, $p \times i \times h$, and $p \times (i:h)$ designs, deviating near 40%, on average, from the parameter values. However, some differences were found between the optimal bootstrap procedures for estimating standard errors suggested by Tong and Brennan (2007) and those determined to be optimal in the current study, particularly for dichotomous data. Therefore, this section presents the results of investigating changes to the workaround proposed by Tong and Brennan (2007) for estimating the standard error of the estimated absolute error variance, based on the optimal bootstrap procedures found in the results this study.

The $p \times i$ Design

For the $p \times i$ design with normal and polytomous data, the workaround estimates of *SE*(Δ) were close to the parameter values; all of the workaround estimates deviated by no more than 15% from the parameter values.

For the $p \times i$ design with dichotomous data, the workaround estimates of *SE*(Δ) were generally not as close to the parameter values as were the boot-*p,i* estimates (Tables G2 and H2); the only workaround estimate that was close was for Variance Component Structure A, Sample Size Pattern 2, for which the workaround estimate was 15% from the parameter value, and the boot-*p,i* estimate was 8% from the parameter value. The rest of the workaround estimates deviated from the parameters by 38% (Variance Component Structure B, Sample Size Pattern 2) to 55% (Variance Component Structure B, Sample Size Pattern 1).

When the data were dichotomous, boot-*p,i* tended to produce better estimates of *SE*(*δ*) than boot-*p*. Therefore, an alternate workaround for calculating the estimate of the standard error of the estimated absolute error variance was also applied to the conditions involving dichotomous data using the following equation.

$$\sqrt{S\hat{E}^2(\delta \mid p,i) + \frac{S\hat{E}^2(i \mid i)}{n_i^2}} \qquad (4.2)$$

Table 9 contains the parameter values and two estimates of *SE*(Δ). The first column describes the combination of conditions for each row; for example, A1 refers to Variance Component Structure A and Sample Size Pattern 1. The second column contains the parameter values. The third column of the table contains the original workaround estimates of SE(Δ), which are also presented in Appendix I, Table I1, and the last column of the table contains the alternate estimates of *SE*(Δ) which were calculated using Equation 4.2. The modified workaround typically produced estimates of $SE(\Delta)$ that were closer to the parameter values than those produced by the workaround suggested by Tong and Brennan (2007).

Table 9. Alternate Estimates of $SE(\Delta)$ for the $p \times i$ Design Using Optimal Bootstrap Procedures

	Parameter $SE(\Delta)$	Tong & Brennan $S\hat{E}(\Delta)$	Moore $S\hat{E}(\Delta)$
		Dichotomous	
A1	0.0011	0.0007	0.0013
A2	0.0002	0.0002	0.0002
B1	0.0021	0.0010	0.0018
B2	0.0002	0.0001	0.0002

The $p \times i \times h$ Design

For the $p \times i \times h$ design with normal data, the workaround estimates of $SE(\Delta)$ typically produced reasonable estimates of $SE(\Delta)$; the estimates deviated from the parameter values by no more than 20%. For dichotomous data, the workaround for estimating $SE(\Delta)$ produced estimates of $SE(\Delta)$ that were, on average, not as close to the parameter values as those produced by the boot-*i*, boot-*h*, boot-*p*,*i*, and boot-*p*,*h* procedures (Tables G5 and H5), differing by 40% on average from the parameter values. For polytomous data, the workaround estimates of $SE(\Delta)$ deviated by 20% on average from the parameter values, but all of the estimates deviated by less than 30%.

In terms of the standard error estimates that enter into the workaround proposed by Tong and Brennan (2007), there were differences between the optimal bootstrap procedures proposed by Tong and Brennan (2007) and the optimal bootstrap estimates that were found to be optimal in the current study. Specifically, when the data were dichotomous, boot-*p*,*h* tended to produce better estimates of $SE(\delta)$ than boot-*p*.

Therefore, for the $p \times i \times h$ design using dichotomous data, the following equation was also used to estimate the standard error of the absolute error variance estimates.

$$\sqrt{S\hat{E}^2(\delta \mid p,h) + \frac{S\hat{E}^2(i \mid i)}{n_i^2} + \frac{S\hat{E}^2(h \mid h)}{n_h^2} + \frac{S\hat{E}^2(ih \mid i)}{n_i^2 n_h^2}} \quad (4.3)$$

Table 10 contains the parameter values and two estimates of $SE(\Delta)$. The first column describes the combination of conditions for each row; for example, A1 refers to Variance Component Structure A and Sample Size Pattern 1. The second column contains the parameter values. The third column of the table contains the original workaround estimates of SE(Δ), which are also presented in Appendix I, Table I2, and the last column of the table contains the alternate estimates of $SE(\Delta)$ which were calculated using Equation 4.3. The modified workaround produced better estimates of $SE(\Delta)$ across all four combinations of variance component structures and sample size patterns; these estimates deviated no more than 15% from the parameter values.

The $i : p$ and $i : h : p$ Designs

The workaround for estimating $SE(\Delta)$ worked well for the $i : p$ and $i : h : p$ designs. However, for these designs, $S\hat{E}(\delta) = S\hat{E}(\Delta)$, meaning that the estimated standard error of the estimated relative error variance is the only component that enters into the equation for calculating $S\hat{E}(\Delta)$; therefore, the workaround is redundant for these designs. Additionally, the estimates of $SE(\delta)$ tended to be close to the parameter values for these designs, thus the estimates of $SE(\Delta)$ by definition were also close to the parameter values.

The $p \times (i : h)$ Design

For the $p \times (i : h)$ design with normal data, the workaround for estimating $SE(\Delta)$ generally produced the optimal estimates as compared to the bootstrap procedures; estimates produced by the workaround deviated from the parameter values by no more than 10%. When the data were dichotomous or polytomous, the workaround for estimating *SE*(Δ) did not perform as well, on average, as boot-*h* and boot-*p,h*, but the workaround estimates of *SE*(Δ) were generally reasonable, deviating, on average, 33% and 12% for dichotomous and polytomous data, respectively.

Table 10. Alternate Estimates of *SE*(Δ) for the $p \times i \times h$ Design Using Optimal Bootstrap Procedures

	Parameter $SE(\Delta)$	Tong & Brennan $S\hat{E}(\Delta)$	Moore $S\hat{E}(\Delta)$
		Dichotomous	
A1	0.0055	0.0036	0.0059
A2	0.0027	0.0017	0.0026
B1	0.0073	0.0042	0.0072
B2	0.0029	0.0016	0.0025

In terms of the standard error estimates that enter into the workaround proposed by Tong and Brennan (2007), there were differences between the optimal bootstrap procedures proposed by Tong and Brennan (2007) and the optimal bootstrap estimates that were found to be optimal in the current study. Specifically, boot-*h* was found to produce more accurate estimates of $SE(i:h)$ than boot-*i* across all three data types; for dichotomous data, boot-*p,h* was found to produce more accurate estimates of $SE(\delta)$ than boot-*p*. Therefore, the following equation was used to estimate the standard error of the absolute error variance estimates when the data were normal or polytomous,

$$\sqrt{S\hat{E}^2(\delta \mid p) + \frac{S\hat{E}^2(h \mid h)}{n_h^2} + \frac{S\hat{E}^2(i:h \mid h)}{n_i^2 n_h^2}}, \qquad (4.4)$$

and the following equation was used to estimate the standard error of the absolute error variance estimates when the data were dichotomous

$$\sqrt{S\hat{E}^2(\delta \mid p,h) + \frac{S\hat{E}^2(h \mid h)}{n_h^2} + \frac{S\hat{E}^2(i:h \mid h)}{n_i^2 n_h^2}} \,. \tag{4.5}$$

Table 11 contains the parameter values and two estimates of *SE*(Δ). The first column describes the combination of conditions for each row; for example, A1 refers to Variance Component Structure A and Sample Size Pattern 1. The second column contains the parameter values. The third column of the table contains the original workaround estimates of SE(Δ), which are also presented in Appendix I, Table I5; and the last column of the table contains the alternate estimates of *SE*(Δ) which were calculated using Equations 4.4 and 4.5. In general, the modified workaround tended to produce slightly better estimates of *SE*(Δ) for some of the combinations of variance component structures and sample size patterns to which they were applied, but did not bring the estimates substantially closer to the parameter values.

Overall, the workaround for estimating the standard error of the estimated absolute error variance produced reasonable estimates. Modification of the workaround to reflect the different optimal bootstrap procedures associated with some of the estimated standard errors of estimated variance components generally resulted in estimates that were closer to the parameter values than the original workaround estimates using the estimates associated with the bootstrap procedures suggested by Tong and Brennan (2007). Additionally, while some of the estimates of *SE*(Δ) deviated from the parameter values by more than 10%, particularly for dichotomous data, the actual differences between the estimates and parameter values were very small (in many cases less than 0.001), and would be good enough for practical use.

Confidence Interval Coverage

Confidence interval coverage for three different types of confidence intervals were calculated for each variance component and relative and absolute error variance, for each bootstrap procedure within each of the 60 combinations of conditions: standard normal, percentile, and bias corrected percentile confidence intervals. Confidence interval coverage was based on ninety percent confidence intervals; therefore, it was expected that 900 of the 1,000 replications would contain the parameter value if the assumptions of the confidence interval were met.

Table 11. Alternate Estimates of $SE(\Delta)$ for the $p \times (i:h)$ Design Using Optimal Bootstrap Procedures

	Parameter $SE(\Delta)$	Tong & Brennan $S\hat{E}(\Delta)$	Moore $S\hat{E}(\Delta)$
		Normal	
A1	0.0021	0.0021	0.0021
A2	0.0018	0.0020	0.0019
B1	0.0088	0.0086	0.0086
B2	0.0051	0.0056	0.0051
		Dichotomous	
A1	0.0003	0.0002	0.0005
A2	0.0004	0.0003	0.0005
B1	0.0010	0.0006	0.0010
B2	0.0009	0.0005	0.0008
		Polytomous	
A1	0.0084	0.0067	0.0067
A2	0.0037	0.0035	0.0036
B1	0.0217	0.0168	0.0168
B2	0.0085	0.0085	0.0082

The differences in coverage among the three types of confidence intervals were minimal for the most part. However, some of the bias corrected confidence interval coverage values were markedly smaller than those for the standard normal and percentile confidence intervals. For example, for the $p \times i$ design with normal data, Variance Component Structure B, and Sample Size Pattern 2, the standard normal confidence interval coverage for $\sigma^2(pi)$ using the boot-*p,i* procedure was 99.5%, the percentile confidence interval coverage was 89.8%, and the bias corrected confidence interval coverage was 20.1%. The parameter value of $\sigma^2(pi)$ was 0.5, the boot-*p,i* estimate of $\sigma^2(pi)$ was 0.5001, the parameter value of *SE*(*pi*) was 0.0102, and the estimated standard error of $\hat{\sigma}^2(pi)$ was 0.0175. Because the boot-*p,i* procedure over-estimated the standard error of $\sigma^2(pi)$, it would be expected that the confidence interval coverage would be wider than 90%, which was the case for the standard normal confidence interval coverage. However, the bias corrected confidence interval coverage was much lower, casting doubt on the adequacy of the bias corrected confidence interval coverage in capturing the variability of $\sigma^2(pi)$, at least for this combination of conditions.

Carpenter and Bithell (2000) suggested that the bias corrected confidence interval, while an improvement over the percentile confidence interval for asymmetric distributions, may still be inaccurate if the skewness of the distribution varies with θ. Because of this, and because of the underestimation of the variability of some of the estimated variance components using bias corrected confidence intervals, this section focuses on the percentile confidence interval coverage. Tables containing the percentile confidence interval coverage can be found in Appendix J, Tables J1 – J15. Within each of these tables, the coverage for the estimates that would have been selected based on Tong and Brennan's (2007) rules are bolded, and the estimates that would have been selected based on the results of the current study are boxed.

In general, the percentile confidence interval coverage provides support for the optimal bootstrap procedures for estimating standard errors of estimated variance components and relative and absolute error variances. For example, percentile confidence interval coverage for each estimated variance component and relative and absolute error variance for the $p \times i$ design can be found in Tables J1 – J3. The confidence interval coverage provides support for boot-*p* as the optimal bootstrap procedure for estimating *SE*(*p*), *SE*(*pi*), and *SE*(δ), and boot-*i* as the optimal procedure for estimating *SE*(i) for normal data; the coverage was near 90% across the four combinations of variance component structure and sample size pattern. The non-optimal estimates tended to have lower coverage where the estimated standard errors were too small, for example for the boot-p estimates of $\sigma^2(i)$, or higher coverage where the estimated standard errors were too large, for example for the boot-p,i estimates of $\sigma^2(pi)$.

For dichotomous and polytomous data, and particularly for estimates for which the sample size is small (such as the h facet, and in some cases the i facet), the confidence interval coverage values were not close to 90%. For example, most of the estimates of $\sigma^2(h)$ in the $p \times i \times h$ and $p \times (i:h)$ designs were found to have confidence interval coverage below 90%. Because of the small sample sizes for the h facet, the estimates are expected to be less accurate, and it is not surprising that fewer of the intervals contained the parameter value.

Overall, though, it appears that the confidence interval coverage reflected the optimal bootstrap procedure for estimating the standard error of each variance component and relative and absolute error variance, such that the optimal estimates for a given variance component or relative or absolute error variance were closer to 90% relative to the estimates calculated by other bootstrap procedures.

Normality-Based Equations for Estimating Standard Errors of Estimated Variance Components and Relative and Absolute Error Variances

One question that may arise from reading this dissertation is the extent to which bootstrap procedures produce estimates of standard errors of variance components and relative and absolute error variances that are superior to those that would be obtained using the equations that assume normality of score effects (Table A23 – A27, A30, and A31). Tables G2 – G15 in Appendix G contain the estimated standard errors of the estimated variance components and relative and absolute error variances resulting from applying those equations to dichotomous and polytomous data. These are estimates that are calculated within each dataset and averaged across the 1,000 datasets for each combination of conditions.

It can be seen that for the $p \times i$ design using dichotomous data, the estimates calculated using the normality-based equations tended to be too low; and for the $p \times i$ design using polytomous data, the estimates of $SE(p)$, $SE(i)$, and $SE(\Delta)$ tended to be too low, while the estimates of $SE(pi)$ and $SE(\delta)$ tended to be too high. For the $p \times i \times h$ design, all of the standard error estimates calculated using the normality-based equations tended to be too low using dichotomous data, except for the estimates of $SE(h)$ for Variance Component Structure A, which slightly overestimated the parameter values; and all of the estimates using polytomous data tended to be too low, except for the estimates of $SE(ph)$, $SE(pih)$, and $SE(\delta)$ for Variance Component Structure A. For the $i : p$ design using dichotomous data, the estimates of $SE(p)$, $SE(i : p)$, $SE(\delta)$, and $SE(\Delta)$ tended to be too low; and for the $i : p$ design using polytomous data, the estimates of $SE(p)$ tended to be too low, while the estimates of $SE(i : p)$, $SE(\delta)$, and $SE(\Delta)$ tended to be too high. For the $i : h : p$ design using dichotomous data, all of the standard error estimates tended to be too low, and for the $i : h : p$ design using polytomous data, all of the standard error estimates tended to be too low, except for the estimates of $SE(i : h : p)$, which tended to be too high. For the $p \times (i : h)$ design using dichotomous data, all of the standard error estimates tended to be too low; and for the $p \times (i : h)$ design using polytomous data, all of the standard error estimates tended to be too low, except for the estimates of $SE(pi : h)$, which tended to be too high.

In summary, when the data were dichotomous, the equations for estimating standard errors that are based on normality tended to produce estimates that were too low across all five G theory designs. When the data were polytomous, the normality-based equations for estimating standard errors produced estimates that tended to be too low for estimated standard errors of main effects; the estimates of some interaction effects and nested effects tended to be too high, but it is unclear whether there is a pattern to these effects. However, these results show that in many cases, when the score effects are not normally distributed, bootstrap estimates of standard errors of estimated variance components and relative and absolute error variances are indeed superior to the estimates based on equations that assume normality.

Summary of Results

This section summarizes the results of this study with respect to the research questions posed in Chapter 1. The first research question asked "How well do the rules work across various G theory designs, including the $p \times i$, $p \times i \times h$, $i : p$, $i : h : p$, and $p \times (i : h)$ designs?" This study examined five G theory designs, and found common patterns across the designs with respect to the facets entering into the variance component estimates for which the standard errors were being estimated. For example, rules were proposed for non-nested and nested main effects, non-nested and nested interactions, etc.

The second research question asked "How well do the rules work when the variance components for the various effects are relatively similar to one another (i.e., each variance component contributes a similar amount of variance to the total variance)? When they are relatively dissimilar (i.e., most of the total variance is contained within a smaller subset of variance components)?" This study included two variance component structures for each G theory design; one containing relatively similar magnitudes of variance across the various effects, and one containing relatively dissimilar magnitudes of variance across the various effects. It was found that, for the most part, the rules followed similar patterns across variance component structures, and that sample size patterns tended to have a larger impact on the extent to which the rules produced adequate estimates.

The third research question asked "How well do the rules work when the sample sizes for the various effects are similar to one another (i.e., the number of persons is similar to the number of items and/or raters)? When the sample sizes are dissimilar to one another?" This study found that sample size did indeed influence the extent to which the rules held up across the various combinations of conditions. In particular, for the $i:p$ design with normal or polytomous data, when the number of persons was larger than the number of items, the boot-*p* bootstrapping procedure outperformed the boot-*i* procedure for estimating $SE(i:p)$, $SE(\delta)$, and $SE(\Delta)$. However, when the $n_p = n_i$, both boot-*p* and boot-*i* produced reasonable estimates of $SE(i:p)$, $SE(\delta)$, and $SE(\Delta)$.

The fourth research question asked "How well does the workaround for the bootstrap estimate of the standard error of the absolute error variance [$SE(\Delta)$] proposed by Tong and Brennan (2007) hold up across various G theory designs, variance component structures, sample size patterns, and data types?" This study found that the workaround, on average, outperformed the bootstrap estimates of $SE(\Delta)$ across the various combinations of conditions studied.

The fifth research question asked "To what extent do standard normal, percentile, and bias corrected percentile confidence intervals adequately capture the variability of estimated variance components and relative and absolute error variances? Is one type of confidence interval preferred to the others?" It was found that the three confidence intervals performed similarly overall; however, some problems with the bias-corrected intervals emerged, particularly when calculating coverage of estimated variance components for which the sample size was small. Additionally, the known asymmetry of variance components calls into question the adequacy of the standard normal confidence intervals. For these reasons, percentile confidence intervals were determined to be preferable to standard normal and bias corrected percentile confidence intervals in this study.

CHAPTER 5: SUMMARY AND CONCLUSIONS

The purpose of this study was to test the extent to which rules proposed by Tong and Brennan (2007) for estimating standard errors of estimated variance components held up across a variety of G theory designs, variance component structures, sample size patterns, and data types. Several different bootstrap procedures were applied to each of the 60 combinations of conditions to calculate point estimates, standard errors, and three types of confidence intervals for each variance component and relative and absolute error variance.

The bootstrap procedures generally produced reasonably accurate estimates of variance components and relative and absolute error variances. Estimates of standard errors were reasonable, depending on the associated variance component estimate, and depending on the bootstrap procedure that was being implemented.

Evaluation of Tong and Brennan's (2007) Rules

This study provides support for some of the rules proposed by Tong and Brennan (2007) (Appendix D), but suggests refinements of others. Furthermore, some of the optimal bootstrap procedures were dependent on the data type, which complicates the rules somewhat.

Tong and Brennan's (2007) first rule was that for non-nested main effects, the bootstrap procedure corresponding to that facet should be used. For example, boot-*p* should be used to estimate $SE(p)$. This rule was largely supported in this study across the three types of data.

The second rule was that for nested main effects, the bootstrap procedure corresponding to the nested facet should be used. For example, boot-*i* should be used to estimate $SE(i:p)$. The results of this study generally did not support this rule, and appeared to be dependent on the sample sizes of the facets involved. This study found that, in general, the bootstrap procedure corresponding to the non-nested facet tended to produce better estimates of standard errors of nested main effects than did the nested facet. For example, boot-*p* tended to produce better estimates of $SE(i:p)$. However, when $n_p = n_i$, both boot-*p* and boot-*i* tended to produce reasonable estimates of $SE(i:p)$ for normal and polytomous data.

The third rule was that for non-nested interaction effects, the bootstrap procedure corresponding to the facet with the larger sample size should be used. For example, boot-*p* should be used to estimate $SE(pi)$ when $n_p > n_i$. The results of this study found that this rule was supported for normal and polytomous data, but the results were different for dichotomous data. Like Brennan, Harris, and Hanson (1987), this study found that for the $p \times i$ design with dichotomous data, boot-*p,i* produced better estimates of $SE(pi)$ than did boot-*p*. Additionally, for two-facet designs with dichotomous data, Tong and Brennan's (2007) third rule was supported for the two-way interactions (i.e., $SE(pi)$, $SE(ph)$, and $SE(ih)$), but was not supported for the highest order interactions. For example, boot-*p,i* tended to produce reasonable estimates of $SE(pi)$ for the $p \times i$ design and $SE(pih)$ for the $p \times i \times h$ design.

The fourth rule was that for nested interactions, the bootstrap procedure corresponding to the nested facet with the larger sample size should be used. The results of this study provided some support for this rule, dependent on the data type. For the $p \times (i:h)$ design, boot-*p* tended to produce the best estimates of $SE(pi:h)$ for normal and polytomous data. However, for dichotomous data, boot-*h* tended to produce better estimates of $SE(pi:h)$ than did boot-*p*.

The fifth rule was that for estimating the standard error of the estimated relative error variance, the bootstrap procedure corresponding to the objects of measurement facet should be used, which is usually the persons facet. This study provided support for this rule for normal and polytomous data; however, results were different for dichotomous data. For dichotomous data, boot-*p,i* tended to produce better estimates of $SE(\delta)$ for the $p \times i$ design, and boot-*p,h* tended to produce better estimates of $SE(\delta)$ for the $p \times i \times h$ and

$p \times (i:h)$ designs. For the $i:p$ and $i:h:p$ designs, boot-*p* produced the best estimates of *SE*(δ) across all three data types.

Tong and Brennan (2007) also proposed a workaround for estimating the standard error of the estimated absolute error variance. In general, the workaround performed reasonably well across the five G theory designs. The workaround was based on taking the optimal bootstrap estimates for each component that enters into *SE*(Δ), where the optimal estimates were based on their proposed rules. In this study it was found that the bootstrap procedures suggested by their rules did not always provide the optimal estimates. When the rules did not provide the best estimates, the workarounds for the standard error estimates of the absolute error variance estimates were modified to include the best estimates for each component entering into the estimate. These modified workarounds tended to produce better estimates for the $p \times i$, $p \times i \times h$, and $p \times (i:h)$ designs. The workaround for estimating *SE*(Δ) worked well for the $i:p$ and $i:h:p$ designs; however, because *SE*(δ) = *SE*(Δ) for these designs, the workaround is redundant to the estimates of *SE*(δ).

Because the current study also included a G theory design containing two levels of nesting (i.e., the $i:h:p$ design), this dissertation proposes an additional rule to deal with estimating standard errors of estimated variance components involving more than one level of nesting. Based on the results of this study, it is proposed that for these estimates, the bootstrap procedure corresponding to the non-nested facet be used. For example, in the $i:h:p$ design, boot-*p* should be used to estimate $SE(i:h:p)$.

<u>Suggested Modifications to Tong and Brennan's (2007)</u>

<u>Rules</u>

Based on the results of the current study, the following is a list of suggested rules for choosing which bootstrap procedures to use to estimate standard errors of estimated variance components and relative and absolute error variances. It should be noted that, although they are based on a more extensive set of conditions than those of Tong and Brennan (2007), they are still somewhat ad hoc in nature, and further research is needed to make definitive statements about the extent to which they produce reasonable estimates.

Rule 1: For nonnested main effects, use the corresponding bootstrap procedure. For example, to estimate $SE\left[\hat{\sigma}^2(p)\right]$, use boot–*p*.

Rule 2: For nested main effects (e.g., *i*:*h* in the *p* × (*i*:*h*) design), use the bootstrap procedure for the nonnested index. For example, to estimate $SE\left[\hat{\sigma}^2(i:h)\right]$ for the *p* ×(*i*:*h*) design, use boot–*h*.

Rule 3a: For nonnested interaction effects when the data are normal or polytomous, use the one-dimensional bootstrap procedure for the facet in the interaction that has the largest sample size. For example, to estimate $SE\left[\hat{\sigma}^2(ph)\right]$, use boot–*p* if $n_p > n_h$ or use boot–*h* if $n_p < n_h$.

Rule 3b: For nonnested interaction effects when the data are dichotomous, use the one-dimensional bootstrap procedure for the facet in the interaction that has the largest sample size for all but the highest order interaction. For example, to estimate $SE\left[\hat{\sigma}^2(ph)\right]$, use boot–*p* if $n_p > n_h$ or use boot–*h* if $n_p < n_h$. For the highest order nonnested interaction effect, use boot-*p,i*.

Rule 4a: For nested interaction effects when the data are normal or polytomous, use the one-dimensional bootstrap procedure for the primary facet in the interaction that has the largest sample size. For example, to estimate $SE\left[\hat{\sigma}^2(pi:h)\right]$, use boot–*p* if $n_p > n_i$ or use boot–*i* if $n_p <$ *n*i.

Rule 4b: For nested interaction effects when the data are dichotomous, use the bootstrap procedure for the nonnested index. For example, to estimate $SE\left[\hat{\sigma}^2(pi:h)\right]$ for the *p* ×(*i*:*h*) design, use boot–*h*.

Rule 5: For effects involving more than one level of nesting, use the bootstrap procedure corresponding to the non-nested facet. For example, in the $i:h:p$ design, use boot-*p* to estimate $SE(i:h:p)$

Rule 6a: For $SE(\delta)$ when the data are normal or polytomous, use the one-dimensional bootstrap procedure for the objects-of-measurement facet, typically boot-*p*.

Rule 6b: For $SE(\delta)$ when the data are dichotomous, use boot-*p,i* for the $p \times i$ design, and use boot-*p,h* for the $p \times i \times h$ and $p \times (i:h)$ designs.

Rule 7: For $SE(\Delta)$, use a combination of bootstrap procedures, such that the optimal bootstrap procedure suggested by the rules above is selected for each standard error component entering into the equation for estimating $SE(\Delta)$.

Confidence Interval Coverage

This study also investigated three types of confidence interval coverage: standard normal, percentile, and bias corrected percentile. The results of this study provided support for the use of confidence intervals as additional evidence of the variability of estimated variance components and relative and absolute error variances. However, for those variance components involving facets with small sample sizes, the confidence interval coverage tended to be inaccurate. Researchers who use confidence intervals in this context would want to keep in mind that as sample size decreases, the error associated with estimated variance components increases.

One of the goals of this dissertation was to investigate whether one of the three different types of confidence intervals would be preferable to the other two in terms of its utility in describing the variability of estimated variance components and relative and absolute error variances. It was found that for the most part, the three confidence intervals were comparable. However, the bias corrected percentile confidence interval coverage for some of the conditions were substantially lower than the other two types of confidence intervals, casting some doubt on the adequacy of the bias corrected percentile confidence intervals for use in this context. Additionally, because in many cases the data were not normally distributed, the standard normal confidence intervals were also suspect; therefore, the percentile confidence intervals were chosen as the preferable confidence intervals for use in this context.

Hypothetical Example of Use of Standard Error Estimates

It is important to have information regarding the accuracy of estimates of variance components. Therefore, this section presents a hypothetical example illustrating the use of standard error estimates, as well as showing some of the possible consequences resulting from not taking the standard errors into account.

This example is based on the results found for the $p \times i \times h$ design, Variance Component Structure B, Sample Size Pattern 2, with polytomous data. For this example, consider the following scenario: researchers are developing an exam consisting of constructed-response questions, and are interested in determining the optimal numbers of items and raters that should be used to obtain a generalizability coefficient of at least 0.80. A generalizability coefficient ($\mathrm{E}\rho^2$) is similar to a reliability coefficient in Classical Test Theory, and can be calculated from the following equation:

$$\mathrm{E}\rho^2 = \frac{\sigma^2(p)}{\sigma^2(p)+\sigma^2(\delta)} \tag{5.1}$$

To accomplish their goal, the researchers created an exam consisting of 10 constructed-response questions, and it was administered to 100 students. Each question was rated by three raters.

The results presented in Appendix F, Table F6 can be used to create a D study to answer the researchers' question. This D study would be for the $p \times I \times H$ design, where items and raters are averaged across numbers of conditions specified by the researchers. By varying the number of items and raters that enter into the equation to estimate the relative error variance (Equation A.152 in Appendix A, Table A28), hypothetical

$E\rho^2$ estimates can be calculated to determine the optimal number of items and raters to use to obtain the desired reliability. Table 12 contains the results of this hypothetical D study.

Table 12. Generalizability Coefficients for Hypothetical D Study

	Number of Raters			
Number of Items	2	3	4	5
10	0.72	0.79	0.82	0.84
15	0.75	0.81	0.84	0.87
20	0.76	0.82	0.85	0.88
25	0.77	0.83	0.86	0.88
30	0.77	0.83	0.87	0.89

As can be seen in Table 12, the researchers could decide to use 15 items and 3 raters for their test to obtain a reliability estimate of at least 0.80. However, these results are estimates, and do not take into account the error associated with the estimates. Table G6 in Appendix G contains the estimated standard errors for this hypothetical example. If the standard errors are taken into account by creating intervals, in this case by taking the estimate plus or minus 1.645 standard errors, the upper and lower bounds can be used to estimate reasonable ranges (analogous to 90% confidence bands) for the generalizability coefficients produced by the D study. Table 13 contains the resulting ranges of generalizability coefficients produced when the standard error is taken into account.

As can be seen in Table 13, if the standard errors are taken into account, the researchers might conclude that if they chose 15 items and three raters for their test instrument, there is a possibility that the resulting reliability may only be 0.78. They might decide that it would be preferable to use 25 items and three raters, or 15 items and four raters, depending on the cost of increasing items and raters, respectively, to increase their confidence that the resulting reliability is at least as high as 0.80. The results presented in Table 13 provide more information regarding the variability of the estimated generalizability coefficients, and therefore should lead the researchers to make more informed decisions about their test instrument.

Table 13. Generalizability Coefficient Ranges for Hypothetical D Study

	Number of Raters			
Number of Items	2	3	4	5
10	0.68 - 0.74	0.75 - 0.80	0.79 - 0.84	0.81 - 0.86
15	0.71 - 0.76	0.78 - 0.82	0.82 - 0.86	0.84 - 0.88
20	0.73 - 0.78	0.79 - 0.83	0.83 - 0.87	0.85 - 0.89
25	0.74 - 0.78	0.80 - 0.84	0.84 - 0.87	0.86 - 0.89
30	0.74 - 0.79	0.81 - 0.84	0.85 - 0.88	0.87 - 0.90

Limitations and Future Research

One limitation of this study is that only certain G theory designs, variance component structures, and sample size patterns were considered. It does, however, provide a more comprehensive set of results when partnered with the previous research that has been done on this topic, because it is an extension of Tong and Brennan (2007), using different variance component structures and sample size patterns than those used in their study. Additionally, the methods used to estimate standard errors in this study are not exhaustive. Obviously, other procedures may yield different results.

The reliance on simulated data is another limitation of this study. Simulated data are by definition contrived, and the extent to which the results will apply in a situation where data are collected from actual examinees administered actual test items is unknown. Real data can have unpredictable properties sometimes. Further research should be done using empirical data. Further research should also investigate bias corrected and accelerated confidence intervals, which were beyond the scope of this study due to the complexity of calculating them and the large number of replications and bootstrap samples within replications.

The methods that were used to generate the simulated data are also potential limitations of this study. Dichotomous data were generated by dichotomizing data with normally distributed score effects. It is probably not realistic to assume that the underlying distributions of the score effects of dichotomous data are normal. Additionally, the method used to generate polytomous data produced data that are not strictly polytomous; meaning that the simulated data do not consist of finite integers, but are non-normal data with a finite range. Therefore, the nature of the simulated data is a potential shortcoming of this study.

Other ways of obtaining simulated dichotomous and polytomous data would be worthwhile to investigate as well to make sure that the results are not dependent on how the data were derived. Because of how the dichotomous data were derived (i.e., dichotomizing normal data), it is unclear the extent to which they represent empirical dichotomous data. It may or may not be a reasonable assumption that dichotomous data has an underlying normal distribution. However, it should be noted that the dichotomous data in this study were generated in a different fashion than the dichotomous data generated in Brennan, Harris, and Hanson’s (1987) study, and both studies found similar results with respect to which bootstrap procedures produced the optimal standard error estimates.

Another limitation is that this study was confined to univariate G theory. It is not certain whether these results would also apply to multivariate G theory. Multivariate G theory is more complex than univariate G theory because each object of measurement has more than one universe score (the *universe score* is the expected value of a person’s mean score for every combination of measurement procedures). For example, consider a student who takes a battery of tests. Instead of having a single universe score, the student has a *profile* of universe scores (e.g., scores for math, English, and science), and there are covariances between the various pairs of universe scores, which makes the issue much more complicated (Brennan, 2001). However, it is important to establish procedures that work well for univariate G theory, because if they do not work well for univariate G theory, it is unlikely that they will be adequate for more complicated multivariate G theory designs.

Further research should be done with respect to nesting within G theory designs. In this study, although a variety of sample sizes were utilized, there were no cases in which $n_p < n_i$ in the $i:p$ design or $n_p < n_i$ or $n_p < n_h$ in the $i:h:p$ design. It may be that when the sample size of the nested facet is larger than the sample size of the non-nested facet(s), the bootstrap procedure corresponding to the nested facet may be preferable to the bootstrap procedure in which the non-nested facet is resampled.

A further limitation of this study is that it focused solely on bootstrap resampling rather than also exploring jackknife resampling. The jackknife involves systematically deleting one observation from a dataset and estimating the variance components for each resulting subset of the data, or jackknife sample. The standard deviations of the estimated variance components across jackknife samples are the estimated standard errors of the estimated variance components. Brennan (2001) and Feng (2002) have suggested reasons to prefer the jackknife over the bootstrap, including unbiased estimates of variance components. However, the amount of computation required for jackknife resampling makes it an impractical approach in this context, whereas the bootstrap approach is not ideal but is more practical. It would be a worthwhile endeavor for future research to investigate the utility of delete-*d* jackknife procedures for estimating standard errors of estimated variance components. Delete-*d* jackknife procedures involve systematically deleting *n* observations rather than a single observation at a time, resulting in fewer calculations required. However, there are not equations worked out yet for using this approach in a generalizability theory context, so further research needs to be done on this topic.

It might also be worthwhile to investigate whether parametric bootstrap procedures would provide adequate estimates of variance components and their estimated standard errors. However, parametric bootstrap procedures require assumptions about the distributions of the score effects, which is why the bootstrap procedures have been proposed in the first place, because these distributions are often unknown. For parametric bootstrap procedures, the researcher must assign an assumed distribution for each of the score effects, then sample from those distributions. If the data are changed to make them dichotomous or polytomous, the sampling is destroyed, so it is unclear as to what the best way to do this would be.

This study included coverage for three types of confidence intervals, standard normal, percentile, and bias corrected percentile. However, because of the nature of the distributions of variance components, none of these three types of confidence intervals may be very accurate. A possible better alternative may be bias corrected and accelerated confidence intervals, which are very complicated; and due to the large amount of computation required, were beyond the scope of this study. Further research, however, might investigate the performance of these confidence intervals, perhaps in empirical rather than simulated data.

Implications

As stated in Chapter 1 of this dissertation, oftentimes the data that are analyzed via G studies and D studies may violate the assumption of normality of score effects, meaning that the estimated standard errors may be inaccurate. This dissertation attempted to further explore the utility of using bootstrap procedures to more accurately estimate standard errors of estimated variance components and relative and absolute error variances. It is important that researchers are aware of the variability associated with results of their studies, so that the conclusions they draw from their research are appropriate.

Several articles that were cited in Chapter 2 of this dissertation presented estimated standard errors which were based on GENOVA output or Searle's (1971) or Brennan's (2001) equations, which assume multivariate normality. When data are not normally distributed, as was the case for at least some of these articles, the reported standard errors are likely to be inaccurate. Therefore, the authors should not have confidence in them, and decisions based on them are suspect. The authors of these studies need to be cautious when making statements about, for example, the number of items needed to obtain a certain level of reliability, because the reliability coefficients obtained from a D study are estimates.

The results of this study, when partnered with previous studies using bootstrap procedures to estimate standard errors, suggest that the bootstrap approach may hold promise in providing more accurate standard error estimates. However, more research is needed before definitive statements can be made regarding the optimal procedures for estimating the standard errors. A long term goal of this line of research is to establish these rules more firmly, and ultimately create software so that researchers have better tools with which to conduct G studies and D studies.

APPENDIX A: EQUATIONS

Table A1. Linear Models of Observed Scores for the G Theory Designs in This Study

Design	Linear model of observed scores X	
$p \times i$	$X_{pi} = \mu + v_p + v_i + v_{pi}$	(A.1)
$p \times i \times h$	$X_{pih} = \mu + v_p + v_i + v_h + v_{pi} + v_{ph} + v_{ih} + v_{pih}$	(A.2)
$i : p$	$X_{i:p} = \mu + v_p + v_{i:p}$	(A.3)
$i : h : p$	$X_{i:h:p} = \mu + v_p + v_{h:p} + v_{i:h:p}$	(A.4)
$p \times (i : h)$	$X_{pi:h} = \mu + v_p + v_h + v_{ph} + v_{i:h} + v_{pi:h}$	(A.5)

Table A2. Score Effects for the $p \times i$ Design

Effect (α)	Score effect v_α	
p	$\mu_p - \mu$	(A.6)
i	$\mu_i - \mu$	(A.7)
pi	$X_{pi} - \mu_p - \mu_i + \mu$	(A.8)

Table A3. Score Effects for the $p \times i \times h$ Design

Effect (α)	Score effect v_α	
p	$\mu_p - \mu$	(A.9)
i	$\mu_i - \mu$	(A.10)
h	$\mu_h - \mu$	(A.11)
pi	$\mu_{pi} - \mu_p - \mu_i + \mu$	(A.12)
ph	$\mu_{ph} - \mu_p - \mu_h + \mu$	(A.13)
ih	$\mu_{ih} - \mu_i - \mu_h + \mu$	(A.14)
pih	$X_{pih} - \mu_{pi} - \mu_{ph} - \mu_{ih} + \mu_p + \mu_i + \mu_h - \mu$	(A.15)

Table A4. Score Effects for the $i : p$ Design

Effect (α)	Score effect v_α	
p	$\mu_p - \mu$	(A.16)
$i : p$	$X_{pi} - \mu_p$	(A.17)

Table A5. Score Effects for the $i : h : p$ Design

Effect (α)	Score effect v_α	
p	$\mu_p - \mu$	(A.18)
$h : p$	$\mu_{ph} - \mu_p$	(A.19)
$i : h : p$	$X_{pih} - \mu_{ph}$	(A.20)

Table A6. Score Effects for the $p \times (i : h)$ Design

Effect (α)	Score effect v_α	
p	$\mu_p - \mu$	(A.21)
h	$\mu_h - \mu$	(A.22)
ph	$\mu_{ph} - \mu_p - \mu_h + \mu$	(A.23)
$i{:}h$	$\mu_{ih} - \mu_h$	(A.24)
$pi{:}h$	$X_{pih} - \mu_{ph} - \mu_{ih} + \mu_h$	(A.25)

Table A7. Variance of Observed Scores Decomposed into Variance Components

Design	Decomposition of observed score into variance components	
$p \times i$	$\sigma^2(X_{pi}) = \sigma^2(p) + \sigma^2(i) + \sigma^2(pi)$	(A.26)
$p \times i \times h$	$\sigma^2(X_{pih}) = \sigma^2(p) + \sigma^2(i) + \sigma^2(h) + \sigma^2(pi) + \sigma^2(ph) + \sigma^2(ih) + \sigma^2(pih)$	(A.27)
$i : p$	$\sigma^2(X_{i:p}) = \sigma^2(p) + \sigma^2(i{:}p)$	(A.28)
$i : h : p$	$\sigma^2(X_{i:h:p}) = \sigma^2(p) + \sigma^2(h{:}p) + \sigma^2(i{:}h{:}p)$	(A.29)
$p \times (i : h)$	$\sigma^2(X_{pi:h}) = \sigma^2(p) + \sigma^2(h) + \sigma^2(ph) + \sigma^2(i{:}h) + \sigma^2(pi{:}h)$	(A.30)

Table A8. Definitions of Variance Components for the $p \times i$ Design

Effect (α)	Variance component definitions $\sigma^2(\alpha)$	
p	$\underset{p}{\mathrm{E}}(\mu_p - \mu)^2 = \underset{p}{\mathrm{E}}(\nu_p^2)$	(A.31)
i	$\underset{i}{\mathrm{E}}(\mu_i - \mu)^2 = \underset{i}{\mathrm{E}}(\nu_i^2)$	(A.32)
pi	$\underset{p}{\mathrm{E}}\underset{i}{\mathrm{E}}(\mu_{pi} - \mu_p - \mu_i + \mu)^2 = \underset{p}{\mathrm{E}}\underset{i}{\mathrm{E}}(\nu_{pi}^2)$	(A.33)

Table A9. Definitions of Variance Components for the $p \times i \times h$ Design

Effect (α)	Variance component definitions $\sigma^2(\alpha)$	
p	$\underset{p}{\mathrm{E}}(\mu_p - \mu)^2 = \underset{p}{\mathrm{E}}(\nu_p^2)$	(A.34)
i	$\underset{i}{\mathrm{E}}(\mu_i - \mu)^2 = \underset{i}{\mathrm{E}}(\nu_i^2)$	(A.35)
h	$\underset{h}{\mathrm{E}}(\mu_h - \mu)^2 = \underset{h}{\mathrm{E}}(\nu_h^2)$	(A.36)
pi	$\underset{p}{\mathrm{E}}\underset{i}{\mathrm{E}}(\mu_{pi} - \mu_p - \mu_i + \mu)^2 = \underset{p}{\mathrm{E}}\underset{i}{\mathrm{E}}(\nu_{pi}^2)$	(A.37)
ph	$\underset{p}{\mathrm{E}}\underset{h}{\mathrm{E}}(\mu_{ph} - \mu_p - \mu_h + \mu)^2 = \underset{p}{\mathrm{E}}\underset{h}{\mathrm{E}}(\nu_{ph}^2)$	(A.38)
ih	$\underset{i}{\mathrm{E}}\underset{h}{\mathrm{E}}(\mu_{ih} - \mu_i - \mu_h + \mu)^2 = \underset{i}{\mathrm{E}}\underset{h}{\mathrm{E}}(\nu_{ih}^2)$	(A.39)
pih	$\underset{p}{\mathrm{E}}\underset{i}{\mathrm{E}}\underset{h}{\mathrm{E}}(X_{pih} - \mu_{pi} - \mu_{ph} - \mu_{ih} + \mu_p + \mu_i + \mu_h - \mu)^2 = \underset{p}{\mathrm{E}}\underset{i}{\mathrm{E}}\underset{h}{\mathrm{E}}(\nu_{pih}^2)$	(A.40)

Table A10. Definitions of Variance Components for the $i:p$ Design

Effect (α)	Variance component definitions $\sigma^2(\alpha)$	
p	$\underset{p}{\mathrm{E}}(\mu_p - \mu)^2 = \underset{p}{\mathrm{E}}(\nu_p^2)$	(A.41)
$i:p$	$\underset{i}{\mathrm{E}}(X_{pi} - \mu_p)^2 = \underset{i}{\mathrm{E}}(\nu_{i:p}^2)$	(A.42)

Table A11. Definitions of Variance Components for the $i:h:p$ Design

Effect (α)	Variance component definitions $\sigma^2(\alpha)$	
p	$\underset{p}{\mathrm{E}}(\mu_p - \mu)^2 = \underset{p}{\mathrm{E}}(\nu_p^2)$	(A.43)
$h:p$	$\underset{i}{\mathrm{E}}(\mu_{ph} - \mu_p)^2 = \underset{i}{\mathrm{E}}(\nu_{h:p}^2)$	(A.44)
$i:h:p$	$\underset{i}{\mathrm{E}}(X_{pih} - \mu_{ph})^2 = \underset{i}{\mathrm{E}}(\nu_{i:h:p}^2)$	(A.45)

Table A12. Definitions of Variance Components for the $p \times (i:h)$ Design

Effect (α)	Variance component definitions $\sigma^2(\alpha)$	
p	$\underset{p}{\mathrm{E}}(\mu_p - \mu)^2 = \underset{p}{\mathrm{E}}(\nu_p^2)$	(A.46)
h	$\underset{h}{\mathrm{E}}(\mu_h - \mu)^2 = \underset{h}{\mathrm{E}}(\nu_h^2)$	(A.47)
ph	$\underset{p}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(\mu_{ph} - \mu_p - \mu_h + \mu)^2 = \underset{p}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(\nu_{ph}^2)$	(A.48)
$i{:}h$	$\underset{i}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(\mu_{ih} - \mu_h)^2 = \underset{i}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(\nu_{i:h}^2)$	(A.49)
$pi{:}h$	$\underset{p}{\mathrm{E}}\,\underset{i}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(X_{pih} - \mu_{ph} - \mu_{ih} + \mu_h)^2 = \underset{p}{\mathrm{E}}\,\underset{i}{\mathrm{E}}\,\underset{h}{\mathrm{E}}(\nu_{pi:h}^2)$	(A.50)

Table A13. Estimators of Variance Components for the $p \times i$ Design

Effect (α)	Estimator of variance component $\sigma^2(\alpha)$	
p	$[MS(p) - MS(pi)] / n_i$	(A.51)
i	$[MS(i) - MS(pi)] / n_p$	(A.52)
pi	$MS(pi)$	(A.53)

Table A14. Estimators of Variance Components for the $p \times i \times h$ Design

Effect (α)	Estimator of variance component $\sigma^2(\alpha)$	
p	$[MS(p) - MS(pi) - MS(ph) + MS(pih)] / n_i n_h$	(A.54)
i	$[MS(i) - MS(pi) - MS(ih) + MS(pih)] / n_p n_h$	(A.55)
h	$[MS(h) - MS(ph) - MS(ih) + MS(pih)] / n_p n_i$	(A.56)
pi	$[MS(pi) - MS(pih)] / n_h$	(A.57)
ph	$[MS(ph) - MS(pih)] / n_i$	(A.58)
ih	$[MS(ih) - MS(pih)] / n_p$	(A.59)
pih	$MS(pih)$	(A.60)

Table A15. Estimators of Variance Components for the $i : p$ Design

Effect (α)	Estimator of variance component $\sigma^2(\alpha)$	
p	$[MS(p) - MS(i{:}p)] / n_i$	(A.61)
$i{:}p$	$MS(i{:}p)$	(A.62)

Table A16. Estimators of Variance Components for the $i : h : p$ Design

Effect (α)	Estimator of variance component $\sigma^2(\alpha)$	
p	$[MS(p) - MS(h{:}p)] / n_i n_h$	(A.63)
$h{:}p$	$[MS(h{:}p) - MS(i{:}h{:}p)] / n_i$	(A.64)
$i{:}h{:}p$	$MS(i{:}h{:}p)$	(A.65)

Table A17. Estimators of Variance Components for the $p \times (i:h)$ Design

Effect (α)	Estimator of variance component $\sigma^2(\alpha)$	
p	$[MS(p) - MS(ph)] / n_i n_h$	(A.66)
h	$[MS(h) - MS(i{:}h) - MS(ph) + MS(pi{:}h)] / n_p n_i$	(A.67)
ph	$[MS(ph) - MS(pi{:}h)] / n_i$	(A.68)
$i{:}h$	$[MS(i{:}h) - MS(pi{:}h)] / n_p$	(A.69)
$pi{:}h$	$MS(pi{:}h)$	(A.70)

Table A18. *T* Terms, Sums of Squares, and Degrees of Freedom for the $p \times i$ Design

Effect α	*T* terms $T(\alpha)$	Sums of squares $SS(\alpha)$	Degrees of freedom $df(\alpha)$	
p	$n_i \sum \bar{X}_p^2$	$T(p) - T(\mu)$	$n_p - 1$	(A.71-73)
i	$n_p \sum \bar{X}_i^2$	$T(i) - T(\mu)$	$n_i - 1$	(A.74-76)
pi	$\sum \sum X_{pi}^2$	$T(pi) - T(p) - T(i) + T(\mu)$	$(n_p - 1)(n_i - 1)$	(A.77-79)

Table A19. *T* Terms, Sums of Squares, and Degrees of Freedom for the $p \times i \times h$ Design

Effect α	*T* terms $T(\alpha)$	Sums of squares $SS(\alpha)$	Degrees of freedom $df(\alpha)$	
p	$n_i n_h \sum \bar{X}_p^2$	$T(p) - T(\mu)$	$n_p - 1$	(A.80-82)
i	$n_p n_h \sum \bar{X}_i^2$	$T(i) - T(\mu)$	$n_i - 1$	(A.83-85)
h	$n_p n_i \sum \bar{X}_h^2$	$T(h) - T(\mu)$	$n_h - 1$	(A.86-88)
pi	$n_h \sum \sum \bar{X}_{pi}^2$	$T(pi) - T(p) - T(i) + T(\mu)$	$(n_p - 1)(n_i - 1)$	(A.89-91)
ph	$n_i \sum \sum \bar{X}_{ph}^2$	$T(ph) - T(p) - T(h) + T(\mu)$	$(n_p - 1)(n_h - 1)$	(A.92-94)
ih	$n_p \sum \sum \bar{X}_{ih}^2$	$T(ih) - T(i) - T(h) + T(\mu)$	$(n_i - 1)(n_h - 1)$	(A.95-97)
pih	$\sum \sum \sum X_{pih}^2$	$T(pih) - T(pi) - T(ph) - T(ih) + T(p) + T(i) + T(h) - T(\mu)$	$(n_p - 1)(n_i - 1)(n_h - 1)$	(A.98-100)

Table A20. *T* Terms, Sums of Squares, and Degrees of Freedom for the $i : p$ Design

Effect α	*T* terms $T(\alpha)$	Sums of squares $SS(\alpha)$	Degrees of freedom $df\,(\alpha)$	
p	$n_i \sum \bar{X}_p^2$	$T(p) - T(\mu)$	$n_p - 1$	(A.101-103)
$i{:}p$	$\sum \sum X_{pi}^2$	$T(i{:}p) - T(p)$	$n_p\,(n_i - 1)$	(A.104-106)

Table A21. *T* Terms, Sums of Squares, and Degrees of Freedom for the $i : h : p$ Design

Effect α	*T* terms $T(\alpha)$	Sums of squares $SS(\alpha)$	Degrees of freedom $df\,(\alpha)$	
p	$n_i n_h \sum \bar{X}_p^2$	$T(p) - T(\mu)$	$n_p - 1$	(A.107-109)
$h{:}p$	$n_i \sum \sum \bar{X}_{h:p}^2$	$T(h{:}p) - T(p)$	$n_p(n_h - 1)$	(A.110-112)
$i{:}h{:}p$	$\sum \sum \sum X_{pih}^2$	$T(i{:}h{:}p) - T(h{:}p)$	$n_p n_h(n_i - 1)$	(A.113-115)

Table A22. *T* Terms, Sums of Squares, and Degrees of Freedom for the $p \times (i : h)$ Design

Effect α	*T* terms $T(\alpha)$	Sums of squares $SS(\alpha)$	Degrees of freedom $df\,(\alpha)$	
p	$n_i n_h \sum \bar{X}_p^2$	$T(p) - T(\mu)$	$n_p - 1$	(A.116-118)
h	$n_p n_i \sum \bar{X}_h^2$	$T(h) - T(\mu)$	$n_h - 1$	(A.119-121)
ph	$n_i \sum \sum \bar{X}_{ph}^2$	$T(ph) - T(p) - T(h) + T(\mu)$	$(n_p - 1)(n_h - 1)$	(A.122-124)
$i{:}h$	$n_p \sum \sum \bar{X}_{i:h}^2$	$T(i{:}h) - \mathrm{T}(h)$	$n_h(n_i - 1)$	(A.125-127)
$pi{:}h$	$\sum \sum \sum X_{pih}^2$	$T(pi{:}h) - T(ph) - T(i{:}h) + T(h)$	$n_h(n_p - 1)(n_i - 1)$	(A.128-130)

Table A23. Estimated Standard Errors of Estimated Variance Components for the $p \times i$ Design under Normality

Effect (α)	Estimated standard error of $\hat{\sigma}^2(\alpha)$, $S\hat{E}(\alpha)$	
p	$\frac{1}{n_i}\sqrt{\frac{2[MS(p)]^2}{(n_p-1)+2}+\frac{2[MS(pi)]^2}{(n_p-1)(n_i-1)+2}}$	(A.131)
i	$\frac{1}{n_p}\sqrt{\frac{2[MS(i)]^2}{(n_i-1)+2}+\frac{2[MS(pi)]^2}{(n_p-1)(n_i-1)+2}}$	(A.132)
pi	$\sqrt{\frac{2[MS(pi)]^2}{(n_p-1)(n_i-1)+2}} = MS(pi)\sqrt{\frac{2}{(n_p-1)(n_i-1)+2}}$	(A.133)

Table A24. Standard Errors of Estimated Variance Components for the $p \times i \times h$ Design under Normality

Effect (α)	Estimated standard error of $\hat{\sigma}^2(\alpha)$, $S\hat{E}(\alpha)$	
p	$\frac{1}{n_i n_h}\sqrt{\frac{2[MS(p)]^2}{(n_p-1)+2}+\frac{2[MS(pi)]^2}{(n_p-1)(n_i-1)+2}+\frac{2[MS(ph)]^2}{(n_p-1)(n_h-1)+2}+\frac{2[MS(pih)]^2}{(n_p-1)(n_i-1)(n_h-1)+2}}$	(A.134)
i	$\frac{1}{n_p n_h}\sqrt{\frac{2[MS(i)]^2}{(n_i-1)+2}+\frac{2[MS(pi)]^2}{(n_p-1)(n_i-1)+2}+\frac{2[MS(ih)]^2}{(n_i-1)(n_h-1)+2}+\frac{2[MS(pih)]^2}{(n_p-1)(n_i-1)(n_h-1)+2}}$	(A.135)
h	$\frac{1}{n_p n_i}\sqrt{\frac{2[MS(h)]^2}{(n_h-1)+2}+\frac{2[MS(ph)]^2}{(n_p-1)(n_h-1)+2}+\frac{2[MS(ih)]^2}{(n_i-1)(n_h-1)+2}+\frac{2[MS(pih)]^2}{(n_p-1)(n_i-1)(n_h-1)+2}}$	(A.136)
pi	$\frac{1}{n_h}\sqrt{\frac{2[MS(pi)]^2}{(n_p-1)(n_i-1)+2}+\frac{2[MS(pih)]^2}{(n_p-1)(n_i-1)(n_h-1)+2}}$	(A.137)
ph	$\frac{1}{n_i}\sqrt{\frac{2[MS(ph)]^2}{(n_p-1)(n_h-1)+2}+\frac{2[MS(pih)]^2}{(n_p-1)(n_i-1)(n_h-1)+2}}$	(A.138)
ih	$\frac{1}{n_p}\sqrt{\frac{2[MS(ih)]^2}{(n_i-1)(n_h-1)+2}+\frac{2[MS(pih)]^2}{(n_p-1)(n_i-1)(n_h-1)+2}}$	(A.139)
pih	$\sqrt{\frac{2[MS(pih)]^2}{(n_p-1)(n_i-1)(n_h-1)+2}} = MS(pih)\sqrt{\frac{2}{(n_p-1)(n_i-1)(n_h-1)+2}}$	(A.140)

Table A25. Estimated Standard Errors of Estimated Variance Components for the $i:p$ Design under Normality

Effect (α)	Estimated standard error of $\hat{\sigma}^2(\alpha)$, $S\hat{E}(\alpha)$	
p	$\frac{1}{n_i}\sqrt{\frac{2[MS(p)]^2}{(n_p-1)+2}+\frac{2[MS(i:p)]^2}{n_p(n_i-1)+2}}$	(A.141)
i:p	$\sqrt{\frac{2[MS(i:p)]^2}{n_p(n_i-1)+2}} = MS(i:p)\sqrt{\frac{2}{n_p(n_i-1)+2}}$	(A.142)

Table A26. Estimated Standard Errors of Estimated Variance Components for the $i:h:p$ Design under Normality

Effect (α)	Estimated standard error of $\hat{\sigma}^2(\alpha)$, $S\hat{E}(\alpha)$	
p	$\frac{1}{n_i n_h}\sqrt{\frac{2[MS(p)]^2}{(n_p-1)+2}+\frac{2[MS(h:p)]^2}{n_p(n_h-1)+2}}$	(A.143)
h:p	$\frac{1}{n_i}\sqrt{\frac{2[MS(h:p)]^2}{n_p(n_h-1)+2}+\frac{2[MS(i:h:p)]^2}{n_p n_h(n_i-1)+2}}$	(A.144)
i:h:p	$\sqrt{\frac{2[MS(i:h:p)]^2}{n_p n_h(n_i-1)+2}} = MS(i:h:p)\sqrt{\frac{2}{n_p n_h(n_i-1)+2}}$	(A.145)

Table A27. Estimated Standard Errors of Estimated Variance Components for the $p \times (i:h)$ Design under Normality

Effect (α)	Estimated standard error of $\hat{\sigma}^2(\alpha)$, $S\hat{E}(\alpha)$	
p	$\dfrac{1}{n_i n_h}\sqrt{\dfrac{2[MS(p)]^2}{(n_p-1)+2}+\dfrac{2[MS(ph)]^2}{(n_p-1)(n_h-1)+2}}$	(A.146)
h	$\dfrac{1}{n_p n_i}\sqrt{\dfrac{2[MS(h)]^2}{(n_h-1)+2}+\dfrac{2[MS(ph)]^2}{(n_p-1)(n_h-1)+2}+\dfrac{2[MS(i:h)]^2}{n_h(n_i-1)+2}+\dfrac{2[MS(pi:h)]^2}{n_h(n_p-1)(n_i-1)+2}}$	(A.147)
ph	$\dfrac{1}{n_i}\sqrt{\dfrac{2[MS(ph)]^2}{(n_p-1)(n_h-1)+2}+\dfrac{2[MS(pi:h)]^2}{n_h(n_p-1)(n_i-1)+2}}$	(A.148)
i:h	$\dfrac{1}{n_p}\sqrt{\dfrac{2[MS(i:h)]^2}{n_h(n_i-1)+2}+\dfrac{2[MS(pi:h)]^2}{n_h(n_p-1)(n_i-1)+2}}$	(A.149)
pi:h	$\sqrt{\dfrac{2[MS(pi:h)]^2}{n_h(n_p-1)(n_i-1)+2}} = MS(pi:h)\sqrt{\dfrac{2}{n_h(n_p-1)(n_i-1)+2}}$	(A.150)

Table A28. Relative Error Variances

Design	Relative error variance $\sigma^2(\delta)$	
$p \times i$	$\dfrac{\sigma^2(pi)}{n_i}$	(A.151)
$p \times i \times h$	$\dfrac{\sigma^2(pi)}{n_i} + \dfrac{\sigma^2(ph)}{n_h} + \dfrac{\sigma^2(pih)}{n_i n_h}$	(A.152)
$i : p$	$\dfrac{\sigma^2(i:p)}{n_i}$	(A.153)
$i : h : p$	$\dfrac{\sigma^2(h:p)}{n_h} + \dfrac{\sigma^2(i:h:p)}{n_i n_h}$	(A.154)
$p \times (i : h)$	$\dfrac{\sigma^2(ph)}{n_h} + \dfrac{\sigma^2(pi:h)}{n_i n_h}$	(A.155)

Table A29. Absolute Error Variances

Design	Absolute error variance $\sigma^2(\Delta)$	
$p \times i$	$\sigma^2(\delta) + \dfrac{\sigma^2(i)}{n_i}$	(A.156)
$p \times i \times h$	$\sigma^2(\delta) + \dfrac{\sigma^2(i)}{n_i} + \dfrac{\sigma^2(h)}{n_h} + \dfrac{\sigma^2(ih)}{n_i n_h}$	(A.157)
$i : p$	$\sigma^2(\delta)$	(A.158)
$i : h : p$	$\sigma^2(\delta)$	(A.159)
$p \times (i : h)$	$\sigma^2(\delta) + \dfrac{\sigma^2(h)}{n_h} + \dfrac{\sigma^2(i:h)}{n_i n_h}$	(A.160)

Table A30. Standard Errors of Relative Error Variances

Design	Standard error of relative error variance $SE(\delta)$	
$p \times i$	$\frac{1}{n_i}\sqrt{\frac{2[\mathrm{E}MS(pi)]^2}{(n_p-1)(n_i-1)}}$	(A.161)
$p \times i \times h$	$\frac{1}{n_i n_h}\sqrt{\frac{2[\mathrm{E}MS(pi)]^2}{(n_p-1)(n_i-1)}+\frac{2[\mathrm{E}MS(ph)]^2}{(n_p-1)(n_h-1)}+\frac{2[\mathrm{E}MS(pih)]^2}{(n_p-1)(n_i-1)(n_h-1)}}$	(A.162)
$i:p$	$\frac{1}{n_i}\sqrt{\frac{2[\mathrm{E}MS(i:p)]^2}{n_p(n_i-1)}}$	(A.163)
$i:h:p$	$\frac{1}{n_i n_h}\sqrt{\frac{2[\mathrm{E}MS(h:p)]^2}{n_p(n_h-1)}}$	(A.164)
$p \times (i:h)$	$\frac{1}{n_i n_h}\sqrt{\frac{2[\mathrm{E}MS(ph)]^2}{(n_p-1)(n_h-1)}}$	(A.165)

Table A31. Standard Errors of Absolute Error Variances

Design	Standard error of absolute error variance $SE(\Delta)$	
$p \times i$	$\frac{1}{n_p n_i}\sqrt{\frac{2[\mathrm{E}MS(i)]^2}{(n_i-1)}+\frac{2[(n_p-1)\mathrm{E}MS(pi)]^2}{(n_p-1)(n_i-1)}}$	(A.166)
$p \times i \times h$	$\frac{1}{n_p n_i n_h}\sqrt{\frac{2[\mathrm{E}MS(i)]^2}{(n_i-1)}+\frac{2[\mathrm{E}MS(h)]^2}{(n_h-1)}+\frac{2[(n_p-1)\mathrm{E}MS(pi)]^2}{(n_p-1)(n_i-1)}+\frac{2[(n_p-1)\mathrm{E}MS(ph)]^2}{(n_p-1)(n_h-1)}+\frac{2[\mathrm{E}MS(ih)]^2}{(n_i-1)(n_h-1)}+\frac{2[(n_p-1)\mathrm{E}MS(pih)]^2}{(n_p-1)(n_i-1)(n_h-1)}}$	(A.167)
$i:p$	$\frac{1}{n_i}\sqrt{\frac{2[\mathrm{E}MS(i:p)]^2}{n_p(n_i-1)}}$	(A.168)
$i:h:p$	$\frac{1}{n_i n_h}\sqrt{\frac{2[\mathrm{E}MS(h:p)]^2}{n_p(n_h-1)}}$	(A.169)
$p \times (i:h)$	$\frac{1}{n_p n_i n_h}\sqrt{\frac{2[\mathrm{E}MS(h)]^2}{(n_h-1)}+\frac{2[(n_p-1)\mathrm{E}MS(ph)]^2}{(n_p-1)(n_h-1)}}$	(A.170)

Table A32. Estimated Bootstrap Standard Errors of Estimated Absolute Error Variances Based on Tong and Brennan's (2007) Proposed Workaround

Design	Estimated bootstrap standard error of estimated absolute error variance $S\hat{E}(\Delta)$	
$p \times i$	$\sqrt{S\hat{E}^2(\delta \mid p) + \frac{S\hat{E}^2(i \mid i)}{n_i^2}}$	(A.171)
$p \times i \times h$	$\sqrt{S\hat{E}^2(\delta \mid p) + \frac{S\hat{E}^2(i \mid i)}{n_i^2} + \frac{S\hat{E}^2(h \mid h)}{n_h^2} + \frac{S\hat{E}^2(ih \mid i*)}{n_i^2 n_h^2}}$	(A.172)
$i : p$	$\sqrt{S\hat{E}^2(\delta \mid p)}$	(A.173)
$i : h : p$	$\sqrt{S\hat{E}^2(\delta \mid p)}$	(A.174)
$p \times (i : h)$	$\sqrt{S\hat{E}^2(\delta \mid p) + \frac{S\hat{E}^2(h \mid h)}{n_h^2} + \frac{S\hat{E}^2(i : h \mid i)}{n_i^2 n_h^2}}$	(A.175)

*Use boot-*i* to estimate $SE^2(ih)$ assuming $n_i \geq n_h$. If $n_i < n_h$, then boot-*h* would be used.

APPENDIX B: PREVIOUS STUDIES OF BOOTSTRAP ESTIMATED STANDARD ERRORS OF ESTIMATED VARIANCE COMPONENTS

Table B1. Details of Previous Studies of Bootstrap Estimated Standard Errors of Estimated Variance Components

Study	Brennan, Harris, & Hanson (1987) (normal)	Brennan, Harris, & Hanson (1987) (dichotomous)	Luecht & Smith (1989)	Othman (1995)	Othman (1995)
G theory designs	$p \times i$	$p \times i$	$p \times i$	$p \times i$	$p \times i$
Sample size patterns	200 × 20	200 × 20 (from pop. of 2000 × 200)	20 × 20, 150 × 20, 20 × 150, 150 × 150	30 × 5, 30 × 20, 600 × 5, 600 × 20	30 × 5, 30 × 20, 600 × 5, 600 × 20
Variance component structures	$\sigma^2(p) = 4$, $\sigma^2(i) = 16$, $\sigma^2(pi) = 64$	$\sigma^2(p) = .0068$, $\sigma^2(i) = .0346$, $\sigma^2(pi) = .1902$	$\sigma^2(p) = .25$, $\sigma^2(i) = .25$, $\sigma^2(pi) = .5$	$\sigma^2(p) = 0.2$, $\sigma^2(i) = 0.1$, $\sigma^2(pi) = 0.7$	(see Table B2)
Standard error estimation methods	jackknife, bootstrap (boot-*p*, boot-*i*, boot-*p*,*i*, boot-*p*,*i*,*r*), traditional	jackknife, bootstrap (boot-*p*, boot-*i*, boot-*p*,*i*, boot-*p*,*i*,*r*), traditional	boot-*p*, boot-*p*,*i*	boot-*p*, boot-*i*, boot-*p*,*i*, traditional	boot-*p*, boot-*i*, boot-*p*,*i*, traditional
Data types	normal	dichotomous	normal	normal, skewed	dichotomous
Estimates	point estimates, SE's, & CI's	point estimates, SE's, & CI's	point estimates, SE's, & CI's	point estimates, SE's, & CI's	point estimates, SE's, & CI's
Number of datasets (replications)	1	100; 2000 to get "parameters"	476	1000	1000
Number of bootstrap samples	1000	100	200	500	500

Table B1. Continued

Study	Wiley (2001)	Tong & Brennan (2007)
G theory designs	$p \times i$	$p \times i \times h$, $p \times (i:h)$
Sample size patterns	30 × 5, 30 × 20, 600 × 5, 600 × 20	100 × 20 ×2 100 × 20 ×4
Variance component structures	$\sigma^2(p) = 0.2$, $\sigma^2(i) = 0.1$, $\sigma^2(pi) = 0.7$	(see Tables B3 and B4)
Standard error estimation methods	basic bootstrap, percentile bootstrap, BC bootstrap, traditional	bootstrap (see note below*)
Data types	normal, polytomous	normal, dichotomous, polytomous
Estimates	point estimates, SE's, & CI's	point estimates & SE's
Number of datasets (replications)	1000	1000; 5000 to get "parameters"
Number of bootstrap samples	1000	1000

* For the $p \times i \times h$ design: boot-*p*, boot-*i* boot-*h*, boot-*p*,*i*, boot-*p*,*h*, boot-*i*,*h*, boot-*p*,*i*,*h*. For the $p \times (i:h)$ design: boot-*p*, boot-*h*, boot-*p*,*h*, boot-*i*,*h*, boot-*p*,*i*,*h*.

Table B2. Othman (1995) Variance Component Structures for Dichotomous Data

	Variance components		
Sample sizes	*p*	*i*	*pi*
30 × 5	0.05	0.0083	0.25
30 × 20	0.0125	0.0083	0.25
600 × 5	0.02	0.0004	0.25
600 × 20	0.0125	0.0004	0.25

Table B3. Tong and Brennan's (2007) Variance Component Structures for the $p \times i \times h$ Design

	Variance components						
Data types	*p*	*i*	*h*	*pi*	*ph*	*ih*	*pih*
Normal	16	4	1	64	2	3	144
Dichotomous	0.0109	0.0028	0.0007	0.0449	0.0014	0.0021	0.1837
Polytomous	0.3241	0.127	0.012	0.393	0.014	0.0025	0.317

Table B4. Tong and Brennan's (2007) Variance Component Structures for the $p \times (i : h)$ Design

	Variance components				
Data types	*p*	*h*	*i*:*h*	*ph*	*pi*:*h*
Normal	16	1	7	2	208
Dichotomous	0.0108	0.0006	0.0048	0.0014	0.2323
Polytomous	0.2046	0.1324	0.4093	0.0651	1.1655

APPENDIX C: BRENNAN'S (2007) BIAS CORRECTIONS FOR BOOTSTRAP ESTIMATED VARIANCE COMPONENTS

In the following tables, where λ refers to the facet(s) being bootstrapped, and λ_j is any of the facets in λ,

$$s_{\lambda_j} = \frac{2n_{\lambda_j} - 1}{n_{\lambda_j}} \text{ and } t_{\lambda_j} = \frac{n_{\lambda_j} - 1}{n_{\lambda_j}} \quad \text{(C.1-2)}$$

Table C1. Bias-Corrected Variance Component Estimates for the $p \times i$ Design, Boot-*p*

α	$\hat{\sigma}^2(\alpha)$	
p	$\dfrac{\hat{\sigma}^2(p \mid p)}{t_p}$	(C.3)
i	$\hat{\sigma}^2(i \mid p) - \dfrac{\hat{\sigma}^2(pi \mid p)}{n_p - 1}$	(C.4)
pi	$\dfrac{\hat{\sigma}^2(pi \mid p)}{t_p}$	(C.5)

Table C2. Bias-Corrected Variance Component Estimates for the $p \times i$ Design, Boot-*i*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p \mid i) - \dfrac{\hat{\sigma}^2(pi \mid i)}{n_i - 1}$	(C.6)
i	$\dfrac{\hat{\sigma}^2(i \mid i)}{t_i}$	(C.7)
pi	$\dfrac{\hat{\sigma}^2(pi \mid i)}{t_i}$	(C.8)

Table C3. Bias-Corrected Variance Component Estimates for the $p \times i$ Design, Boot-*p*,*i*

α	$\hat{\sigma}^2(\alpha)$	
p	$\frac{1}{t_p}\left[\hat{\sigma}^2(p \mid p,i) - \frac{\hat{\sigma}^2(pi \mid p,i)}{n_i - 1}\right]$	(C.9)
i	$\frac{1}{t_i}\left[\hat{\sigma}^2(i \mid p,i) - \frac{\hat{\sigma}^2(pi \mid p,i)}{n_p - 1}\right]$	(C.10)
pi	$\frac{\hat{\sigma}^2(pi \mid p,i)}{t_p t_i}$	(C.11)

Table C4. Bias-Corrected Variance Component Estimates for the $p \times i \times h$ Design, Boot-*p*

α	$\hat{\sigma}^2(\alpha)$	
p	$\dfrac{\hat{\sigma}^2(p \mid p)}{t_p}$	(C.12)
i	$\hat{\sigma}^2(i \mid p) - \dfrac{\hat{\sigma}^2(pi \mid p)}{n_p - 1}$	(C.13)
h	$\hat{\sigma}^2(h \mid p) - \dfrac{\hat{\sigma}^2(ph \mid p)}{n_p - 1}$	(C.14)
pi	$\dfrac{\hat{\sigma}^2(pi \mid p)}{t_p}$	(C.15)
ph	$\dfrac{\hat{\sigma}^2(ph \mid p)}{t_p}$	(C.16)
ih	$\hat{\sigma}^2(ih \mid p) - \dfrac{\hat{\sigma}^2(pih \mid p)}{n_p - 1}$	(C.17)
pih	$\dfrac{\hat{\sigma}^2(pih \mid p)}{t_p}$	(C.18)

Table C5. Bias-Corrected Variance Component Estimates for the $p \times i \times h$ Design, Boot-*i*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p \mid i) - \dfrac{\hat{\sigma}^2(pi \mid i)}{n_i - 1}$	(C.19)
i	$\dfrac{\hat{\sigma}^2(i \mid i)}{t_i}$	(C.20)
h	$\hat{\sigma}^2(h \mid i) - \dfrac{\hat{\sigma}^2(ih \mid i)}{n_i - 1}$	(C.21)
pi	$\dfrac{\hat{\sigma}^2(pi \mid i)}{t_i}$	(C.22)
ph	$\hat{\sigma}^2(ph \mid i) - \dfrac{\hat{\sigma}^2(pih \mid i)}{n_i - 1}$	(C.23)
ih	$\dfrac{\hat{\sigma}^2(ih \mid i)}{t_i}$	(C.24)
pih	$\dfrac{\hat{\sigma}^2(pih \mid i)}{t_i}$	(C.25)

Table C6. Bias-Corrected Variance Component Estimates for the $p \times i \times h$ Design, Boot-*h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p \mid h) - \dfrac{\hat{\sigma}^2(ph \mid h)}{n_h - 1}$	(C.26)
i	$\hat{\sigma}^2(i \mid h) - \dfrac{\hat{\sigma}^2(ih \mid h)}{n_h - 1}$	(C.27)
h	$\dfrac{\hat{\sigma}^2(h \mid h)}{t_h}$	(C.28)
pi	$\hat{\sigma}^2(pi \mid h) - \dfrac{\hat{\sigma}^2(pih \mid h)}{n_h - 1}$	(C.29)
ph	$\dfrac{\hat{\sigma}^2(ph \mid h)}{t_h}$	(C.30)
ih	$\dfrac{\hat{\sigma}^2(ih \mid h)}{t_h}$	(C.31)
pih	$\dfrac{\hat{\sigma}^2(pih \mid h)}{t_h}$	(C.32)

Table C7. Bias-Corrected Variance Component Estimates for the $p \times i \times h$ Design, Boot-*p*,*i*

α	$\hat{\sigma}^2(\alpha)$	
p	$\frac{1}{t_p}\left[\hat{\sigma}^2(p \mid p,i) - \frac{\hat{\sigma}^2(pi \mid p,i)}{n_i - 1}\right]$	(C.33)
i	$\frac{1}{t_i}\left[\hat{\sigma}^2(i \mid p,i) - \frac{\hat{\sigma}^2(pi \mid p,i)}{n_p - 1}\right]$	(C.34)
h	$\hat{\sigma}^2(h \mid p,i) - \frac{\hat{\sigma}^2(ph \mid p,i)}{n_p - 1} - \frac{\hat{\sigma}^2(ih \mid p,i)}{n_i - 1} + \frac{\hat{\sigma}^2(pih \mid p,i)}{(n_p - 1)(n_i - 1)}$	(C.35)
pi	$\frac{\hat{\sigma}^2(pi \mid p,i)}{t_p t_i}$	(C.36)
ph	$\frac{1}{t_p}\left[\hat{\sigma}^2(ph \mid p,i) - \frac{\hat{\sigma}^2(pih \mid p,i)}{n_i - 1}\right]$	(C.37)
ih	$\frac{1}{t_i}\left[\hat{\sigma}^2(ih \mid p,i) - \frac{\hat{\sigma}^2(pih \mid p,i)}{n_p - 1}\right]$	(C.38)
pih	$\frac{\hat{\sigma}^2(pih \mid p,i)}{t_p t_i}$	(C.39)

Table C8. Bias-Corrected Variance Component Estimates for the $p \times i \times h$ Design, Boot-*p*,*h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\frac{1}{t_p}\left[\hat{\sigma}^2(p \mid p,h) - \frac{\hat{\sigma}^2(ph \mid p,h)}{n_h - 1}\right]$	(C.40)
i	$\hat{\sigma}^2(i \mid p,h) - \frac{\hat{\sigma}^2(pi \mid p,h)}{n_p - 1} - \frac{\hat{\sigma}^2(ih \mid p,h)}{n_h - 1} + \frac{\hat{\sigma}^2(pih \mid p,h)}{(n_p - 1)(n_h - 1)}$	(C.41)
h	$\frac{1}{t_h}\left[\hat{\sigma}^2(h \mid p,h) - \frac{\hat{\sigma}^2(ph \mid p,h)}{n_p - 1}\right]$	(C.42)
pi	$\frac{1}{t_p}\left[\hat{\sigma}^2(pi \mid p,h) - \frac{\hat{\sigma}^2(pih \mid p,h)}{n_h - 1}\right]$	(C.43)
ph	$\frac{\hat{\sigma}^2(ph \mid p,h)}{t_p t_h}$	(C.44)
ih	$\frac{1}{t_h}\left[\hat{\sigma}^2(ih \mid p,h) - \frac{\hat{\sigma}^2(pih \mid p,h)}{n_p - 1}\right]$	(C.45)
pih	$\frac{\hat{\sigma}^2(pih \mid p,h)}{t_p t_h}$	(C.46)

Table C9. Bias-Corrected Variance Component Estimates for the $p \times i \times h$ Design, Boot-*i,h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p \mid i,h) - \dfrac{\hat{\sigma}^2(pi \mid i,h)}{n_i - 1} - \dfrac{\hat{\sigma}^2(ph \mid i,h)}{n_h - 1} + \dfrac{\hat{\sigma}^2(pih \mid i,h)}{(n_i - 1)(n_h - 1)}$	(C.47)
i	$\dfrac{1}{t_i}\left[\hat{\sigma}^2(i \mid i,h) - \dfrac{\hat{\sigma}^2(ih \mid i,h)}{n_h - 1}\right]$	(C.48)
h	$\dfrac{1}{t_h}\left[\hat{\sigma}^2(h \mid i,h) - \dfrac{\hat{\sigma}^2(ih \mid i,h)}{n_i - 1}\right]$	(C.49)
pi	$\dfrac{1}{t_i}\left[\hat{\sigma}^2(pi \mid i,h) - \dfrac{\hat{\sigma}^2(pih \mid i,h)}{n_h - 1}\right]$	(C.50)
ph	$\dfrac{1}{t_h}\left[\hat{\sigma}^2(ph \mid i,h) - \dfrac{\hat{\sigma}^2(pih \mid i,h)}{n_i - 1}\right]$	(C.51)
ih	$\dfrac{\hat{\sigma}^2(ih \mid i,h)}{t_i t_h}$	(C.52)
pih	$\dfrac{\hat{\sigma}^2(pih \mid i,h)}{t_i t_h}$	(C.53)

Table C10. Bias-Corrected Variance Component Estimates for the $p \times i \times h$ Design, Boot-*p*,*i*,*h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\frac{1}{t_p}\left[\hat{\sigma}^2(p \mid p,i,h) - \frac{\hat{\sigma}^2(pi \mid p,i,h)}{n_i - 1} - \frac{\hat{\sigma}^2(ph \mid p,i,h)}{n_h - 1} + \frac{\hat{\sigma}^2(pih \mid p,i,h)}{(n_i - 1)(n_h - 1)}\right]$	(C.54)
i	$\frac{1}{t_i}\left[\hat{\sigma}^2(i \mid p,i,h) - \frac{\hat{\sigma}^2(pi \mid p,i,h)}{n_p - 1} - \frac{\hat{\sigma}^2(ih \mid p,i,h)}{n_h - 1} + \frac{\hat{\sigma}^2(pih \mid p,i,h)}{(n_p - 1)(n_h - 1)}\right]$	(C.55)
h	$\frac{1}{t_h}\left[\hat{\sigma}^2(h \mid p,i,h) - \frac{\hat{\sigma}^2(ph \mid p,i,h)}{n_p - 1} - \frac{\hat{\sigma}^2(ih \mid p,i,h)}{n_i - 1} + \frac{\hat{\sigma}^2(pih \mid p,i,h)}{(n_p - 1)(n_i - 1)}\right]$	(C.56)
pi	$\frac{1}{t_p t_i}\left[\hat{\sigma}^2(pi \mid p,i,h) - \frac{\hat{\sigma}^2(pih \mid p,i,h)}{n_h - 1}\right]$	(C.57)
ph	$\frac{1}{t_p t_h}\left[\hat{\sigma}^2(ph \mid p,i,h) - \frac{\hat{\sigma}^2(pih \mid p,i,h)}{n_i - 1}\right]$	(C.58)
ih	$\frac{1}{t_i t_h}\left[\hat{\sigma}^2(ih \mid p,i,h) - \frac{\hat{\sigma}^2(pih \mid p,i,h)}{n_p - 1}\right]$	(C.59)
pih	$\frac{\hat{\sigma}^2(pih \mid p,i,h)}{t_p t_i t_h}$	(C.60)

Table C11. Bias-Corrected Variance Component Estimates for the $i : p$ Design, Boot-*p*

α	$\hat{\sigma}^2(\alpha)$	
p	$\frac{1}{t_p}\left[\hat{\sigma}^2(p \mid p) - \frac{(t_p - 1)\hat{\sigma}^2(i : p \mid p)}{n_i}\right]$	(C.61)
i:p	$\hat{\sigma}^2(i : p \mid p)$	(C.62)

Table C12. Bias-Corrected Variance Component Estimates for the $i:p$ Design, Boot-*i*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p \mid i) - \dfrac{\hat{\sigma}^2(i:p \mid i)}{n_i - 1}$	(C.63)
$i{:}p$	$\dfrac{\hat{\sigma}^2(i:p \mid i)}{t_i}$	(C.64)

Table C13. Bias-Corrected Variance Component Estimates for the $i:p$ Design, Boot-*p,i*

α	$\hat{\sigma}^2(\alpha)$	
p	$\dfrac{1}{t_p}\left[\hat{\sigma}^2(p \mid p,i) - \left(\dfrac{s_i t_p - t_i}{n_i - 1}\right)\hat{\sigma}^2(i:p \mid p,i)\right]$	(C.65)
$i{:}p$	$\dfrac{\hat{\sigma}^2(i:p \mid p,i)}{t_i}$	(C.66)

Table C14. Bias-Corrected Variance Component Estimates for the $i:h:p$ Design, Boot-*p*

α	$\hat{\sigma}^2(\alpha)$	
p	$\dfrac{1}{t_p}\left[\hat{\sigma}^2(p \mid p) - \left(\dfrac{t_p - 1}{n_h}\right)\hat{\sigma}^2(h:p \mid p) - \left(\dfrac{t_p - 1}{n_i n_h}\right)\hat{\sigma}^2(i:h:p \mid p)\right]$	(C.67)
$h{:}p$	$\hat{\sigma}^2(h:p \mid p)$	(C.68)
$i{:}h{:}p$	$\hat{\sigma}^2(i:h:p \mid p)$	(C.69)

Table C15. Bias-Corrected Variance Component Estimates for the $i:h:p$ Design, Boot-*i*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p\,|\,i)$	(C.70)
$h{:}p$	$\hat{\sigma}^2(h:p\,|\,i)-\dfrac{\hat{\sigma}^2(i:h:p\,|\,i)}{n_i-1}$	(C.71)
$i{:}h{:}p$	$\dfrac{\hat{\sigma}^2(i:h:p\,|\,i)}{t_i}$	(C.72)

Table C16. Bias-Corrected Variance Component Estimates for the $i:h:p$ Design, Boot-*h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p\,|\,h)-\dfrac{\hat{\sigma}^2(h:p\,|\,h)}{n_h-1}-\dfrac{\hat{\sigma}^2(i:h:p\,|\,h)}{n_i(n_h-1)}$	(C.73)
$h{:}p$	$\dfrac{1}{t_h}\left[\hat{\sigma}^2(h:p\,|\,h)-\left(\dfrac{t_h-1}{n_i}\right)\hat{\sigma}^2(i:h:p\,|\,h)\right]$	(C.74)
$i{:}h{:}p$	$\hat{\sigma}^2(i:h:p\,|\,h)$	(C.75)

Table C17. Bias-Corrected Variance Component Estimates for the $i:h:p$ Design, Boot-*p,i*

α	$\hat{\sigma}^2(\alpha)$	
p	$\dfrac{1}{t_p}\left\{ \hat{\sigma}^2(p \mid p,i) - \left(\dfrac{t_p - 1}{n_h}\right)\hat{\sigma}^2(h:p \mid p,i) - \left[\dfrac{(s_i - 1)(t_p - 1)}{(n_i - 1)n_h}\right]\hat{\sigma}^2(i:h:p \mid p,i) \right\}$	(C.76)
h:p	$\hat{\sigma}^2(h:p \mid p,i) - \dfrac{\hat{\sigma}^2(i:h:p \mid p,i)}{n_i - 1}$	(C.77)
i:h:p	$\dfrac{\hat{\sigma}^2(i:h:p \mid p,i)}{t_i}$	(C.78)

Table C18. Bias-Corrected Variance Component Estimates for the $i:h:p$ Design, Boot-*p,h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\dfrac{1}{t_p}\left[\hat{\sigma}^2(p \mid p,h) - (s_h t_p - t_h)\left[\dfrac{\hat{\sigma}^2(h:p \mid p,h)}{n_h - 1} + \dfrac{\hat{\sigma}^2(i:h:p \mid p,h)}{n_i(n_h - 1)}\right]\right]$	(C.79)
h:p	$\dfrac{1}{t_h}\left[\hat{\sigma}^2(h:p \mid p,h) - \left(\dfrac{t_h - 1}{n_i}\right)\hat{\sigma}^2(i:h:p \mid p,h)\right]$	(C.80)
i:h:p	$\hat{\sigma}^2(i:h:p \mid p,h)$	(C.81)

Table C19. Bias-Corrected Variance Component Estimates for the $i:h:p$ Design, Boot-*i,h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p \mid i,h) - \dfrac{\hat{\sigma}^2(h:p \mid i,h)}{(n_h - 1)} - \dfrac{\hat{\sigma}^2(i:h:p \mid i,h)}{n_i(n_h - 1)}$	(C.82)
$h{:}p$	$\dfrac{1}{t_h}\left[\hat{\sigma}^2(h:p \mid i,h) - \left(\dfrac{s_i t_h - t_i}{n_i - 1}\right)\hat{\sigma}^2(i:h:p \mid i,h)\right]$	(C.83)
$i{:}h{:}p$	$\dfrac{\hat{\sigma}^2(i:h:p \mid i,h)}{t_i}$	(C.84)

Table C20. Bias-Corrected Variance Component Estimates for the $i:h:p$ Design, Boot-*p,i,h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\dfrac{1}{t_p}\left\{\begin{array}{l}\hat{\sigma}^2(p \mid p,i,h) - \left(\dfrac{s_h t_p - t_h}{n_h - 1}\right)\hat{\sigma}^2(h:p \mid p,i,h) \\ -\left[\dfrac{s_h t_p - t_h}{n_i(n_h - 1)}\right]\hat{\sigma}^2(i:h:p \mid p,i,h)\end{array}\right\}$	(C.85)
$h{:}p$	$\dfrac{1}{t_h}\left[\hat{\sigma}^2(h:p \mid p,i,h) - \left(\dfrac{s_i t_h - t_i}{n_i - 1}\right)\hat{\sigma}^2(i:h:p \mid p,i,h)\right]$	(C.86)
$i{:}h{:}p$	$\dfrac{\hat{\sigma}^2(i:h:p \mid p,i,h)}{t_i}$	(C.87)

Table C21. Bias-Corrected Variance Component Estimates for the $p \times (i:h)$ Design, Boot-p

α	$\hat{\sigma}^2(\alpha)$	
p	$\dfrac{\hat{\sigma}^2(p \mid p)}{t_p}$	(C.88)
h	$\hat{\sigma}^2(h \mid p) - \dfrac{\hat{\sigma}^2(ph \mid p)}{n_p - 1}$	(C.89)
ph	$\dfrac{\hat{\sigma}^2(ph \mid p)}{t_p}$	(C.90)
$i{:}h$	$\hat{\sigma}^2(i:h \mid p) - \dfrac{\hat{\sigma}^2(pi:h \mid p)}{n_p - 1}$	(C.91)
$pi{:}h$	$\dfrac{\hat{\sigma}^2(pi:h \mid p)}{t_p}$	(C.92)

Table C22. Bias-Corrected Variance Component Estimates for the $p \times (i:h)$ Design, Boot-i

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p \mid i)$	(C.93)
h	$\hat{\sigma}^2(h \mid i) - \dfrac{\hat{\sigma}^2(i:h \mid i)}{n_i - 1}$	(C.94)
ph	$\hat{\sigma}^2(ph \mid i) - \dfrac{\hat{\sigma}^2(pi:h \mid i)}{n_i - 1}$	(C.95)
$i{:}h$	$\dfrac{\hat{\sigma}^2(i:h \mid i)}{t_i}$	(C.96)
$pi{:}h$	$\dfrac{\hat{\sigma}^2(pi:h \mid i)}{t_i}$	(C.97)

Table C23. Bias-Corrected Variance Component Estimates for the $p \times (i:h)$ Design, Boot-*h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p\,|\,h) - \dfrac{\hat{\sigma}^2(ph\,|\,h)}{n_h - 1} - \dfrac{\hat{\sigma}^2(pi:h\,|\,h)}{n_i(n_h - 1)}$	(C.98)
h	$\dfrac{1}{t_h}\left[\hat{\sigma}^2(h\,|\,h) - \dfrac{\hat{\sigma}^2(i:h\,|\,h)}{n_i n_h}\right]$	(C.99)
ph	$\dfrac{1}{t_h}\left[\hat{\sigma}^2(ph\,|\,h) + \dfrac{\hat{\sigma}^2(pi:h\,|\,h)}{n_i n_h}\right]$	(C.100)
$i{:}h$	$\hat{\sigma}^2(i:h\,|\,h)$	(C.101)
$pi{:}h$	$\hat{\sigma}^2(pi:h\,|\,h)$	(C.102)

Table C24. Bias-Corrected Variance Component Estimates for the $p \times (i:h)$ Design, Boot-*p*,*i*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p\,|\,p,i)$	(C.103)
h	$\hat{\sigma}^2(h\,|\,p,i) - \dfrac{\hat{\sigma}^2(ph\,|\,p,i)}{n_p - 1} - \dfrac{\hat{\sigma}^2(i:h\,|\,p,i)}{n_i - 1} + \dfrac{\hat{\sigma}^2(pi:h\,|\,p,i)}{(n_p - 1)(n_i - 1)}$	(C.104)
ph	$\dfrac{1}{t_p}\left[\hat{\sigma}^2(ph\,|\,p,i) - \dfrac{\hat{\sigma}^2(pi:h\,|\,p,i)}{(n_i - 1)}\right]$	(C.105)
$i{:}h$	$\dfrac{1}{t_i}\left[\hat{\sigma}^2(i:h\,|\,p,i) - \dfrac{\hat{\sigma}^2(pi:h\,|\,p,i)}{(n_p - 1)}\right]$	(C.106)
$pi{:}h$	$\dfrac{\hat{\sigma}^2(pi:h\,|\,p,i)}{t_p t_i}$	(C.107)

Table C25. Bias-Corrected Variance Component Estimates for the $p \times (i:h)$ Design, Boot-*p*,*h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\frac{1}{t_p}\left[\hat{\sigma}^2(p \mid p,h) - \frac{\hat{\sigma}^2(ph \mid p,h)}{n_h - 1} - \frac{\hat{\sigma}^2(pi:h \mid p,h)}{n_i(n_h - 1)}\right]$	(C.108)
h	$\frac{1}{t_h}\left[\hat{\sigma}^2(h \mid p,h) - \frac{\hat{\sigma}^2(ph \mid p,h)}{n_p - 1} + \frac{\hat{\sigma}^2(i:h \mid p,h)}{n_i n_h} - \frac{\hat{\sigma}^2(pi:h \mid p,h)}{n_i n_h (n_p - 1)}\right]$	(C.109)
ph	$\frac{1}{t_p t_h}\left[\hat{\sigma}^2(ph \mid p,h) + \frac{\hat{\sigma}^2(pi:h \mid p,h)}{n_i n_h}\right]$	(C.110)
$i{:}h$	$\hat{\sigma}^2(i:h \mid p,h) - \frac{\hat{\sigma}^2(pi:h \mid p,h)}{n_p - 1}$	(C.111)
$pi{:}h$	$\frac{\hat{\sigma}^2(pi:h \mid p,h)}{t_p}$	(C112)

Table C26. Bias-Corrected Variance Component Estimates for the $p \times (i:h)$ Design, Boot-*i,h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\hat{\sigma}^2(p \mid i,h) - \dfrac{\hat{\sigma}^2(ph \mid i,h)}{n_h - 1} - \dfrac{\hat{\sigma}^2(pi:h \mid i,h)}{n_i(n_h - 1)}$	(C.113)
h	$\dfrac{1}{t_h}\left[\hat{\sigma}^2(h \mid i,h) - \left(\dfrac{s_i t_h - t_i}{n_i - 1}\right)\hat{\sigma}^2(i:h \mid i,h)\right]$	(C.114)
ph	$\dfrac{1}{t_h}\left[\hat{\sigma}^2(ph \mid i,h) + \left(\dfrac{s_i t_h - t_i}{n_i - 1}\right)\hat{\sigma}^2(pi:h \mid i,h)\right]$	(C.115)
i:h	$\dfrac{\hat{\sigma}^2(i:h \mid i,h)}{t_i}$	(C.116)
pi:h	$\dfrac{\hat{\sigma}^2(pi:h \mid i,h)}{t_i}$	(C.117)

Table C27. Bias-Corrected Variance Component Estimates for the $p \times (i:h)$ Design, Boot-*p*,*i*,*h*

α	$\hat{\sigma}^2(\alpha)$	
p	$\frac{1}{t_p}\left[\hat{\sigma}^2(p \mid p,i,h) - \frac{\hat{\sigma}^2(ph \mid p,i,h)}{n_h - 1} - \frac{\hat{\sigma}^2(pi:h \mid p,i,h)}{n_i(n_h - 1)}\right]$	(C.118)
h	$\frac{1}{t_h}\left\{\hat{\sigma}^2(h \mid p,i,h) - \frac{\hat{\sigma}^2(ph \mid p,i,h)}{n_p - 1} - \left(\frac{s_i t_h - t_i}{n_i - 1}\right)\left[\hat{\sigma}^2(i:h \mid p,i,h) - \frac{\hat{\sigma}^2(pi:h \mid p,i,h)}{n_p - 1}\right]\right\}$	(C.119)
ph	$\frac{1}{t_p t_h}\left[\hat{\sigma}^2(ph \mid p,i,h) - \left(\frac{s_i t_h - t_i}{n_i - 1}\right)\hat{\sigma}^2(pi:h \mid p,i,h)\right]$	(C.120)
$i:h$	$\frac{1}{t_i}\left[\hat{\sigma}^2(i:h \mid p,i,h) - \frac{\hat{\sigma}^2(pi:h \mid p,i,h)}{n_p - 1}\right]$	(C.121)
$pi:h$	$\frac{\hat{\sigma}^2(pi:h \mid p,i,h)}{t_p t_i}$	(C.122)

APPENDIX D: TONG AND BRENNAN'S (2007) RULES FOR PICKING A BOOTSTRAP PROCEDURE TO ESTIMATE STANDARD ERRORS FOR ESTIMATED VARIANCE COMPONENTS

Rule 1: For nonnested main effects, use the corresponding bootstrap procedure. For example, to estimate $SE\left[\hat{\sigma}^2(p)\right]$, use boot–*p*.

Rule 2: For nested main effects (e.g., *i*:*h* in the *p* × (*i*:*h*) design), use the bootstrap procedure for the primary index (i.e., the index before the colon, following the notational conventions in Brennan, 2001). For example, to estimate $SE\left[\hat{\sigma}^2(i:h)\right]$ for the *p* ×(*i*:*h*) design, use boot–*i*.

Rule 3: For nonnested interaction effects, use the one-dimensional bootstrap procedure for the facet in the interaction that has the largest sample size. For example, to estimate $SE\left[\hat{\sigma}^2(ph)\right]$, use boot–*p* if $n_p > n_h$ or use boot–*h* if $n_p < n_h$.

Rule 4: For nested interaction effects, use the one-dimensional bootstrap procedure for the primary facet in the interaction that has the largest sample size. For example, to estimate $SE\left[\hat{\sigma}^2(pi:h)\right]$, use boot–*p* if $n_p > n_i$ or use boot–*i* if $n_p < n_i$.

Rule 5: For $SE(\delta)$,use the one-dimensional bootstrap procedure for the objects-of-measurement facet. In this article, it is assumed that *p* is the objects-of-measurement facet.

Proposed rule for $SE(\Delta)$: Approximate $SE(\Delta)$ using a combination of bootstrap procedures, where the bootstrap procedure used to estimate the each of the component standard errors are determined by the rules above.

APPENDIX E: SUMMARY OF EMPIRICAL STUDIES USING GENERALIZABILITY THEORY

Table E1. Variance Component Estimates From Empirical Studies Using the $p \times i$ Design

				Proportions of Variance Explained		
Study	score scale	n_p	n_i	*p*	*i*	*pi*
Lane, Liu, Ankenmann, & Stone (1996)	0-4	73	9	32	10	58
Lane, Liu, Ankenmann, & Stone (1996)	0-4	103	9	29	17	54
Lane, Liu, Ankenmann, & Stone (1996)	0-4	106	9	21	9	70
Lane, Liu, Ankenmann, & Stone (1996)	0-4	105	9	24	11	65
Lane, Liu, Ankenmann, & Stone (1996)	0-4	170	9	27	12	62
Lane, Liu, Ankenmann, & Stone (1996)	0-4	209	9	31	10	58
Lane, Liu, Ankenmann, & Stone (1996)	0-4	234	9	20	11	69
Lane, Liu, Ankenmann, & Stone (1996)	0-4	209	9	27	11	62
Lane, Liu, Ankenmann, & Stone (1996)	0-4	106	9	22	13	65
Lane, Liu, Ankenmann, & Stone (1996)	0-4	110	9	31	16	53
Lane, Liu, Ankenmann, & Stone (1996)	0-4	117	9	20	7	73
Lane, Liu, Ankenmann, & Stone (1996)	0-4	101	9	30	10	59
Lane, Liu, Ankenmann, & Stone (1996)	0-4	190	9	25	14	61
Lane, Liu, Ankenmann, & Stone (1996)	0-4	238	9	34	9	57
Lane, Liu, Ankenmann, & Stone (1996)	0-4	214	9	27	7	67
Lane, Liu, Ankenmann, & Stone (1996)	0-4	229	9	33	8	59
Lane, Stone, Ankenmann, & Liu (1994)	0-4	306	9	26	10	64
Lane, Stone, Ankenmann, & Liu (1994)	0-4	331	9	31	14	55
Lane, Stone, Ankenmann, & Liu (1994)	0-4	329	9	21	11	68
Lane, Stone, Ankenmann, & Liu (1994)	0-4	331	9	22	18	60
Turner, Lozano-Nieto, & Bouffard (2003)	kHz	50	20	99.6	0.1	0.4
Turner, Lozano-Nieto, & Bouffard (2003)	kHz	50	20	99.8	0.0	0.2
Turner, Lozano-Nieto, & Bouffard (2003)	kHz	50	20	99.9	0.1	0.0
Turner, Lozano-Nieto, & Bouffard (2003)	kHz	50	20	99.8	0.1	0.1
Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams (2009)	Angoff prob	6	200	3	40	57

Table E1. Continued

Study	score scale	n_p	n_i	Proportions of Variance Explained p	i	pi
Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams (2009)	Angoff prob	6	200	1	77	23
Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams (2009)	Angoff prob	6	34	1	37	61
Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams (2009)	Angoff prob	6	34	0	70	30
Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams (2009)	Angoff prob	6	34	0	80	19
Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams (2009)	Angoff prob	6	195	3	40	57
Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams (2009)	Angoff prob	6	195	2	75	24
Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams (2009)	Angoff prob	6	43	0	41	59
Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams (2009)	Angoff prob	6	43	1	68	31
Clauser, Harik, Margolis, McManus, Mollon, Chis, & Williams (2009)	Angoff prob	6	43	1	84	15
Harik, Clauser, Grabovsky, Nungester, Swanson, & Nandakumar (2009)	6 pt.	1200	11	25	28	47
Harik, Clauser, Grabovsky, Nungester, Swanson, & Nandakumar (2009)	6 pt.	3700	11	25	26	49
Harik, Clauser, Grabovsky, Nungester, Swanson, & Nandakumar (2009)	6 pt.	1200	11	13	28	59
Harik, Clauser, Grabovsky, Nungester, Swanson, & Nandakumar (2009)	6 pt.	3700	11	14	30	57
Harik, Clauser, Grabovsky, Nungester, Swanson, & Nandakumar (2009)	6 pt.	1200	11	15	23	62
Harik, Clauser, Grabovsky, Nungester, Swanson, & Nandakumar (2009)	6 pt.	3700	11	16	24	60
Harik, Clauser, Grabovsky, Nungester, Swanson, & Nandakumar (2009)	6 pt.	1200	11	76	4	20
Harik, Clauser, Grabovsky, Nungester, Swanson, & Nandakumar (2009)	6 pt.	3700	11	76	5	19

Table E2. Variance Component Estimates From Empirical Studies Using the $p \times i \times h$ Design

		Proportions of Variance Explained									
Study	score scale	n_p	n_i	n_h	*p*	*i*	*h*	*pi*	*ph*	*ih*	*pih*
Lane, Liu, Ankenmann, & Stone (1996)	0-4	26	9	2	23	0	14	0	54	0	9
Lane, Liu, Ankenmann, & Stone (1996)	0-4	25	9	2	22	0	14	0	55	0	8
Lane, Liu, Ankenmann, & Stone (1996)	0-4	33	9	2	43	0	9	0	41	0	7
Lane, Liu, Ankenmann, & Stone (1996)	0-4	38	9	2	18	1	10	0	62	0	10
Lane, Liu, Ankenmann, & Stone (1996)	0-4	27	9	2	37	0	8	1	42	0	12
Lane, Liu, Ankenmann, & Stone (1996)	0-4	51	9	2	13	0	12	0	62	1	11
Lane, Liu, Ankenmann, & Stone (1996)	0-4	43	9	2	20	0	15	0	56	0	8
Lane, Liu, Ankenmann, & Stone (1996)	0-4	46	9	2	32	0	8	0	52	0	8
Lane, Stone, Ankenmann, & Liu (1994)	0-4	306	4	2	20	0	15	0	53	0	11
Lane, Stone, Ankenmann, & Liu (1994)	0-4	306	4	2	37	0	12	0	40	0	10

Table E2. Continued

Study	score scale				Proportions of Variance Explained						
		n_p	n_i	n_h	*p*	*i*	*h*	*pi*	*ph*	*ih*	*pih*
Lane, Stone, Ankenmann, & Liu (1994)	0-4	306	4	2	13	0	7	0	65	1	14
Lane, Stone, Ankenmann, & Liu (1994)	0-4	306	4	2	21	0	7	0	59	0	12
Lane, Stone, Ankenmann, & Liu (1994)	0-4	331	4	2	34	0	8	0	43	0	14
Lane, Stone, Ankenmann, & Liu (1994)	0-4	331	4	2	35	0	3	0	56	0	5
Lane, Stone, Ankenmann, & Liu (1994)	0-4	331	4	2	42	0	6	0	47	0	5
Lane, Stone, Ankenmann, & Liu (1994)	0-4	329	5	2	23	0	7	0	63	0	6
Lane, Stone, Ankenmann, & Liu (1994)	0-4	329	5	2	23	2	11	0	55	0	9
Lane, Stone, Ankenmann, & Liu (1994)	0-4	329	5	2	15	0	9	0	70	0	6
Lane, Stone, Ankenmann, & Liu (1994)	0-4	331	3	2	22	0	35	0	36	0	6
Lane, Stone, Ankenmann, & Liu (1994)	0-4	331	3	2	13	0	39	0	40	0	8

Table E2. Continued

		Proportions of Variance Explained									
Study	score scale	n_p	n_i	n_h	*p*	*i*	*h*	*pi*	*ph*	*ih*	*pih*
Lane, Stone, Ankenmann, & Liu (1994)	0-4	331	3	2	16	0	35	0	42	0	7
Shavelson, Baxter, & Gao (1993)	dichot & 6 pt.	186	2	4	19	16	16	16	3	3	29
Heitman, Kovaleski, & Pugh (2009)	mm & °	20	2	2	0	0	84	0	4	0	11
Heitman, Kovaleski, & Pugh (2009)		20	2	2	2	0	89	0	2	0	7
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	11	19	11	7	22	2	28
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	9	3	60	1	15	0	12
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	6	26	1	9	17	1	39
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	9	5	34	4	15	3	30
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	14	16	18	6	10	1	35
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	2	11	18	8	16	2	43
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	23	2	10	7	13	0	45
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	10	10	23	1	28	0	28
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	10	8	10	8	27	0	37
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	9	8	16	4	27	2	35
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	12	6	0	4	26	0	52

Table E2. Continued

Study	score scale				Proportions of Variance Explained						
		n_p	n_i	n_h	*p*	*i*	*h*	*pi*	*ph*	*ih*	*pih*
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	15	2	31	3	18	5	27
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	8	19	0	11	26	1	34
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	1	13	8	6	34	4	33
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	26	5	6	6	20	1	37
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	11	2	12	9	23	0	44
Lakes & Hoyt (2008)	sqrt(prop)	27	3	5	25	4	0	13	8	0	50
Lee & Kantor (2007)	1-5	160	6	6	49	8	1	18	1	1	21
Yin & Shavelson (2008)	0-4	16	2	26	10	0	9	0	16	2	62
Yin & Shavelson (2008)	0-4	16	2	21	19	0	5	0	25	0	50
Gao & Brennan (2001)	0-5	50	12	3	27	3	9	1	34	1	26
Gao & Brennan (2001)	0-5	50	12	3	27	0	4	1	32	4	32
Gao & Brennan (2001)	0-5	50	12	3	26	0	16	2	32	1	22
Gao & Brennan (2001)	0-5	50	12	3	59	2	2	3	12	1	20
Gao & Brennan (2001)	0-5	50	12	3	53	0	3	2	18	0	23
Gao & Brennan (2001)	0-5	50	12	3	65	0	1	6	12	0	15
Gao & Brennan (2001)	0-5	167	6	3	23	1	20	1	43	0	13
Gao & Brennan (2001)	0-5	167	6	3	23	0	25	1	36	0	14
Gao & Brennan (2001)	0-5	167	6	3	48	0	2	2	27	1	20

Table E2. Continued

		Proportions of Variance Explained									
Study	score scale	n_p	n_i	n_h	*p*	*i*	*h*	*pi*	*ph*	*ih*	*pih*
Gao & Brennan (2001)	0-5	167	6	3	48	1	1	3	31	0	16
Huang (2008)	6 pt.	323	3	2	41	1	3	33	0	1	21
Huang (2008)	6 pt.	425	3	2	50	1	0	27	0	1	21
Huang (2008)	6 pt.	357	3	2	37	2	3	39	0	1	18
Huang (2008)	6 pt.	323	3	2	49	3	4	32	0	0	11
Huang (2008)	6 pt.	425	3	2	58	1	1	28	0	0	12
Huang (2008)	6 pt.	357	3	2	64	1	1	28	0	0	6
Moscoso Pablo Tello Lopez (2006)	5 pt.	3	28	42	0	9	26	2	0	40	23
Moscoso Pablo Tello Lopez (2006)	5 pt.	3	28	42	0	9	25	2	0	42	22
Chafouleas, Christ, & Riley-Tillman (2009)	6 pt.	6	106	3	16	17	5	8	15	1	39
Chafouleas, Christ, & Riley-Tillman (2009)	10 pt.	6	106	3	12	21	4	11	17	1	34
Chafouleas, Christ, & Riley-Tillman (2009)	14 pt.	6	106	3	15	20	6	11	11	2	35
Chafouleas, Christ, & Riley-Tillman (2009)	6 pt.	6	106	3	13	22	5	10	15	0	36

Table E2. Continued

		Proportions of Variance Explained									
Study	score scale	n_p	n_i	n_h	*p*	*i*	*h*	*pi*	*ph*	*ih*	*pih*
Chafouleas, Christ, & Riley-Tillman (2009)	10 pt.	6	106	3	11	21	3	11	17	1	36
Chafouleas, Christ, & Riley-Tillman (2009)	14 pt.	6	106	3	13	23	4	11	12	1	35
Solano-Flores & Li (2009)	1-3	27	4	4	6	12	0	50	0	2	29
Solano-Flores & Li (2009)	1-3	55	4	4	5	6	0	55	4	2	28
Solano-Flores & Li (2009)	1-3	41	4	4	6	9	2	49	2	3	29
Solano-Flores & Li (2009)	1-3	123	4	4	9	8	1	50	3	1	28

Table E3. Variance Component Estimates From Empirical Studies Using the $i : p$ Design

Study	score scale	n_p	n_i	Proportions of Variance Explained	
				p	$i{:}p$
Burch, Norman, Schmidt, & van der Vleuten (2008)	percent	69	8	15	85
Burch, Norman, Schmidt, & van der Vleuten (2008)	percent	69	3	32	68
Burch, Norman, Schmidt, & van der Vleuten (2008)	percent	69	20	8	92

Table E4. Variance Component Estimates From Empirical Studies Using the $i : h : p$ Design

					Proportions of Variance Explained		
Study	score scale	n_p	n_i	n_h	*p*	*h:p*	*i:h:p*
Lee & Lewis (2008)	bookmark	6	4-6	2	15	15	71
Lee & Lewis (2008)	bookmark	6	4-6	2	95	0	5
Lee & Lewis (2008)	bookmark	6	4-6	2	77	0	23
Lee & Lewis (2008)	bookmark	6	4-6	2	52	0	48
Lee & Lewis (2008)	bookmark	6	4-6	2	29	2	68
Lee & Lewis (2008)	bookmark	6	4-6	2	75	0	25
Lee & Lewis (2008)	bookmark	6	4-6	2	51	0	49
Lee & Lewis (2008)	bookmark	6	4-6	2	48	40	12
Lee & Lewis (2008)	bookmark	6	4-6	2	0	50	50
Yin & Sconing (2008)	percent	2	5-6	2	0	9	91
Yin & Sconing (2008)	percent	2	5-6	2	11	0	89
Yin & Sconing (2008)	percent	2	5-6	2	0	23	77
Yin & Sconing (2008)	percent	2	5-6	2	0	0	100
Yin & Sconing (2008)	percent	2	5-6	2	0	4	96
Yin & Sconing (2008)	percent	2	5-6	2	0	0	100
Yin & Sconing (2008)	bookmark	2	5-6	2	0	55	45
Yin & Sconing (2008)	bookmark	2	5-6	2	23	0	77
Yin & Sconing (2008)	bookmark	2	5-6	2	0	25	75
Yin & Sconing (2008)	bookmark	2	5-6	2	2	0	98
Yin & Sconing (2008)	bookmark	2	5-6	2	0	55	45
Yin & Sconing (2008)	bookmark	2	5-6	2	0	49	51

Table E5. Variance Component Estimates From Empirical Studies Using the $p \times (i:h)$ Design

Study	score scale	n_p	n_i	n_h	Proportions of Variance Explained				
					p	*h*	*i:h*	*ph*	*pi:h*
Christophersen, Helseth, & Lund (2008)	1-5	239	4	6	11	6	15	13	55
Lee & Frisbie (1999)	0,1	3032	2-6	9	13	5	6	7	69
Lee & Frisbie (1999)	0,1	3074	3-12	9	16	2	6	4	70
Lee & Frisbie (1999)	0,1	3003	6-7	4	15	0	3	8	74
Lee & Frisbie (1999)	0,1	3007	6-7	5	12	1	4	5	78
Lee & Frisbie (1999)	0,1	2919	9	5	20	0	4	4	71
Gagnon, Charlin, Lambert, Carriere, & Van der Vleuten (2009)	partial credits	30	3	16	6	0	6	2	86
Gagnon, Charlin, Lambert, Carriere, & Van der Vleuten (2009)	partial credits	49	2	22	3	1	10	0	85
Gagnon, Charlin, Lambert, Carriere, & Van der Vleuten (2009)	partial credits	106	3	30	8	0	0	0	92

APPENDIX F: ESTIMATED VARIANCE COMPONENTS

Table F1. Estimated Variance Components for the $p \times i$ Design, Normal Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.3500	0.0500	0.6000	0.0600	0.0650
Empirical	0.3495	0.0511	0.6002	0.0600	0.0651
p	0.3495	0.0511	0.6002	0.0600	0.0651
i	0.3496	0.0511	0.6001	0.0600	0.0651
p,i	0.3495	0.0512	0.6001	0.0600	0.0651
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.3500	0.0500	0.6000	0.0120	0.0130
Empirical	0.3517	0.0496	0.6002	0.0120	0.0130
p	0.3516	0.0496	0.6002	0.0120	0.0130
i	0.3517	0.0496	0.6002	0.0120	0.0130
p,i	0.3516	0.0496	0.6002	0.0120	0.0130
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.3000	0.2000	0.5000	0.0500	0.0700
Empirical	0.2996	0.2056	0.5001	0.0500	0.0706
p	0.2996	0.2056	0.5001	0.0500	0.0706
i	0.2996	0.2056	0.5001	0.0500	0.0706
p,i	0.2996	0.2057	0.5001	0.0500	0.0706
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.3000	0.2000	0.5000	0.0100	0.0140
Empirical	0.3014	0.1982	0.5002	0.0100	0.0140
p	0.3014	0.1982	0.5001	0.0100	0.0140
i	0.3015	0.1981	0.5002	0.0100	0.0140
p,i	0.3014	0.1982	0.5001	0.0100	0.0140

Table F2. Estimated Variance Components for the $p \times i$ Design, Dichotomous Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.0242	0.0030	0.1065	0.0107	0.0110
Empirical	0.0242	0.0030	0.1063	0.0106	0.0109
p	0.0242	0.0030	0.1063	0.0106	0.0109
i	0.0242	0.0030	0.1062	0.0106	0.0109
p,i	0.0242	0.0030	0.1062	0.0106	0.0109
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.0243	0.0030	0.1063	0.0021	0.0022
Empirical	0.0244	0.0030	0.1060	0.0021	0.0022
p	0.0243	0.0030	0.1060	0.0021	0.0022
i	0.0244	0.0030	0.1060	0.0021	0.0022
p,i	0.0204	0.0129	0.1004	0.0020	0.0023
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.0203	0.0128	0.1006	0.0101	0.0113
Empirical	0.0203	0.0131	0.1006	0.0101	0.0114
p	0.0203	0.0131	0.1006	0.0101	0.0114
i	0.0203	0.0131	0.1005	0.0101	0.0114
p,i	0.0203	0.0131	0.1006	0.0101	0.0114
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.0204	0.0129	0.1004	0.0020	0.0023
Empirical	0.0204	0.0127	0.1001	0.0020	0.0023
p	0.0204	0.0127	0.1001	0.0020	0.0023
i	0.0204	0.0127	0.1001	0.0020	0.0023
p,i	0.0204	0.0127	0.1001	0.0020	0.0023

Table F3. Estimated Variance Components for the $p \times i$ Design, Polytomous Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.3500	0.0500	0.6000	0.0600	0.0650
Empirical	0.3516	0.0504	0.5988	0.0599	0.0649
p	0.3516	0.0504	0.5988	0.0599	0.0649
i	0.3516	0.0504	0.5988	0.0599	0.0649
p,i	0.3517	0.0504	0.5988	0.0599	0.0649
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.3500	0.0500	0.6000	0.0120	0.0130
Empirical	0.3485	0.0491	0.6007	0.0120	0.0130
p	0.3485	0.0491	0.6007	0.0120	0.0130
i	0.3485	0.0492	0.6007	0.0120	0.0130
p,i	0.3485	0.0491	0.6007	0.0120	0.0130
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.3000	0.2000	0.5000	0.0500	0.0700
Empirical	0.2985	0.2008	0.4999	0.0500	0.0701
p	0.2985	0.2009	0.4999	0.0500	0.0701
i	0.2985	0.2008	0.4999	0.0500	0.0701
p,i	0.2986	0.2009	0.4999	0.0500	0.0701
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.3000	0.2000	0.5000	0.0100	0.0140
Empirical	0.3010	0.2016	0.4999	0.0100	0.0140
p	0.3010	0.2016	0.4999	0.0100	0.0140
i	0.3010	0.2017	0.4999	0.0100	0.0140
p,i	0.3010	0.2016	0.4999	0.0100	0.0140

Table F4. Estimated Variance Components for the $p \times i \times h$ Design, Normal Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
	VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)								
Parameter	0.1900	0.0100	0.0400	0.0100	0.2400	0.0100	0.5000	0.1720	0.1950
Empirical	0.1916	0.0102	0.0409	0.0102	0.2400	0.0098	0.4999	0.1720	0.1955
p	0.1917	0.0102	0.0409	0.0102	0.2400	0.0098	0.4999	0.1720	0.1955
i	0.1916	0.0102	0.0409	0.0102	0.2400	0.0098	0.4999	0.1720	0.1955
h	0.1918	0.0103	0.0410	0.0105	0.2399	0.0098	0.4996	0.1720	0.1955
p,i	0.1917	0.0102	0.0410	0.0101	0.2401	0.0098	0.4999	0.1721	0.1956
p,h	0.1918	0.0103	0.0410	0.0103	0.2399	0.0098	0.4998	0.1720	0.1955
i,h	0.1914	0.0102	0.0408	0.0097	0.2402	0.0098	0.5003	0.1721	0.1955
p,i,h	0.1919	0.0103	0.0409	0.0103	0.2398	0.0097	0.4999	0.1719	0.1954
	VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)								
Parameter	0.1900	0.0100	0.0400	0.0100	0.2400	0.0100	0.5000	0.0977	0.1123
Empirical	0.1893	0.0097	0.0387	0.0101	0.2403	0.0101	0.4999	0.0978	0.1120
p	0.1894	0.0097	0.0387	0.0101	0.2403	0.0101	0.5000	0.0978	0.1120
i	0.1893	0.0097	0.0387	0.0101	0.2403	0.0101	0.5000	0.0978	0.1120
h	0.1894	0.0097	0.0388	0.0104	0.2402	0.0101	0.4996	0.0977	0.1120
p,i	0.1893	0.0097	0.0387	0.0101	0.2403	0.0101	0.4999	0.0978	0.1120
p,h	0.1894	0.0097	0.0387	0.0102	0.2403	0.0101	0.4998	0.0978	0.1120
i,h	0.1895	0.0097	0.0387	0.0104	0.2401	0.0101	0.4996	0.0977	0.1119
p,i,h	0.1893	0.0098	0.0387	0.0099	0.2404	0.0101	0.5001	0.0978	0.1120

Table F4. Continued

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
Parameter	0.2000	0.1500	0.0500	0.1000	0.1000	0.0500	0.3500	0.1050	0.1650
Empirical	0.2010	0.1494	0.0519	0.1001	0.1000	0.0502	0.3499	0.1050	0.1659
p	0.2010	0.1494	0.0519	0.1001	0.1000	0.0502	0.3500	0.1050	0.1659
i	0.2010	0.1494	0.0519	0.1001	0.1000	0.0502	0.3499	0.1050	0.1659
h	0.2010	0.1494	0.0520	0.1003	0.1000	0.0501	0.3497	0.1050	0.1659
p,i	0.2011	0.1494	0.0520	0.1000	0.1001	0.0502	0.3499	0.1050	0.1659
p,h	0.2011	0.1494	0.0520	0.1002	0.1000	0.0502	0.3499	0.1050	0.1659
i,h	0.2009	0.1493	0.0517	0.0998	0.1001	0.0503	0.3502	0.1050	0.1658
p,i,h	0.2011	0.1494	0.0518	0.1002	0.0999	0.0502	0.3499	0.1050	0.1658
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
Parameter	0.2000	0.1500	0.0500	0.1000	0.1000	0.0500	0.3500	0.0550	0.0883
Empirical	0.1994	0.1480	0.0479	0.1003	0.1003	0.0504	0.3500	0.0551	0.0876
p	0.1995	0.1480	0.0479	0.1003	0.1002	0.0504	0.3500	0.0551	0.0876
i	0.1994	0.1481	0.0479	0.1003	0.1002	0.0504	0.3500	0.0551	0.0876
h	0.1995	0.1481	0.0480	0.1005	0.1002	0.0503	0.3497	0.0551	0.0876
p,i	0.1994	0.1480	0.0480	0.1003	0.1003	0.0504	0.3499	0.0551	0.0876
p,h	0.1995	0.1481	0.0479	0.1004	0.1003	0.0504	0.3499	0.0551	0.0876
i,h	0.1995	0.1480	0.0479	0.1005	0.1002	0.0504	0.3497	0.0551	0.0875
p,i,h	0.1994	0.1481	0.0479	0.1002	0.1003	0.0504	0.3501	0.0551	0.0876

Table F5. Estimated Variance Components for the $p \times i \times h$ Design, Dichotomous Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
Parameter	0.0123	0.0006	0.0023	0.0008	0.0199	0.0007	0.0971	0.0198	0.0211
Empirical	0.0125	0.0005	0.0025	0.0008	0.0201	0.0006	0.0978	0.0200	0.0214
p	0.0125	0.0005	0.0025	0.0008	0.0201	0.0006	0.0978	0.0200	0.0214
i	0.0126	0.0005	0.0025	0.0008	0.0202	0.0006	0.0978	0.0200	0.0214
h	0.0126	0.0005	0.0025	0.0008	0.0201	0.0006	0.0977	0.0200	0.0214
p,i	0.0125	0.0005	0.0025	0.0008	0.0201	0.0006	0.0978	0.0200	0.0214
p,h	0.0125	0.0005	0.0025	0.0007	0.0202	0.0006	0.0978	0.0200	0.0214
i,h	0.0126	0.0005	0.0025	0.0007	0.0201	0.0006	0.0978	0.0200	0.0214
p,i,h	0.0126	0.0005	0.0025	0.0008	0.0201	0.0006	0.0978	0.0200	0.0214
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
Parameter	0.0121	0.0006	0.0024	0.0008	0.0198	0.0006	0.0966	0.0099	0.0108
Empirical	0.0124	0.0006	0.0023	0.0007	0.0200	0.0006	0.0973	0.0100	0.0108
p	0.0124	0.0006	0.0023	0.0007	0.0200	0.0006	0.0973	0.0100	0.0108
i	0.0124	0.0006	0.0023	0.0007	0.0200	0.0006	0.0973	0.0100	0.0108
h	0.0124	0.0006	0.0023	0.0006	0.0200	0.0006	0.0973	0.0100	0.0108
p,i	0.0124	0.0006	0.0023	0.0007	0.0200	0.0006	0.0973	0.0100	0.0108
p,h	0.0124	0.0006	0.0023	0.0008	0.0200	0.0006	0.0972	0.0100	0.0108
i,h	0.0124	0.0006	0.0023	0.0008	0.0200	0.0006	0.0972	0.0100	0.0108
p,i,h	0.0124	0.0006	0.0023	0.0007	0.0200	0.0006	0.0973	0.0100	0.0108

Table F5. Continued

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
Parameter	0.0130	0.0096	0.0029	0.0105	0.0083	0.0042	0.0855	0.0148	0.0186
Empirical	0.0133	0.0091	0.0032	0.0104	0.0086	0.0040	0.0865	0.0150	0.0188
p	0.0133	0.0091	0.0032	0.0104	0.0086	0.0040	0.0865	0.0150	0.0188
i	0.0133	0.0091	0.0032	0.0104	0.0086	0.0040	0.0865	0.0150	0.0188
h	0.0133	0.0091	0.0032	0.0104	0.0086	0.0040	0.0865	0.0150	0.0188
p,i	0.0133	0.0091	0.0032	0.0104	0.0086	0.0040	0.0865	0.0150	0.0188
p,h	0.0133	0.0091	0.0032	0.0103	0.0086	0.0040	0.0865	0.0150	0.0188
i,h	0.0133	0.0091	0.0032	0.0103	0.0085	0.0040	0.0865	0.0150	0.0188
p,i,h	0.0133	0.0091	0.0032	0.0104	0.0086	0.0040	0.0865	0.0150	0.0188
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
Parameter	0.0127	0.0092	0.0030	0.0103	0.0083	0.0040	0.0849	0.0066	0.0087
Empirical	0.0132	0.0093	0.0030	0.0106	0.0084	0.0041	0.0857	0.0067	0.0088
p	0.0132	0.0093	0.0030	0.0106	0.0084	0.0041	0.0857	0.0067	0.0088
i	0.0132	0.0093	0.0030	0.0106	0.0083	0.0041	0.0857	0.0067	0.0088
h	0.0132	0.0093	0.0030	0.0105	0.0084	0.0041	0.0858	0.0067	0.0088
p,i	0.0132	0.0093	0.0030	0.0106	0.0083	0.0041	0.0857	0.0067	0.0088
p,h	0.0132	0.0093	0.0030	0.0106	0.0083	0.0041	0.0857	0.0067	0.0088
i,h	0.0132	0.0093	0.0030	0.0106	0.0083	0.0041	0.0857	0.0067	0.0088
p,i,h	0.0132	0.0093	0.0030	0.0105	0.0083	0.0041	0.0857	0.0067	0.0088

Table F6. Estimated Variance Components for the $p \times i \times h$ Design, Polytomous Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
Parameter	0.1900	0.0100	0.0400	0.0100	0.2400	0.0100	0.5000	0.1720	0.1950
Empirical	0.1898	0.0106	0.0300	0.0093	0.2399	0.0094	0.5001	0.1718	0.1898
p	0.1898	0.0106	0.0300	0.0093	0.2399	0.0094	0.5001	0.1718	0.1899
i	0.1898	0.0106	0.0300	0.0093	0.2396	0.0094	0.5004	0.1717	0.1897
h	0.1899	0.0106	0.0299	0.0096	0.2397	0.0093	0.4999	0.1718	0.1898
p,i	0.1898	0.0106	0.0299	0.0094	0.2399	0.0094	0.5000	0.1718	0.1898
p,h	0.1898	0.0106	0.0300	0.0094	0.2397	0.0094	0.5001	0.1718	0.1898
i,h	0.1898	0.0105	0.0298	0.0095	0.2399	0.0094	0.5000	0.1718	0.1898
p,i,h	0.1902	0.0105	0.0300	0.0097	0.2395	0.0094	0.4997	0.1717	0.1897
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
Parameter	0.1900	0.0100	0.0400	0.0100	0.2400	0.0100	0.5000	0.0977	0.1123
Empirical	0.1914	0.0083	0.0367	0.0102	0.2403	0.0103	0.5000	0.0978	0.1112
p	0.1914	0.0083	0.0367	0.0101	0.2404	0.0103	0.5000	0.0978	0.1112
i	0.1914	0.0083	0.0367	0.0102	0.2403	0.0104	0.5000	0.0978	0.1112
h	0.1914	0.0083	0.0368	0.0100	0.2403	0.0104	0.5001	0.0978	0.1112
p,i	0.1915	0.0083	0.0367	0.0101	0.2403	0.0104	0.5001	0.0978	0.1112
p,h	0.1913	0.0083	0.0368	0.0101	0.2403	0.0104	0.5001	0.0978	0.1112
i,h	0.1914	0.0084	0.0364	0.0103	0.2403	0.0104	0.4998	0.0978	0.1111
p,i,h	0.1913	0.0084	0.0368	0.0101	0.2404	0.0103	0.5001	0.0978	0.1113

Table F6. Continued

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
	VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)								
Parameter	0.2000	0.1500	0.0500	0.1000	0.1000	0.0500	0.3500	0.1050	0.1650
Empirical	0.2026	0.1480	0.0409	0.0999	0.0996	0.0512	0.3510	0.1049	0.1600
p	0.2026	0.1480	0.0409	0.0999	0.0996	0.0512	0.3510	0.1049	0.1600
i	0.2026	0.1480	0.0409	0.0999	0.0997	0.0511	0.3509	0.1049	0.1600
h	0.2026	0.1479	0.0409	0.0999	0.0995	0.0513	0.3510	0.1048	0.1600
p,i	0.2026	0.1478	0.0409	0.0999	0.0995	0.0511	0.3511	0.1049	0.1600
p,h	0.2026	0.1477	0.0408	0.0996	0.0996	0.0513	0.3512	0.1048	0.1599
i,h	0.2023	0.1481	0.0409	0.0992	0.0999	0.0513	0.3516	0.1050	0.1602
p,i,h	0.2027	0.1482	0.0404	0.1001	0.0994	0.0512	0.3508	0.1048	0.1598
	VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)								
Parameter	0.2000	0.1500	0.0500	0.1000	0.1000	0.0500	0.3500	0.0550	0.0883
Empirical	0.2009	0.1513	0.0464	0.0999	0.0997	0.0493	0.3498	0.0549	0.0871
p	0.2010	0.1513	0.0464	0.0999	0.0997	0.0493	0.3498	0.0549	0.0871
i	0.2009	0.1511	0.0464	0.0999	0.0996	0.0493	0.3498	0.0549	0.0871
h	0.2009	0.1513	0.0464	0.0999	0.0996	0.0493	0.3498	0.0549	0.0871
p,i	0.2009	0.1515	0.0464	0.1000	0.0996	0.0492	0.3498	0.0549	0.0871
p,h	0.2009	0.1514	0.0464	0.1001	0.0996	0.0492	0.3496	0.0549	0.0871
i,h	0.2010	0.1512	0.0462	0.1001	0.0996	0.0493	0.3496	0.0549	0.0870
p,i,h	0.2010	0.1513	0.0463	0.1000	0.0996	0.0493	0.3497	0.0549	0.0871

Table F7. Estimated Variance Components for the $i : p$ Design, Normal Data

	$\sigma^2(p)$	$\sigma^2(i{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.1000	0.9000	0.1800	0.1800
Empirical	0.0987	0.9005	0.1801	0.1801
p	0.0988	0.9004	0.1801	0.1801
i	0.0986	0.9006	0.1801	0.1801
p,i	0.0986	0.9005	0.1801	0.1801
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.1000	0.9000	0.0180	0.0180
Empirical	0.0988	0.9013	0.0180	0.0180
p	0.0988	0.9014	0.0180	0.0180
i	0.0988	0.9013	0.0180	0.0180
p,i	0.0988	0.9013	0.0180	0.0180
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.3000	0.7000	0.1400	0.1400
Empirical	0.2981	0.7004	0.1401	0.1401
p	0.2982	0.7003	0.1401	0.1401
i	0.2979	0.7005	0.1401	0.1401
p,i	0.2979	0.7004	0.1401	0.1401
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.3000	0.7000	0.0140	0.0140
Empirical	0.2969	0.7010	0.0140	0.0140
p	0.2969	0.7011	0.0140	0.0140
i	0.2969	0.7010	0.0140	0.0140
p,i	0.2969	0.7010	0.0140	0.0140

Table F8. Estimated Variance Components for the $i : p$ Design, Dichotomous Data

	$\sigma^2(p)$	$\sigma^2(i{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.0060	0.1273	0.0255	0.0255
Empirical	0.0060	0.1278	0.0256	0.0256
p	0.0060	0.1278	0.0256	0.0256
i	0.0059	0.1278	0.0256	0.0256
p,i	0.0060	0.1278	0.0256	0.0256
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.0062	0.1274	0.0025	0.0025
Empirical	0.0059	0.1273	0.0025	0.0025
p	0.0059	0.1273	0.0025	0.0025
i	0.0059	0.1273	0.0025	0.0025
p,i	0.0059	0.1273	0.0025	0.0025
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.0203	0.1132	0.0226	0.0226
Empirical	0.0201	0.1141	0.0228	0.0228
p	0.0201	0.1141	0.0228	0.0228
i	0.0201	0.1141	0.0228	0.0228
p,i	0.0200	0.1141	0.0228	0.0228
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.0202	0.1132	0.0023	0.0023
Empirical	0.0198	0.1127	0.0023	0.0023
p	0.0198	0.1127	0.0023	0.0023
i	0.0198	0.1127	0.0023	0.0023
p,i	0.0198	0.1127	0.0023	0.0023

Table F9. Estimated Variance Components for the $i : p$ Design, Polytomous Data

	$\sigma^2(p)$	$\sigma^2(i{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.1000	0.9000	0.1800	0.1800
Empirical	0.1003	0.9027	0.1806	0.1806
p	0.1002	0.9028	0.1806	0.1806
i	0.1001	0.9031	0.1806	0.1806
p,i	0.1003	0.9026	0.1805	0.1805
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.1000	0.9000	0.0180	0.0180
Empirical	0.0975	0.8991	0.0180	0.0180
p	0.0975	0.8991	0.0180	0.0180
i	0.0975	0.8991	0.0180	0.0180
p,i	0.0976	0.8991	0.0180	0.0180
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.3000	0.7000	0.1400	0.1400
Empirical	0.2972	0.6965	0.1393	0.1393
p	0.2973	0.6965	0.1393	0.1393
i	0.2972	0.6966	0.1393	0.1393
p,i	0.2972	0.6965	0.1393	0.1393
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.3000	0.7000	0.0140	0.0140
Empirical	0.2964	0.7000	0.0140	0.0140
p	0.2965	0.7000	0.0140	0.0140
i	0.2964	0.7000	0.0140	0.0140
p,i	0.2965	0.7000	0.0140	0.0140

Table F10. Estimated Variance Components for the $i:h:p$ Design, Normal Data

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.2800	0.0200	0.7000	0.0110	0.0110
Empirical	0.2787	0.0200	0.6998	0.0110	0.0110
p	0.2787	0.0200	0.6998	0.0110	0.0110
i	0.2787	0.0200	0.6998	0.0110	0.0110
h	0.2787	0.0200	0.6998	0.0110	0.0110
p,i	0.2805	0.0199	0.7001	0.0110	0.0110
p,h	0.2788	0.0200	0.6998	0.0110	0.0110
i,h	0.2787	0.0200	0.6998	0.0110	0.0110
p,i,h	0.2786	0.0200	0.6998	0.0110	0.0110
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.2800	0.0200	0.7000	0.0800	0.0800
Empirical	0.2804	0.0194	0.6967	0.0794	0.0794
p	0.2804	0.0195	0.6966	0.0794	0.0794
i	0.2803	0.0197	0.6965	0.0795	0.0795
h	0.2806	0.0192	0.6967	0.0793	0.0793
p,i	0.2816	0.0208	0.6990	0.0803	0.0803
p,h	0.2804	0.0195	0.6968	0.0794	0.0794
i,h	0.2806	0.0194	0.6967	0.0794	0.0794
p,i,h	0.2805	0.0195	0.6966	0.0794	0.0794

Table F10. Continued

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.1500	0.1500	0.7000	0.0370	0.0370
Empirical	0.1491	0.1499	0.6998	0.0370	0.0370
p	0.1491	0.1499	0.6998	0.0370	0.0370
i	0.1491	0.1499	0.6998	0.0370	0.0370
h	0.1490	0.1500	0.6998	0.0370	0.0370
p,i	0.1504	0.1495	0.7001	0.0369	0.0369
p,h	0.1491	0.1499	0.6998	0.0370	0.0370
i,h	0.1490	0.1499	0.6998	0.0370	0.0370
p,i,h	0.1490	0.1499	0.6998	0.0370	0.0370
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.1500	0.1500	0.7000	0.1450	0.1450
Empirical	0.1506	0.1499	0.7010	0.1451	0.1451
p	0.1536	0.1499	0.7010	0.1451	0.1451
i	0.1505	0.1501	0.7010	0.1452	0.1452
h	0.1508	0.1497	0.7010	0.1449	0.1449
p,i	0.1549	0.1488	0.7000	0.1444	0.1444
p,h	0.1543	0.1500	0.7009	0.1451	0.1451
i,h	0.1504	0.1503	0.7008	0.1452	0.1452
p,i,h	0.1551	0.1500	0.7009	0.1451	0.1451

Table F11. Estimated Variance Components for the $i : h : p$ Design, Dichotomous Data

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.0187	0.0015	0.1131	0.0014	0.0014
Empirical	0.0186	0.0015	0.1130	0.0014	0.0014
p	0.0186	0.0015	0.1130	0.0014	0.0014
i	0.0186	0.0015	0.1130	0.0014	0.0014
h	0.0186	0.0015	0.1130	0.0014	0.0014
p,i	0.0187	0.0016	0.1129	0.0014	0.0014
p,h	0.0186	0.0015	0.1130	0.0014	0.0014
i,h	0.0186	0.0015	0.1130	0.0014	0.0014
p,i,h	0.0187	0.0015	0.1130	0.0014	0.0014
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.0187	0.0016	0.1133	0.0121	0.0121
Empirical	0.0186	0.0016	0.1129	0.0121	0.0121
p	0.0186	0.0016	0.1130	0.0121	0.0121
i	0.0186	0.0016	0.1129	0.0121	0.0121
h	0.0186	0.0015	0.1129	0.0121	0.0121
p,i	0.0188	0.0019	0.1115	0.0121	0.0121
p,h	0.0186	0.0016	0.1129	0.0121	0.0121
i,h	0.0186	0.0016	0.1129	0.0121	0.0121
p,i,h	0.0186	0.0016	0.1129	0.0121	0.0121

Table F11. Continued

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.0094	0.0108	0.1131	0.0033	0.0033
Empirical	0.0094	0.0108	0.1131	0.0033	0.0033
p	0.0094	0.0108	0.1131	0.0033	0.0033
i	0.0094	0.0108	0.1131	0.0033	0.0033
h	0.0094	0.0108	0.1131	0.0033	0.0033
p,i	0.0095	0.0108	0.1129	0.0033	0.0033
p,h	0.0094	0.0108	0.1131	0.0033	0.0033
i,h	0.0094	0.0108	0.1131	0.0033	0.0033
p,i,h	0.0094	0.0108	0.1132	0.0033	0.0033
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.0094	0.0107	0.1134	0.0167	0.0167
Empirical	0.0095	0.0106	0.1133	0.0166	0.0166
p	0.0095	0.0106	0.1133	0.0166	0.0166
i	0.0095	0.0106	0.1133	0.0166	0.0166
h	0.0096	0.0105	0.1133	0.0166	0.0166
p,i	0.0098	0.0109	0.1120	0.0166	0.0166
p,h	0.0095	0.0106	0.1133	0.0166	0.0166
i,h	0.0096	0.0106	0.1133	0.0166	0.0166
p,i,h	0.0095	0.0106	0.1133	0.0166	0.0166

Table F12. Estimated Variance Components for the $i : h : p$ Design, Polytomous Data

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.2800	0.0200	0.7000	0.0110	0.0110
Empirical	0.2770	0.0200	0.6999	0.0110	0.0110
p	0.2770	0.0200	0.6999	0.0110	0.0110
i	0.2770	0.0201	0.6999	0.0110	0.0110
h	0.2770	0.0200	0.7000	0.0110	0.0110
p,i	0.2801	0.0200	0.6997	0.0110	0.0110
p,h	0.2770	0.0201	0.6999	0.0110	0.0110
i,h	0.2769	0.0201	0.6999	0.0110	0.0110
p,i,h	0.2770	0.0200	0.6999	0.0110	0.0110
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.2800	0.0200	0.7000	0.0800	0.0800
Empirical	0.2771	0.0210	0.6996	0.0804	0.0804
p	0.2771	0.0210	0.6997	0.0804	0.0804
i	0.2768	0.0212	0.6997	0.0806	0.0806
h	0.2771	0.0210	0.6996	0.0805	0.0805
p,i	0.2749	0.0223	0.7006	0.0812	0.0812
p,h	0.2771	0.0210	0.6997	0.0805	0.0805
i,h	0.2772	0.0210	0.6996	0.0805	0.0805
p,i,h	0.2770	0.0210	0.6995	0.0805	0.0805

Table F12. Continued

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.1500	0.1500	0.7000	0.0370	0.0370
Empirical	0.1502	0.1499	0.6999	0.0370	0.0370
p	0.1503	0.1499	0.6999	0.0370	0.0370
i	0.1503	0.1499	0.6999	0.0370	0.0370
h	0.1502	0.1500	0.6999	0.0370	0.0370
p,i	0.1484	0.1507	0.7004	0.0372	0.0372
p,h	0.1503	0.1499	0.6999	0.0370	0.0370
i,h	0.1503	0.1499	0.6999	0.0370	0.0370
p,i,h	0.1502	0.1499	0.6999	0.0370	0.0370
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.1500	0.1500	0.7000	0.1450	0.1450
Empirical	0.1544	0.1489	0.7001	0.1445	0.1445
p	0.1544	0.1489	0.7002	0.1445	0.1445
i	0.1547	0.1485	0.7002	0.1443	0.1443
h	0.1544	0.1489	0.7001	0.1445	0.1445
p,i	0.1483	0.1496	0.7002	0.1448	0.1448
p,h	0.1543	0.1490	0.7000	0.1445	0.1445
i,h	0.1546	0.1488	0.7001	0.1444	0.1444
p,i,h	0.1546	0.1489	0.7001	0.1444	0.1444

Table F13. Estimated Variance Components for the $p \times (i:h)$ Design, Normal Data

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i{:}h)$	$\sigma^2(pi{:}h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.2000	0.0100	0.0500	0.0400	0.7000	0.0193	0.0219
Empirical	0.2000	0.0099	0.0500	0.0401	0.7003	0.0193	0.0218
p	0.2000	0.0099	0.0500	0.0401	0.7003	0.0193	0.0218
i	0.2000	0.0099	0.0500	0.0401	0.7003	0.0193	0.0218
h	0.2000	0.0098	0.0500	0.0401	0.7003	0.0193	0.0218
p,i	0.1990	0.0099	0.0500	0.0401	0.7003	0.0193	0.0218
p,h	0.2000	0.0099	0.0500	0.0401	0.7003	0.0193	0.0218
i,h	0.2000	0.0099	0.0500	0.0401	0.7003	0.0193	0.0218
p,i,h	0.1999	0.0099	0.0500	0.0401	0.7003	0.0193	0.0218
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.2000	0.0100	0.0500	0.0400	0.7000	0.0283	0.0307
Empirical	0.1996	0.0093	0.0501	0.0402	0.7007	0.0284	0.0306
p	0.1996	0.0093	0.0501	0.0402	0.7007	0.0284	0.0306
i	0.1996	0.0093	0.0500	0.0403	0.7008	0.0284	0.0306
h	0.1996	0.0093	0.0500	0.0403	0.7007	0.0284	0.0306
p,i	0.1977	0.0093	0.0498	0.0403	0.7009	0.0283	0.0306
p,h	0.1997	0.0093	0.0501	0.0402	0.7007	0.0284	0.0306
i,h	0.1996	0.0093	0.0498	0.0402	0.7010	0.0283	0.0306
p,i,h	0.1996	0.0093	0.0500	0.0402	0.7008	0.0284	0.0306

Table F13. Continued

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i{:}h)$	$\sigma^2(pi{:}h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.1000	0.0500	0.1500	0.1500	0.5500	0.0373	0.0493
Empirical	0.0999	0.0492	0.1501	0.1503	0.5502	0.0374	0.0492
p	0.0999	0.0492	0.1501	0.1503	0.5502	0.0374	0.0492
i	0.0999	0.0492	0.1501	0.1503	0.5502	0.0374	0.0492
h	0.0999	0.0492	0.1501	0.1503	0.5502	0.0374	0.0492
p,i	0.0994	0.0492	0.1501	0.1503	0.5503	0.0374	0.0492
p,h	0.0999	0.0492	0.1501	0.1503	0.5502	0.0374	0.0492
i,h	0.0999	0.0492	0.1501	0.1503	0.5502	0.0374	0.0492
p,i,h	0.0998	0.0492	0.1501	0.1503	0.5503	0.0374	0.0492
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.1000	0.0500	0.1500	0.1500	0.5500	0.0333	0.0433
Empirical	0.0997	0.0478	0.1501	0.1507	0.5506	0.0334	0.0432
p	0.0996	0.0478	0.1501	0.1507	0.5506	0.0334	0.0432
i	0.0997	0.0478	0.1500	0.1507	0.5506	0.0334	0.0432
h	0.0997	0.0478	0.1500	0.1507	0.5506	0.0334	0.0432
p,i	0.0987	0.0478	0.1499	0.1508	0.5507	0.0333	0.0432
p,h	0.0997	0.0478	0.1501	0.1507	0.5506	0.0334	0.0432
i,h	0.0997	0.0478	0.1499	0.1507	0.5508	0.0333	0.0431
p,i,h	0.0997	0.0478	0.1500	0.1507	0.5506	0.0334	0.0432

Table F14. Estimated Variance Components for the $p \times (i : h)$ Design, Dichotomous Data

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i{:}h)$	$\sigma^2(pi{:}h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.0128	0.0006	0.0038	0.0024	0.1136	0.0023	0.0024
Empirical	0.0128	0.0006	0.0038	0.0024	0.1137	0.0023	0.0024
p	0.0128	0.0006	0.0038	0.0024	0.1137	0.0023	0.0024
i	0.0128	0.0006	0.0038	0.0024	0.1137	0.0023	0.0024
h	0.0128	0.0006	0.0038	0.0024	0.1137	0.0023	0.0024
p,i	0.0127	0.0006	0.0038	0.0024	0.1137	0.0023	0.0024
p,h	0.0128	0.0006	0.0038	0.0024	0.1137	0.0023	0.0024
i,h	0.0128	0.0006	0.0038	0.0024	0.1137	0.0023	0.0024
p,i,h	0.0128	0.0006	0.0038	0.0024	0.1137	0.0023	0.0024
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.0130	0.0006	0.0037	0.0024	0.1140	0.0042	0.0043
Empirical	0.0128	0.0006	0.0039	0.0024	0.1134	0.0042	0.0043
p	0.0128	0.0006	0.0039	0.0024	0.1134	0.0042	0.0043
i	0.0128	0.0006	0.0039	0.0024	0.1135	0.0042	0.0043
h	0.0128	0.0006	0.0039	0.0024	0.1134	0.0042	0.0043
p,i	0.0127	0.0005	0.0039	0.0024	0.1135	0.0042	0.0043
p,h	0.0128	0.0006	0.0039	0.0024	0.1134	0.0042	0.0043
i,h	0.0128	0.0006	0.0039	0.0024	0.1135	0.0042	0.0043
p,i,h	0.0128	0.0005	0.0039	0.0024	0.1135	0.0042	0.0043

Table F14. Continued

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i{:}h)$	$\sigma^2(pi{:}h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.0061	0.0030	0.0111	0.0098	0.1030	0.0036	0.0043
Empirical	0.0060	0.0029	0.0111	0.0098	0.1028	0.0036	0.0043
p	0.0060	0.0029	0.0111	0.0098	0.1028	0.0036	0.0043
i	0.0060	0.0029	0.0111	0.0098	0.1028	0.0036	0.0043
h	0.0060	0.0029	0.0111	0.0098	0.1028	0.0036	0.0043
p,i	0.0060	0.0029	0.0111	0.0098	0.1028	0.0036	0.0043
p,h	0.0060	0.0029	0.0111	0.0098	0.1028	0.0036	0.0043
i,h	0.0060	0.0029	0.0111	0.0098	0.1028	0.0036	0.0043
p,i,h	0.0060	0.0029	0.0111	0.0098	0.1028	0.0036	0.0043
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.0062	0.0030	0.0111	0.0099	0.1034	0.0046	0.0052
Empirical	0.0061	0.0029	0.0113	0.0096	0.1026	0.0045	0.0052
p	0.0061	0.0029	0.0113	0.0096	0.1026	0.0045	0.0052
i	0.0061	0.0029	0.0113	0.0096	0.1026	0.0045	0.0052
h	0.0061	0.0029	0.0113	0.0096	0.1026	0.0045	0.0052
p,i	0.0060	0.0029	0.0112	0.0096	0.1026	0.0045	0.0052
p,h	0.0061	0.0029	0.0113	0.0096	0.1026	0.0045	0.0052
i,h	0.0061	0.0029	0.0112	0.0096	0.1027	0.0045	0.0052
p,i,h	0.0061	0.0029	0.0113	0.0096	0.1026	0.0045	0.0052

Table F15. Estimated Variance Components for the $p \times (i:h)$ Design, Polytomous Data

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i{:}h)$	$\sigma^2(pi{:}h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.2000	0.0100	0.0500	0.0400	0.7000	0.0193	0.0219
Empirical	0.1999	0.0093	0.0502	0.0390	0.6996	0.0194	0.0217
p	0.1999	0.0093	0.0502	0.0390	0.6996	0.0194	0.0217
i	0.1999	0.0093	0.0502	0.0390	0.6996	0.0194	0.0217
h	0.1999	0.0093	0.0502	0.0390	0.6996	0.0194	0.0217
p,i	0.1989	0.0093	0.0502	0.0390	0.6996	0.0194	0.0217
p,h	0.1999	0.0092	0.0502	0.0390	0.6996	0.0194	0.0217
i,h	0.1999	0.0092	0.0502	0.0391	0.6996	0.0194	0.0217
p,i,h	0.1998	0.0093	0.0502	0.0390	0.6996	0.0194	0.0218
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.2000	0.0100	0.0500	0.0400	0.7000	0.0283	0.0307
Empirical	0.1988	0.0090	0.0505	0.0421	0.6996	0.0284	0.0307
p	0.1988	0.0090	0.0505	0.0421	0.6996	0.0284	0.0307
i	0.1987	0.0091	0.0512	0.0421	0.6989	0.0284	0.0307
h	0.1988	0.0091	0.0504	0.0421	0.6996	0.0284	0.0307
p,i	0.1967	0.0091	0.0508	0.0421	0.6993	0.0284	0.0307
p,h	0.1988	0.0090	0.0504	0.0420	0.6996	0.0284	0.0307
i,h	0.1988	0.0090	0.0505	0.0420	0.6995	0.0284	0.0307
p,i,h	0.1988	0.0090	0.0504	0.0421	0.6996	0.0284	0.0307

Table F15. Continued

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i{:}h)$	$\sigma^2(pi{:}h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.1000	0.0500	0.1500	0.1500	0.5500	0.0373	0.0493
Empirical	0.1001	0.0520	0.1502	0.1518	0.5501	0.0374	0.0498
p	0.1001	0.0520	0.1502	0.1518	0.5501	0.0374	0.0498
i	0.1001	0.0520	0.1502	0.1518	0.5501	0.0374	0.0498
h	0.1001	0.0520	0.1502	0.1519	0.5501	0.0374	0.0498
p,i	0.0996	0.0520	0.1502	0.1519	0.5501	0.0374	0.0498
p,h	0.1001	0.0520	0.1502	0.1518	0.5501	0.0374	0.0498
i,h	0.1001	0.0520	0.1503	0.1519	0.5501	0.0374	0.0498
p,i,h	0.1001	0.0519	0.1503	0.1518	0.5501	0.0374	0.0498
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.1000	0.0500	0.1500	0.1500	0.5500	0.0333	0.0433
Empirical	0.0993	0.0531	0.1509	0.1498	0.5492	0.0334	0.0437
p	0.0993	0.0531	0.1510	0.1498	0.5492	0.0334	0.0437
i	0.0993	0.0532	0.1512	0.1497	0.5490	0.0334	0.0437
h	0.0993	0.0532	0.1510	0.1498	0.5492	0.0334	0.0437
p,i	0.0982	0.0531	0.1510	0.1499	0.5491	0.0334	0.0437
p,h	0.0992	0.0530	0.1509	0.1499	0.5492	0.0334	0.0437
i,h	0.0993	0.0530	0.1511	0.1498	0.5490	0.0334	0.0437
p,i,h	0.0992	0.0531	0.1508	0.1499	0.5493	0.0334	0.0437

APPENDIX G: ESTIMATED STANDARD ERRORS OF ESTIMATED VARIANCE COMPONENTS

Table G1. Estimated Standard Errors of Estimated Variance Components for the $p \times i$ design, Normal Data

	SE(*p*)	*SE*(*i*)	*SE*(*pi*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.0412	0.0250	0.0201	0.0020	0.0032
p	**0.0412**	0.0086	**0.0200**	**0.0020**	0.0022
i	0.0373	**0.0245**	0.0354	0.0035	0.0045
p,i	0.0598	0.0273	0.0449	0.0045	0.0054
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.0515	0.0113	0.0122	0.0002	0.0003
p	**0.0514**	0.0054	**0.0121**	**0.0002**	0.0003
i	0.0135	**0.0112**	0.0123	0.0002	0.0003
p,i	0.0547	0.0135	0.0211	0.0004	0.0005
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.0351	0.0955	0.0167	0.0017	0.0097
p	**0.0351**	0.0152	**0.0167**	**0.0017**	0.0022
i	0.0312	**0.0943**	0.0295	0.0029	0.0101
p,i	0.0505	0.0968	0.0374	0.0037	0.0106
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.0441	0.0414	0.0102	0.0002	0.0009
p	**0.0440**	0.0091	**0.0101**	**0.0002**	0.0003
i	0.0114	**0.0411**	0.0102	0.0002	0.0008
p,i	0.0467	0.0430	0.0175	0.0004	0.0009

Table G2. Estimated Standard Errors of Estimated Variance Components for the $p \times i$ design, Dichotomous Data

	$SE(p)$	$SE(i)$	$SE(pi)$	$SE(\delta)$	$SE(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.0055	0.0018	0.0108	0.0011	0.0011
p	**0.0047**	0.0010	**0.0068**	**0.0007**	0.0007
i	0.0074	**0.0017**	0.0099	0.0010	0.0010
p,i	0.0094	0.0022	0.0127	0.0013	0.0013
Normal Eqs.	0.0035	0.0017	0.0036	0.0004	0.0004
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.0056	0.0010	0.0087	0.0002	0.0002
p	**0.0053**	0.0008	**0.0078**	**0.0002**	0.0002
i	0.0024	**0.0009**	0.0045	0.0001	0.0001
p,i	0.0061	0.0014	0.0093	0.0002	0.0002
Normal Eqs.	0.0038	0.0008	0.0022	0.0000	0.0000
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.0065	0.0082	0.0165	0.0016	0.0021
p	**0.0039**	0.0022	**0.0058**	**0.0006**	0.0007
i	0.0091	**0.0076**	0.0148	0.0015	0.0019
p,i	0.0104	0.0082	0.0163	0.0016	0.0021
Normal Eqs.	0.0031	0.0068	0.0034	0.0003	0.0008
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.0051	0.0039	0.0095	0.0002	0.0002
p	**0.0044**	0.0019	**0.0065**	**0.0001**	0.0002
i	0.0029	**0.0036**	0.0071	0.0001	0.0002
p,i	0.0055	0.0042	0.0098	0.0002	0.0002
Normal Eqs.	0.0032	0.0028	0.0020	0.0000	0.0001

Table G3. Estimated Standard Errors of Estimated Variance Components for the $p \times i$ design, Polytomous Data

	SE(*p*)	*SE*(*i*)	*SE*(*pi*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.0511	0.0734	0.0184	0.0018	0.0076
p	**0.0502**	0.0085	**0.0184**	**0.0018**	0.0020
i	0.0371	**0.0618**	0.0347	0.0035	0.0072
p,i	0.0663	0.0630	0.0432	0.0043	0.0077
Normal Eqs.	0.0414	0.0380	0.0200	0.0020	0.0043
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.0657	0.0321	0.0109	0.0002	0.0007
p	**0.0643**	0.0054	**0.0112**	**0.0002**	0.0002
i	0.0135	**0.0307**	0.0113	0.0002	0.0007
p,i	0.0672	0.0316	0.0194	0.0004	0.0007
Normal Eqs.	0.0516	0.0125	0.0122	0.0002	0.0003
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
Parameter	0.0431	0.1271	0.0125	0.0012	0.0128
p	**0.0424**	0.0150	**0.0125**	**0.0012**	0.0019
i	0.0311	**0.1171**	0.0279	0.0028	0.0122
p,i	0.0557	0.1190	0.0331	0.0033	0.0125
Normal Eqs.	0.0351	0.1028	0.0167	0.0017	0.0104
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
Parameter	0.0560	0.0566	0.0073	0.0001	0.0011
p	**0.0560**	0.0092	**0.0074**	**0.0001**	0.0002
i	0.0114	**0.0570**	0.0076	0.0002	0.0011
p,i	0.0582	0.0585	0.0129	0.0003	0.0012
Normal Eqs.	0.0445	0.0425	0.0101	0.0002	0.0009

Table G4. Estimated Standard Errors of Estimated Variance Components for the $p \times i \times h$ Design, Normal Data

	SE(*p*)	*SE*(*i*)	*SE*(*h*)	*SE*(*pi*)	*SE*(*ph*)	*SE*(*ih*)	*SE*(*pih*)	*SE*(*δ*)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
Parameter	0.0403	0.0123	0.0618	0.0181	0.0345	0.0088	0.0251	0.0174	0.0355
p	**0.0404**	0.0055	0.0176	**0.0181**	**0.0344**	0.0058	**0.0251**	**0.0174**	0.0194
i	0.0230	**0.0116**	0.0189	0.0167	0.0834	**0.0081**	0.0824	0.0341	0.0356
h	0.2436	0.0148	**0.0778**	0.5007	0.2423	0.0131	0.5006	0.0722	0.0970
p,i	0.0519	0.0138	0.0277	0.0297	0.0942	0.0113	0.0893	0.0406	0.0431
p,h	0.2506	0.0173	0.0816	0.5019	0.2472	0.0155	0.5019	0.0764	0.1009
i,h	0.2590	0.0214	0.0822	0.5078	0.2695	0.0174	0.5139	0.0940	0.1154
p,i,h	0.2684	0.0246	0.0870	0.5100	0.2764	0.0207	0.5164	0.0992	0.1205
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
Parameter	0.0420	0.0073	0.0439	0.0101	0.0292	0.0050	0.0168	0.0098	0.0176
p	**0.0418**	0.0038	0.0159	**0.0100**	**0.0290**	0.0045	**0.0167**	**0.0097**	0.0109
i	0.0150	**0.0068**	0.0108	0.0096	0.0302	**0.0049**	0.0297	0.0094	0.0101
h	0.1041	0.0072	**0.0416**	0.2051	0.1018	0.0061	0.2050	0.0211	0.0287
p,i	0.0469	0.0085	0.0206	0.0169	0.0454	0.0078	0.0376	0.0147	0.0161
p,h	0.1172	0.0094	0.0469	0.2055	0.1090	0.0086	0.2060	0.0249	0.0323
i,h	0.1072	0.0114	0.0442	0.2059	0.1078	0.0090	0.2079	0.0241	0.0314
p,i,h	0.1222	0.0141	0.0500	0.2073	0.1170	0.0122	0.2100	0.0286	0.0357

Table G4. Continued

	$SE(p)$	$SE(i)$	$SE(h)$	$SE(pi)$	$SE(ph)$	$SE(ih)$	$SE(pih)$	$SE(\delta)$	$SE(\Delta)$
	VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)								
Parameter	0.0319	0.1261	0.0864	0.0163	0.0174	0.0366	0.0175	0.0091	0.0507
p	**0.0320**	0.0163	0.0148	**0.0163**	**0.0174**	0.0096	**0.0176**	**0.0091**	0.0121
i	0.0252	**0.1183**	0.0488	0.0218	0.0578	**0.0348**	0.0577	0.0209	0.0384
h	0.1027	0.0776	**0.1088**	0.3506	0.1014	0.0625	0.3504	0.0177	0.0617
p,i	0.0453	0.1205	0.0525	0.0311	0.0626	0.0372	0.0625	0.0244	0.0417
p,h	0.1102	0.0805	0.1108	0.3516	0.1044	0.0640	0.3513	0.0220	0.0641
i,h	0.1212	0.1520	0.1288	0.3561	0.1302	0.0795	0.3597	0.0440	0.0868
p,i,h	0.1308	0.1549	0.1315	0.3579	0.1345	0.0817	0.3615	0.0477	0.0897
	VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)								
Parameter	0.0365	0.0798	0.0564	0.0110	0.0136	0.0178	0.0117	0.0047	0.0209
p	**0.0363**	0.0128	0.0121	**0.0109**	**0.0136**	0.0065	**0.0117**	**0.0046**	0.0062
i	0.0158	**0.0737**	0.0237	0.0115	0.0201	**0.0176**	0.0208	0.0059	0.0122
h	0.0456	0.0381	**0.0533**	0.1437	0.0428	0.0263	0.1435	0.0060	0.0197
p,i	0.0424	0.0758	0.0276	0.0190	0.0263	0.0198	0.0264	0.0083	0.0144
p,h	0.0616	0.0414	0.0558	0.1443	0.0466	0.0277	0.1442	0.0087	0.0214
i,h	0.0504	0.0887	0.0623	0.1447	0.0487	0.0353	0.1455	0.0094	0.0249
p,i,h	0.0681	0.0912	0.0651	0.1463	0.0538	0.0373	0.1470	0.0123	0.0269

Table G5. Estimated Standard Errors of Estimated Variance Components for the $p \times i \times h$ Design, Dichotomous Data

	SE(*p*)	*SE*(*i*)	*SE*(*h*)	*SE*(*pi*)	*SE*(*ph*)	*SE*(*ih*)	*SE*(*pih*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
Parameter	0.0058	0.0009	0.0040	0.0038	0.0067	0.0009	0.0177	0.0047	0.0055
p	**0.0046**	0.0007	0.0016	**0.0039**	**0.0050**	0.0008	**0.0074**	**0.0026**	0.0028
i	0.0041	**0.0008**	0.0016	0.0036	0.0168	**0.0008**	0.0170	0.0070	0.0071
h	0.0221	0.0012	**0.0049**	0.1000	0.0212	0.0011	0.0993	0.0037	0.0047
p,i	0.0072	0.0013	0.0025	0.0064	0.0183	0.0014	0.0194	0.0079	0.0081
p,h	0.0234	0.0017	0.0054	0.1004	0.0224	0.0016	0.0998	0.0053	0.0063
i,h	0.0284	0.0018	0.0054	0.1015	0.0318	0.0015	0.1022	0.0127	0.0132
p,i,h	0.0302	0.0024	0.0061	0.1022	0.0335	0.0022	0.1030	0.0139	0.0144
Normal Eqs.	0.0040	0.0009	0.0042	0.0035	0.0042	0.0008	0.0050	0.0022	0.0030
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
Parameter	0.0049	0.0006	0.0031	0.0021	0.0054	0.0006	0.0146	0.0022	0.0027
p	**0.0041**	0.0005	0.0014	**0.0021**	**0.0040**	0.0007	**0.0070**	**0.0014**	0.0015
i	0.0023	**0.0005**	0.0009	0.0020	0.0060	**0.0005**	0.0070	0.0019	0.0019
h	0.0100	0.0007	**0.0027**	0.0409	0.0092	0.0006	0.0409	0.0017	0.0020
p,i	0.0052	0.0009	0.0019	0.0036	0.0078	0.0011	0.0106	0.0026	0.0027
p,h	0.0116	0.0011	0.0033	0.0413	0.0106	0.0011	0.0418	0.0025	0.0029
i,h	0.0110	0.0010	0.0030	0.0413	0.0115	0.0009	0.0418	0.0028	0.0031
p,i,h	0.0129	0.0015	0.0037	0.0417	0.0133	0.0016	0.0428	0.0037	0.0040
Normal Eqs.	0.0034	0.0006	0.0028	0.0019	0.0031	0.0005	0.0033	0.0010	0.0014

Table G5. Continued

	SE(*p*)	*SE*(*i*)	*SE*(*h*)	*SE*(*pi*)	*SE*(*ph*)	*SE*(*ih*)	*SE*(*pih*)	*SE*(*δ*)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
Parameter	0.0071	0.0108	0.0061	0.0055	0.0045	0.0042	0.0228	0.0046	0.0073
p	**0.0038**	0.0022	0.0016	**0.0042**	**0.0032**	0.0016	**0.0063**	**0.0018**	0.0021
i	0.0062	**0.0088**	0.0045	0.0048	0.0164	**0.0034**	0.0191	0.0073	0.0084
h	0.0111	0.0076	**0.0068**	0.0904	0.0098	0.0057	0.0895	0.0055	0.0059
p,i	0.0082	0.0093	0.0050	0.0074	0.0173	0.0040	0.0207	0.0078	0.0090
p,h	0.0124	0.0082	0.0072	0.0908	0.0108	0.0061	0.0899	0.0061	0.0067
i,h	0.0213	0.0127	0.0093	0.0927	0.0252	0.0074	0.0935	0.0132	0.0146
p,i,h	0.0229	0.0134	0.0098	0.0933	0.0263	0.0080	0.0942	0.0138	0.0153
Normal Eqs.	0.0034	0.0091	0.0062	0.0036	0.0029	0.0035	0.0045	0.0015	0.0039
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
Parameter	0.0057	0.0069	0.0041	0.0037	0.0032	0.0023	0.0173	0.0017	0.0029
p	**0.0037**	0.0018	0.0012	**0.0027**	**0.0022**	0.0012	**0.0057**	**0.0008**	0.0010
i	0.0039	**0.0060**	0.0022	0.0030	0.0058	**0.0020**	0.0113	0.0021	0.0025
h	0.0056	0.0043	**0.0036**	0.0366	0.0042	0.0026	0.0364	0.0017	0.0022
p,i	0.0058	0.0065	0.0027	0.0047	0.0066	0.0026	0.0131	0.0023	0.0029
p,h	0.0072	0.0049	0.0040	0.0371	0.0051	0.0031	0.0372	0.0020	0.0026
i,h	0.0077	0.0080	0.0047	0.0374	0.0078	0.0037	0.0387	0.0029	0.0036
p,i,h	0.0095	0.0086	0.0051	0.0378	0.0089	0.0043	0.0394	0.0033	0.0041
Normal Eqs.	0.0030	0.0057	0.0038	0.0021	0.0018	0.0018	0.0029	0.0006	0.0015

Table G6. Estimated Standard Errors of Estimated Variance Components for the $p \times i \times h$ Design, Polytomous Data

	SE(*p*)	*SE*(*i*)	*SE*(*h*)	*SE*(*pi*)	*SE*(*ph*)	*SE*(*ih*)	*SE*(*pih*)	*SE*(*δ*)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
Parameter	0.0476	0.0441	0.1567	0.0186	0.0303	0.0329	0.0226	0.0153	0.0801
p	**0.0481**	0.0056	0.0155	**0.0182**	**0.0304**	0.0057	**0.0223**	**0.0154**	0.0171
i	0.0232	**0.0357**	0.0240	0.0168	0.0830	**0.0231**	0.0817	0.0339	0.0371
h	0.2427	0.0403	**0.1282**	0.5008	0.2417	0.0319	0.5006	0.0716	0.1060
p,i	0.0581	0.0365	0.0300	0.0300	0.0924	0.0245	0.0876	0.0396	0.0432
p,h	0.2501	0.0413	0.1301	0.5017	0.2454	0.0329	0.5017	0.0749	0.1088
i,h	0.2581	0.0614	0.1326	0.5079	0.2689	0.0458	0.5139	0.0935	0.1237
p,i,h	0.2681	0.0624	0.1350	0.5095	0.2744	0.0471	0.5159	0.0979	0.1276
Normal Eqs.	0.0403	0.0318	0.1060	0.0181	0.0344	0.0189	0.0251	0.0174	0.0561
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
Parameter	0.0549	0.0321	0.1169	0.0105	0.0224	0.0184	0.0141	0.0075	0.0398
p	**0.0563**	0.0037	0.0155	**0.0104**	**0.0225**	0.0045	**0.0140**	**0.0075**	0.0089
i	0.0152	**0.0245**	0.0101	0.0100	0.0303	**0.0159**	0.0285	0.0094	0.0103
h	0.1031	0.0141	**0.0882**	0.2043	0.1008	0.0139	0.2043	0.0207	0.0382
p,i	0.0602	0.0250	0.0198	0.0175	0.0413	0.0171	0.0347	0.0133	0.0149
p,h	0.1217	0.0154	0.0907	0.2054	0.1059	0.0152	0.2055	0.0234	0.0403
i,h	0.1065	0.0307	0.0899	0.2058	0.1071	0.0242	0.2074	0.0238	0.0405
p,i,h	0.1267	0.0318	0.0923	0.2070	0.1140	0.0254	0.2091	0.0272	0.0431
Normal Eqs.	0.0425	0.0139	0.0887	0.0101	0.0292	0.0075	0.0167	0.0098	0.0311

Table G6. Continued

	SE(*p*)	*SE*(*i*)	*SE*(*h*)	*SE*(*pi*)	*SE*(*ph*)	*SE*(*ih*)	*SE*(*pih*)	*SE*(δ)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
Parameter	0.0377	0.1367	0.1715	0.0175	0.0197	0.0763	0.0199	0.0103	0.0903
p	**0.0380**	0.0163	0.0135	**0.0176**	**0.0199**	0.0096	**0.0199**	**0.0105**	0.0128
i	0.0255	**0.1261**	0.0552	0.0224	0.0580	**0.0638**	0.0582	0.0211	0.0420
h	0.1030	0.1192	**0.1508**	0.3518	0.1016	0.0965	0.3515	0.0183	0.0812
p,i	0.0499	0.1279	0.0577	0.0326	0.0637	0.0650	0.0640	0.0252	0.0451
p,h	0.1130	0.1209	0.1520	0.3530	0.1054	0.0974	0.3527	0.0237	0.0831
i,h	0.1216	0.1936	0.1694	0.3574	0.1308	0.1320	0.3610	0.0446	0.1040
p,i,h	0.1334	0.1956	0.1706	0.3596	0.1357	0.1333	0.3631	0.0488	0.1064
Normal Eqs.	0.0322	0.1338	0.1267	0.0164	0.0174	0.0563	0.0176	0.0091	0.0691
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
Parameter	0.0460	0.0887	0.1375	0.0128	0.0181	0.0437	0.0138	0.0062	0.0470
p	**0.0455**	0.0130	0.0118	**0.0125**	**0.0176**	0.0065	**0.0138**	**0.0060**	0.0072
i	0.0156	**0.0845**	0.0239	0.0128	0.0201	**0.0379**	0.0216	0.0059	0.0132
h	0.0461	0.0481	**0.0957**	0.1435	0.0433	0.0384	0.1434	0.0065	0.0331
p,i	0.0504	0.0863	0.0276	0.0214	0.0286	0.0390	0.0286	0.0092	0.0157
p,h	0.0683	0.0505	0.0970	0.1446	0.0490	0.0394	0.1445	0.0102	0.0345
i,h	0.0508	0.1049	0.1011	0.1447	0.0492	0.0611	0.1456	0.0098	0.0369
p,i,h	0.0742	0.1071	0.1028	0.1469	0.0560	0.0622	0.1477	0.0135	0.0386
Normal Eqs.	0.0369	0.0826	0.0965	0.0110	0.0136	0.0218	0.0117	0.0047	0.0335

Table G7. Estimated Standard Errors of Estimated Variance Components for the $i : p$ Design, Normal Data

	SE(*p*)	*SE*(*i*:*p*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.0594	0.0900	0.0180	0.0180
p	**0.0587**	0.0896	**0.0179**	0.0179
i	0.1604	**0.1646**	0.0329	0.0329
p,i	0.1233	0.1503	0.0301	0.0301
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.0238	0.0257	0.0005	0.0005
p	**0.0234**	0.0254	**0.0005**	0.0005
i	0.0140	**0.0258**	0.0005	0.0005
p,i	0.0301	0.0439	0.0009	0.0009
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.0900	0.0700	0.0140	0.0140
p	**0.0887**	0.0697	**0.0139**	0.0139
i	0.1327	**0.1280**	0.0256	0.0256
p,i	0.1376	0.1169	0.0234	0.0234
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.0634	0.0200	0.0004	0.0004
p	**0.0624**	0.0197	**0.0004**	0.0004
i	0.0191	**0.0200**	0.0004	0.0004
p,i	0.0677	0.0341	0.0007	0.0007

Table G8. Estimated Standard Errors of Estimated Variance Components for the $i:p$ Design, Dichotomous Data

	$SE(p)$	$SE(i{:}p)$	$SE(\delta)$	$SE(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.0074	0.0171	0.0034	0.0034
p	**0.0074**	0.0170	**0.0034**	0.0034
i	0.0231	**0.0246**	0.0049	0.0049
p,i	0.0178	0.0263	0.0053	0.0053
Normal Eqs.	0.0069	0.0128	0.0026	0.0026
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.0021	0.0083	0.0002	0.0002
p	**0.0019**	0.0082	**0.0002**	0.0002
i	0.0017	**0.0046**	0.0001	0.0001
p,i	0.0029	0.0104	0.0002	0.0002
Normal Eqs.	0.0017	0.0036	0.0001	0.0001
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.0112	0.0176	0.0035	0.0035
p	**0.0110**	0.0173	**0.0035**	0.0035
i	0.0217	**0.0222**	0.0044	0.0044
p,i	0.0206	0.0254	0.0051	0.0051
Normal Eqs.	0.0091	0.0115	0.0023	0.0023
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.0064	0.0113	0.0002	0.0002
p	**0.0061**	0.0110	**0.0002**	0.0002
i	0.0026	**0.0040**	0.0001	0.0001
p,i	0.0070	0.0124	0.0002	0.0002
Normal Eqs.	0.0046	0.0032	0.0001	0.0001

Table G9. Estimated Standard Errors of Estimated Variance Components for the $i : p$ Design, Polytomous Data

	SE(*p*)	*SE*(*i*:*p*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.0660	0.0840	0.0168	0.0168
p	**0.0646**	0.0823	**0.0165**	0.0165
i	0.1604	**0.1624**	0.0325	0.0325
p,*i*	0.1247	0.1412	0.0282	0.0282
Normal Eqs.	0.0598	0.0902	0.0180	0.0180
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.0370	0.0230	0.0005	0.0005
p	**0.0357**	0.0228	**0.0005**	0.0005
i	0.0139	**0.0233**	0.0005	0.0005
p,*i*	0.0405	0.0396	0.0008	0.0008
Normal Eqs.	0.0240	0.0257	0.0005	0.0005
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
Parameter	0.1037	0.0614	0.0123	0.0123
p	**0.1001**	0.0605	**0.0121**	0.0121
i	0.1315	**0.1249**	0.0250	0.0250
p,*i*	0.1440	0.1055	0.0211	0.0211
Normal Eqs.	0.0896	0.0696	0.0139	0.0139
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
Parameter	0.0823	0.0169	0.0003	0.0003
p	**0.0777**	0.0167	**0.0003**	0.0003
i	0.0190	**0.0172**	0.0003	0.0003
p,*i*	0.0821	0.0291	0.0006	0.0006
Normal Eqs.	0.0635	0.0200	0.0004	0.0004

Table G10. Estimated Standard Errors of Estimated Variance Components for the $i:h:p$ Design, Normal Data

	SE(*p*)	*SE*(*h*:*p*)	*SE*(*i*:*h*:*p*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.0414	0.0039	0.0102	0.0008	0.0110
p	**0.0411**	0.0039	0.0101	**0.0008**	0.0008
i	0.0090	0.0127	**0.0153**	0.0024	0.0024
h	0.0135	**0.0094**	0.0090	0.0019	0.0019
p,i	0.0430	0.0082	0.0174	0.0016	0.0016
p,h	0.0437	0.0065	0.0163	0.0013	0.0013
i,h	0.0214	0.0172	0.0175	0.0034	0.0034
p,i,h	0.0471	0.0118	0.0249	0.0023	0.0023
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.0745	0.0335	0.0495	0.0160	0.0160
p	**0.0735**	0.0329	0.0486	**0.0157**	0.0157
i	0.0483	0.1204	**0.1194**	0.0496	0.0496
h	0.1660	**0.1621**	0.0345	0.0810	0.0810
p,i	0.1004	0.0800	0.0837	0.0373	0.0373
p,h	0.1036	0.0647	0.0689	0.0316	0.0316
i,h	0.2909	0.2962	0.1160	0.1455	0.1455
p,i,h	0.1534	0.1245	0.1060	0.0598	0.0598

Table G10. Continued

	SE(*p*)	*SE*(*h*:*p*)	*SE*(*i*:*h*:*p*)	*SE*(δ)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.0267	0.0131	0.0102	0.0026	0.0026
p	**0.0265**	0.0130	0.0101	**0.0026**	0.0026
i	0.0073	0.0143	**0.0153**	0.0028	0.0028
h	0.0335	**0.0316**	0.0090	0.0063	0.0063
p,i	0.0285	0.0175	0.0174	0.0035	0.0035
p,h	0.0354	0.0218	0.0163	0.0044	0.0044
i,h	0.0397	0.0387	0.0175	0.0077	0.0077
p,i,h	0.0388	0.0271	0.0249	0.0054	0.0054
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.0663	0.0588	0.0495	0.0290	0.0290
p	**0.0471**	0.0414	0.0348	**0.0204**	0.0204
i	0.0338	0.1188	**0.1156**	0.0488	0.0488
h	0.2944	**0.2930**	0.0249	0.1465	0.1465
p,i	0.0671	0.0740	0.0595	0.0356	0.0356
p,h	0.0951	0.0825	0.0493	0.0410	0.0410
i,h	0.4141	0.4196	0.1070	0.2082	0.2082
p,i,h	0.1316	0.1243	0.0757	0.0609	0.0609

Table G11. Estimated Standard Errors of Estimated Variance Components for the $i:h:p$ Design, Dichotomous Data

	$SE(p)$	$SE(h{:}p)$	$SE(i{:}h{:}p)$	$SE(\delta)$	$SE(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.0039	0.0006	0.0074	0.0002	0.0002
p	**0.0040**	0.0006	0.0074	**0.0002**	0.0002
i	0.0012	0.0021	**0.0027**	0.0004	0.0004
h	0.0017	**0.0013**	0.0020	0.0003	0.0003
p,i	0.0042	0.0015	0.0080	0.0003	0.0003
p,h	0.0043	0.0011	0.0079	0.0002	0.0002
i,h	0.0029	0.0026	0.0035	0.0005	0.0005
p,i,h	0.0049	0.0021	0.0088	0.0004	0.0004
Normal Eqs.	0.0029	0.0005	0.0016	0.0001	0.0001
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.0086	0.0063	0.0140	0.0032	0.0032
p	**0.0083**	0.0061	0.0141	**0.0031**	0.0031
i	0.0075	0.0202	**0.0201**	0.0084	0.0084
h	0.0255	**0.0251**	0.0069	0.0125	0.0125
p,i	0.0136	0.0150	0.0194	0.0071	0.0071
p,h	0.0146	0.0118	0.0174	0.0059	0.0059
i,h	0.0461	0.0474	0.0203	0.0232	0.0232
p,i,h	0.0245	0.0235	0.0236	0.0113	0.0113
Normal Eqs.	0.0068	0.0052	0.0080	0.0025	0.0025

Table G11. Continued

	SE(*p*)	*SE*(*h*:*p*)	*SE*(*i*:*h*:*p*)	*SE*(*δ*)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.0022	0.0016	0.0060	0.0003	0.0003
p	**0.0023**	0.0016	0.0060	**0.0003**	0.0003
i	0.0009	0.0022	**0.0027**	0.0004	0.0004
h	0.0031	**0.0029**	0.0028	0.0006	0.0006
p,i	0.0026	0.0024	0.0066	0.0005	0.0005
p,h	0.0032	0.0025	0.0072	0.0005	0.0005
i,h	0.0041	0.0042	0.0041	0.0008	0.0008
p,i,h	0.0038	0.0034	0.0081	0.0007	0.0007
Normal Eqs.	0.0018	0.0012	0.0016	0.0002	0.0002
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.0072	0.0081	0.0133	0.0041	0.0041
p	**0.0071**	0.0080	0.0133	**0.0041**	0.0041
i	0.0070	0.0205	**0.0201**	0.0086	0.0086
h	0.0347	**0.0343**	0.0076	0.0171	0.0171
p,i	0.0122	0.0167	0.0188	0.0080	0.0080
p,h	0.0169	0.0158	0.0174	0.0079	0.0079
i,h	0.0549	0.0562	0.0206	0.0277	0.0277
p,i,h	0.0264	0.0270	0.0236	0.0131	0.0131
Normal Eqs.	0.0063	0.0069	0.0081	0.0034	0.0034

Table G12. Estimated Standard Errors of Estimated Variance Components for the $i:h:p$ Design, Polytomous Data

	SE(*p*)	*SE*(*h*:*p*)	*SE*(*i*:*h*:*p*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.0542	0.0073	0.0086	0.0015	0.0015
p	**0.0531**	0.0073	0.0086	**0.0014**	0.0014
i	0.0089	0.0127	**0.0144**	0.0024	0.0024
h	0.0138	**0.0105**	0.0078	0.0021	0.0021
p,i	0.0550	0.0102	0.0150	0.0020	0.0020
p,h	0.0552	0.0111	0.0139	0.0022	0.0022
i,h	0.0215	0.0178	0.0156	0.0035	0.0035
p,i,h	0.0579	0.0147	0.0214	0.0029	0.0029
Normal Eqs.	0.0413	0.0039	0.0102	0.0008	0.0008
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.0902	0.0363	0.0432	0.0173	0.0173
p	**0.0855**	0.0358	0.0428	**0.0171**	0.0171
i	0.0475	0.1205	**0.1179**	0.0495	0.0495
h	0.1686	**0.1645**	0.0311	0.0822	0.0822
p,i	0.1115	0.0806	0.0765	0.0374	0.0374
p,h	0.1150	0.0689	0.0610	0.0337	0.0337
i,h	0.2932	0.2984	0.1131	0.1467	0.1467
p,i,h	0.1599	0.1248	0.0966	0.0600	0.0600
Normal Eqs.	0.0745	0.0337	0.0494	0.0161	0.0161

Table G12. Continued

	SE(*p*)	*SE*(*h*:*p*)	*SE*(*i*:*h*:*p*)	*SE*(*δ*)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
Parameter	0.0295	0.0142	0.0086	0.0028	0.0028
p	**0.0292**	0.0141	0.0086	**0.0028**	0.0028
i	0.0073	0.0143	**0.0145**	0.0028	0.0028
h	0.0335	**0.0318**	0.0077	0.0064	0.0064
p,i	0.0309	0.0184	0.0150	0.0037	0.0037
p,h	0.0377	0.0232	0.0139	0.0046	0.0046
i,h	0.0399	0.0390	0.0156	0.0078	0.0078
p,i,h	0.0409	0.0283	0.0215	0.0056	0.0056
Normal Eqs.	0.0268	0.0131	0.0102	0.0026	0.0026
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
Parameter	0.0690	0.0599	0.0432	0.0294	0.0294
p	**0.0681**	0.0592	0.0432	**0.0291**	0.0291
i	0.0479	0.1253	**0.1184**	0.0525	0.0525
h	0.2979	**0.2948**	0.0304	0.1474	0.1474
p,i	0.0954	0.1045	0.0763	0.0501	0.0501
p,h	0.1354	0.1172	0.0611	0.0582	0.0582
i,h	0.4197	0.4243	0.1131	0.2104	0.2104
p,i,h	0.1849	0.1737	0.0969	0.0851	0.0851
Normal Eqs.	0.0671	0.0586	0.0495	0.0289	0.0289

Table G13. Estimated Standard Errors of Estimated Variance Components for the $p \times (i:h)$ Design, Normal Data

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.0220	0.0093	0.0049	0.0074	0.0084	0.0010	0.0021
p	**0.0219**	0.0025	**0.0049**	0.0030	**0.0084**	**0.0010**	0.0011
i	0.0063	0.0064	0.0190	**0.0072**	0.0200	0.0036	0.0039
h	0.0175	**0.0090**	0.0159	0.0067	0.0075	0.0032	0.0040
p,i	0.0235	0.0073	0.0203	0.0083	0.0231	0.0038	0.0042
p,h	0.0293	0.0097	0.0173	0.0078	0.0135	0.0034	0.0042
i,h	0.0256	0.0126	0.0258	0.0117	0.0172	0.0051	0.0061
p,i,h	0.0359	0.0134	0.0275	0.0130	0.0233	0.0054	0.0064
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.0325	0.0133	0.0153	0.0149	0.0222	0.0013	0.0018
p	**0.0324**	0.0065	**0.0153**	0.0085	**0.0221**	**0.0013**	0.0014
i	0.0126	0.0225	0.2870	**0.0212**	0.2869	0.0192	0.0203
h	0.0213	**0.0124**	0.0203	0.0139	0.0211	0.0019	0.0023
p,i	0.0366	0.0249	0.2881	0.0241	0.2884	0.0193	0.0204
p,h	0.0420	0.0153	0.0292	0.0182	0.0371	0.0026	0.0030
i,h	0.4468	0.0963	0.2232	0.2006	0.8372	0.0088	0.0095
p,i,h	0.0526	0.0296	0.1333	0.0290	0.1357	0.0093	0.0100

Table G13. Continued

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(*δ*)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.0139	0.0431	0.0094	0.0258	0.0066	0.0019	0.0088
p	**0.0138**	0.0075	**0.0093**	0.0049	**0.0066**	**0.0019**	0.0024
i	0.0045	0.0254	0.0155	**0.0254**	0.0157	0.0029	0.0061
h	0.0312	**0.0418**	0.0308	0.0235	0.0059	0.0062	0.0114
p,i	0.0151	0.0269	0.0191	0.0263	0.0182	0.0037	0.0068
p,h	0.0356	0.0430	0.0333	0.0244	0.0106	0.0067	0.0119
i,h	0.0371	0.0543	0.0377	0.0412	0.0135	0.0075	0.0143
p,i,h	0.0415	0.0555	0.0406	0.0422	0.0183	0.0081	0.0148
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.0190	0.0514	0.0168	0.0492	0.0175	0.0016	0.0051
p	**0.0190**	0.0131	**0.0167**	0.0132	**0.0174**	**0.0016**	0.0020
i	0.0087	0.0815	0.2257	**0.0757**	0.2255	0.0151	0.0197
h	0.0216	**0.0478**	0.0231	0.0462	0.0166	0.0022	0.0052
p,i	0.0216	0.0478	0.0231	0.0462	0.0166	0.0152	0.0199
p,h	0.0319	0.0512	0.0327	0.0497	0.0291	0.0031	0.0059
i,h	0.0303	0.0922	0.1031	0.0816	0.1018	0.0072	0.0119
p,i,h	0.0405	0.0961	0.1086	0.0853	0.1067	0.0079	0.0125

Table G14. Estimated Standard Errors of Estimated Variance Components for the $p \times (i:h)$ Design, Dichotomous Data

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.0023	0.0006	0.0008	0.0006	0.0080	0.0002	0.0003
p	**0.0020**	0.0002	**0.0007**	0.0004	**0.0046**	**0.0002**	0.0002
i	0.0010	0.0004	0.0031	**0.0005**	0.0043	0.0006	0.0006
h	0.0024	**0.0006**	0.0019	0.0005	0.0060	0.0004	0.0004
p,i	0.0023	0.0005	0.0033	0.0007	0.0065	0.0006	0.0007
p,h	0.0032	0.0006	0.0021	0.0007	0.0078	0.0004	0.0005
i,h	0.0037	0.0008	0.0037	0.0009	0.0076	0.0007	0.0008
p,i,h	0.0044	0.0009	0.0040	0.0011	0.0093	0.0008	0.0009
Normal Eqs.	0.0015	0.0006	0.0006	0.0005	0.0014	0.0001	0.0002
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.0034	0.0010	0.0027	0.0012	0.0096	0.0004	0.0004
p	**0.0031**	0.0007	**0.0027**	0.0010	**0.0073**	**0.0003**	0.0003
i	0.0018	0.0015	0.0467	**0.0014**	0.0467	0.0031	0.0032
h	0.0032	**0.0009**	0.0033	0.0011	0.0072	0.0004	0.0004
p,i	0.0039	0.0019	0.0470	0.0020	0.0475	0.0032	0.0032
p,h	0.0049	0.0013	0.0050	0.0018	0.0111	0.0005	0.0006
i,h	0.0052	0.0019	0.0216	0.0019	0.0227	0.0015	0.0016
p,i,h	0.0069	0.0025	0.0226	0.0026	0.0248	0.0016	0.0017
Normal Eqs.	0.0024	0.0009	0.0023	0.0011	0.0036	0.0002	0.0002

Table G14. Continued

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(δ)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.0017	0.0029	0.0023	0.0029	0.0126	0.0006	0.0010
p	**0.0011**	0.0006	**0.0012**	0.0008	**0.0033**	**0.0002**	0.0003
i	0.0008	0.0018	0.0031	**0.0022**	0.0056	0.0006	0.0008
h	0.0036	**0.0027**	0.0034	0.0026	0.0112	0.0007	0.0011
p,i	0.0015	0.0020	0.0035	0.0024	0.0066	0.0007	0.0009
p,h	0.0039	0.0028	0.0037	0.0028	0.0118	0.0008	0.0011
i,h	0.0048	0.0036	0.0050	0.0039	0.0130	0.0010	0.0015
p,i,h	0.0052	0.0038	0.0054	0.0041	0.0137	0.0011	0.0015
Normal Eqs.	0.0010	0.0027	0.0009	0.0018	0.0012	0.0002	0.0006
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.0022	0.0038	0.0035	0.0044	0.0126	0.0006	0.0009
p	**0.0017**	0.0014	**0.0029**	0.0018	**0.0054**	**0.0003**	0.0004
i	0.0013	0.0057	0.0428	**0.0053**	0.0426	0.0029	0.0032
h	0.0034	**0.0035**	0.0038	0.0040	0.0114	0.0006	0.0008
p,i	0.0025	0.0062	0.0430	0.0058	0.0431	0.0029	0.0032
p,h	0.0042	0.0040	0.0055	0.0047	0.0132	0.0007	0.0009
i,h	0.0053	0.0069	0.0215	0.0065	0.0239	0.0017	0.0019
p,i,h	0.0061	0.0076	0.0224	0.0073	0.0253	0.0018	0.0021
Normal Eqs.	0.0015	0.0034	0.0024	0.0035	0.0033	0.0002	0.0004

Table G15. Estimated Standard Errors of Estimated Variance Components for the $p \times (i:h)$ Design, Polytomous Data

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(*δ*)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.0294	0.0414	0.0083	0.0231	0.0071	0.0017	0.0084
p	**0.0296**	0.0025	**0.0084**	0.0030	**0.0072**	**0.0017**	0.0017
i	0.0064	0.0069	0.0190	**0.0209**	0.0196	0.0036	0.0040
h	0.0177	**0.0325**	0.0168	0.0204	0.0064	0.0033	0.0074
p,i	0.0308	0.0076	0.0214	0.0213	0.0220	0.0041	0.0045
p,h	0.0356	0.0325	0.0199	0.0208	0.0116	0.0040	0.0078
i,h	0.0257	0.0337	0.0263	0.0348	0.0159	0.0052	0.0088
p,i,h	0.0412	0.0340	0.0293	0.0353	0.0209	0.0058	0.0092
Normal Eqs.	0.0221	0.0261	0.0049	0.0080	0.0084	0.0010	0.0053
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.0438	0.0323	0.0168	0.0379	0.0199	0.0015	0.0037
p	**0.0420**	0.0065	**0.0167**	0.0086	**0.0200**	**0.0015**	0.0015
i	0.0123	0.0348	0.2866	**0.0334**	0.2864	0.0015	0.0205
h	0.0214	**0.0302**	0.0211	0.0369	0.0190	0.0191	0.0038
p,i	0.0450	0.0365	0.2880	0.0354	0.2879	0.0020	0.0206
p,h	0.0497	0.0315	0.0311	0.0388	0.0335	0.0193	0.0043
i,h	0.0334	0.0510	0.1278	0.0549	0.1286	0.0028	0.0104
p,i,h	0.0589	0.0532	0.1331	0.0571	0.1343	0.0087	0.0109
Normal Eqs.	0.0326	0.0199	0.0153	0.0188	0.0222	0.0013	0.0023

Table G15. Continued

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(*δ*)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
Parameter	0.0201	0.1079	0.0102	0.0294	0.0063	0.0020	0.0217
p	**0.0198**	0.0077	**0.0101**	0.0049	**0.0063**	**0.0020**	0.0025
i	0.0045	0.0261	0.0156	**0.0284**	0.0156	0.0029	0.0063
h	0.0314	**0.0836**	0.0310	0.0267	0.0057	0.0062	0.0185
p,i	0.0207	0.0276	0.0195	0.0291	0.0179	0.0038	0.0069
p,h	0.0385	0.0840	0.0338	0.0275	0.0102	0.0068	0.0187
i,h	0.0372	0.0910	0.0379	0.0461	0.0132	0.0076	0.0205
p,i,h	0.0440	0.0917	0.0411	0.0470	0.0177	0.0082	0.0209
Normal Eqs.	0.0140	0.0710	0.0094	0.0262	0.0066	0.0019	0.0143
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
Parameter	0.0280	0.0851	0.0172	0.0541	0.0172	0.0016	0.0085
p	**0.0273**	0.0134	**0.0172**	0.0132	**0.0169**	**0.0016**	0.0020
i	0.0084	0.0836	0.2249	**0.0770**	0.2247	0.0150	0.0197
h	0.0216	**0.0789**	0.0234	0.0514	0.0160	0.0023	0.0082
p,i	0.0295	0.0861	0.2265	0.0791	0.2261	0.0152	0.0200
p,h	0.0375	0.0811	0.0333	0.0545	0.0281	0.0032	0.0087
i,h	0.0302	0.1140	0.1028	0.0873	0.1013	0.0072	0.0136
p,i,h	0.0449	0.1172	0.1084	0.0908	0.1060	0.0078	0.0142
Normal Eqs.	0.0192	0.0611	0.0169	0.0494	0.0175	0.0016	0.0061

APPENDIX H: MAGNITUDES OF DIFFERENCE BETWEEN PARAMETERS AND ESTIMATED STANDARD ERRORS OF ESTIMATED VARIANCE COMPONENTS

Table H1. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $p \times i$ Design, Normal Data

	SE(*p*)	*SE*(*i*)	*SE*(*pi*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**0**	-66	**0**	**0**	-32
i	-9	**-2**	77	77	40
p,i	45	9	124	124	68
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**0**	-52	**0**	**0**	-21
i	-74	**-1**	1	1	0
p,i	6	19	73	73	49
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**0**	-84	**0**	**0**	-77
i	-11	**-1**	77	77	4
p,i	44	1	124	124	9
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**0**	-78	**0**	**0**	-69
i	-74	**-1**	1	1	-1
p,i	6	4	73	73	9

Table H2. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $p \times i$ Design, Dichotomous Data

	$SE(p)$	$SE(i)$	$SE(pi)$	$SE(\delta)$	$SE(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**-15**	-43	**-38**	**-38**	-38
i	35	**-4**	-9	-9	-9
p,i	72	23	17	17	16
Normal Eqs.	-35	-6	-67	-67	-66
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**-6**	-16	**-11**	**-11**	-10
i	-58	**-10**	-48	-48	-48
p,i	8	40	7	7	8
Normal Eqs.	-32	-15	-75	-75	-75
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**-40**	-73	**-65**	**-65**	-68
i	40	**-7**	-10	-10	-10
p,i	60	0	-1	-1	-2
Normal Eqs.	-52	-16	-79	-79	-64
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**-13**	-53	**-32**	**-32**	-34
i	-42	**-8**	-25	-25	-24
p,i	9	-8	4	4	3
Normal Eqs.	-36	-29	-78	-78	-71

Table H3. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $p \times i$ Design, Polytomous Data

	SE(*p*)	*SE*(*i*)	*SE*(*pi*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**-2**	-88	**0**	**0**	-74
i	-27	**-16**	88	88	-5
p,i	30	-14	134	134	2
Normal Eqs.	-19	-48	8	8	-44
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**-2**	-83	**2**	**2**	-64
i	-80	**-5**	4	4	-4
p,i	2	-2	77	77	9
Normal Eqs.	-21	-61	12	12	-49
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**-2**	-88	**0**	**0**	-85
i	-28	**-8**	123	123	-4
p,i	29	-6	165	165	-2
Normal Eqs.	-19	-19	34	34	-19
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**0**	-84	**1**	**1**	-80
i	-80	**1**	5	5	1
p,i	4	3	77	77	5
Normal Eqs.	-21	-25	39	39	-24

Table H4. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $p \times i \times h$ Design, Normal Data

	$SE(p)$	$SE(i)$	$SE(h)$	$SE(pi)$	$SE(ph)$	$SE(ih)$	$SE(pih)$	$SE(\delta)$	$SE(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
p	**0**	-55	-72	**0**	**0**	-34	**0**	**0**	-45
i	-43	**-6**	-69	-7	142	**-8**	229	96	0
h	505	20	**26**	2670	603	48	1897	314	173
p,i	29	12	-55	64	173	28	256	133	21
p,h	523	40	32	2676	617	75	1902	338	184
i,h	543	74	33	2709	682	97	1950	440	225
p,i,h	567	100	41	2721	702	134	1960	470	239
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
p	**-1**	-48	-64	**0**	**-1**	-11	**0**	**-1**	-38
i	-64	**-8**	-75	-5	3	**-2**	77	-4	-42
h	148	-2	**-5**	1938	249	23	1124	116	64
p,i	12	17	-53	67	55	57	125	51	-8
p,h	179	28	7	1943	273	72	1130	155	84
i,h	155	56	1	1947	269	80	1141	147	79
p,i,h	191	93	14	1961	301	144	1154	193	103

Table H4. Continued

	$SE(p)$	$SE(i)$	$SE(h)$	$SE(pi)$	$SE(ph)$	$SE(ih)$	$SE(pih)$	$SE(\delta)$	$SE(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
p	**0**	-87	-83	**0**	**0**	**-74**	**0**	**0**	-76
i	-21	**-6**	-44	34	232	-5	229	128	-24
h	222	-38	**26**	2046	483	71	1897	94	22
p,i	42	-4	-39	90	260	2	256	167	-18
p,h	245	-36	28	2052	500	75	1902	141	26
i,h	280	21	49	2079	649	117	1950	382	71
p,i,h	310	23	52	2091	673	123	1960	422	77
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
p	**-1**	-84	-78	**0**	**-1**	-63	**0**	**-1**	-70
i	-57	**-8**	-58	5	48	**-1**	77	26	-42
h	25	-52	**-6**	1208	215	47	1124	28	-6
p,i	16	-5	-51	73	93	11	125	78	-31
p,h	69	-48	-1	1214	242	55	1130	87	2
i,h	38	11	11	1217	258	98	1141	103	19
p,i,h	86	14	15	1232	295	109	1154	165	29

Table H5. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $p \times i \times h$ Design, Dichotomous Data

	SE(*p*)	*SE*(*i*)	*SE*(*h*)	*SE*(*pi*)	*SE*(*ph*)	*SE*(*ih*)	*SE*(*pih*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
p	**-20**	-26	-60	**2**	**-25**	-1	**-58**	**-43**	-49
i	-30	**-13**	-60	-6	151	**-10**	-4	50	29
h	285	32	**22**	2523	218	26	462	-21	-15
p,i	26	33	-37	67	174	62	10	70	47
p,h	307	84	34	2533	235	89	465	14	14
i,h	393	87	35	2564	377	80	479	173	139
p,i,h	425	155	51	2582	401	161	483	197	161
Normal Eqs.	-31	-4	4	-7	-37	-3	-72	-54	-46
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
p	**-15**	-16	-55	**-1**	**-27**	23	**-52**	**-36**	-43
i	-53	**-9**	-69	-5	10	**-10**	-52	-41	-27
h	104	15	**-12**	1819	69	-3	180	-23	-23
p,i	7	47	-39	67	43	89	-27	17	1
p,h	137	85	7	1841	94	90	187	14	7
i,h	124	76	-3	1840	112	55	187	30	17
p,i,h	164	164	20	1857	144	172	194	68	49
Normal Eqs.	-30	-3	-10	-9	-44	-8	-77	-53	-48

Table H5. Continued

	$SE(p)$	$SE(i)$	$SE(h)$	$SE(pi)$	$SE(ph)$	$SE(ih)$	$SE(pih)$	$SE(\delta)$	$SE(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
p	**-46**	-80	-74	**-24**	**-29**	-62	**-73**	**-62**	-71
i	-12	**-18**	-26	-13	261	**-19**	-16	58	15
h	56	-30	**12**	1541	115	35	293	18	-18
p,i	15	-14	-18	34	280	-4	-9	69	23
p,h	75	-24	18	1547	137	46	295	31	-9
i,h	201	18	53	1582	454	78	311	184	100
p,i,h	223	24	61	1593	479	91	314	197	110
Normal Eqs.	-52	-16	1	-35	-37	-17	-80	-67	-47
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
p	**-35**	-73	-70	**-28**	**-34**	47	**-67**	**-54**	-64
i	-31	**-12**	-45	-18	80	**-13**	-35	18	-12
h	-1	-37	**-10**	893	29	15	111	0	-23
p,i	3	-5	-33	27	105	12	-24	35	0
p,h	27	-28	-2	906	56	35	115	18	-11
i,h	37	16	15	914	124	62	124	69	27
p,i,h	68	25	26	924	173	88	128	89	41
Normal Eqs.	-46	-17	-7	-42	-45	-23	-83	-64	-47

Table H6. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $p \times i \times h$ Design, Polytomous Data

	SE(*p*)	*SE*(*i*)	*SE*(*h*)	*SE*(*pi*)	*SE*(*ph*)	*SE*(*ih*)	*SE*(*pih*)	*SE*(δ)	*SE*(Δ)
	VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)								
p	**1**	-87	-90	**-2**	**0**	-83	**-1**	**1**	-79
i	-51	**-19**	-85	-9	174	**-30**	262	122	-54
h	410	-9	**-18**	2593	698	-3	2115	368	32
p,i	22	-17	-81	61	205	-26	287	159	-46
p,h	425	-6	-17	2598	710	0	2120	390	36
i,h	442	39	-15	2632	788	39	2174	511	54
p,i,h	463	42	-14	2640	806	43	2183	540	59
Normal Eqs.	-15	-28	-32	-3	14	-43	11	14	-30
	VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)								
p	**3**	-89	-87	**0**	**0**	-76	**0**	**0**	-78
i	-72	**-24**	-91	-5	35	**-13**	103	25	-74
h	88	-56	**-25**	1851	350	-24	1352	175	-4
p,i	10	-22	-83	67	85	-7	146	78	-62
p,h	122	-52	-22	1861	373	-17	1361	211	1
i,h	91	-4	-23	1865	378	31	1375	216	2
p,i,h	131	-1	-21	1877	409	38	1387	263	8
Normal Eqs.	-22	-57	-24	-4	30	-59	19	30	-22

Table H6. Continued

	$SE(p)$	$SE(i)$	$SE(h)$	$SE(pi)$	$SE(ph)$	$SE(ih)$	$SE(pih)$	$SE(\delta)$	$SE(\Delta)$
	VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)								
p	**1**	-88	-92	**1**	**1**	-87	**0**	**2**	-86
i	-32	**-8**	-68	28	195	**-16**	193	105	-53
h	173	-13	**-12**	1912	417	26	1669	79	-10
p,i	32	-6	-66	86	224	-15	222	145	-50
p,h	200	-12	-11	1919	436	28	1674	131	-8
i,h	222	42	-1	1944	565	73	1716	334	15
p,i,h	254	43	-1	1957	590	75	1727	375	18
Normal Eqs.	-15	-2	-26	-6	-11	-26	-11	-11	-23
	VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)								
p	**-1**	-85	-91	**-2**	**-2**	-85	**-1**	**-3**	-85
i	-66	**-5**	-83	1	11	**-13**	56	-4	-72
h	0	-46	**-30**	1025	140	-12	937	5	-29
p,i	10	-3	-80	68	59	-11	107	49	-67
p,h	48	-43	-29	1034	172	-10	945	65	-27
i,h	10	18	-26	1035	172	40	952	59	-21
p,i,h	61	21	-25	1052	210	42	968	119	-18
Normal Eqs.	-20	-7	-30	-14	-24	-50	-15	-25	-29

Table H7. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $i:p$ Design, Normal Data

	SE(*p*)	*SE*(*i*:*p*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 ($n_p = 50$, $n_i = 5$)				
p	**-1**	0	**0**	0
i	170	**83**	83	83
p,i	108	67	67	67
VC Structure A, Sample Size Pattern 2 ($n_p = 50$, $n_i = 50$)				
p	**-2**	-1	**-1**	-1
i	-41	**0**	0	0
p,i	26	71	71	71
VC Structure B, Sample Size Pattern 1 ($n_p = 50$, $n_i = 5$)				
p	**-1**	0	**0**	0
i	47	**83**	83	83
p,i	53	67	67	67
VC Structure B, Sample Size Pattern 2 ($n_p = 50$, $n_i = 50$)				
p	**-2**	-1	**-1**	-1
i	-70	**0**	0	0
p,i	7	71	71	71

Table H8. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $i : p$ Design, Dichotomous Data

	SE(*p*)	*SE*(*i*:*p*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
p	**0**	-1	**-1**	-1
i	213	**43**	43	43
p,i	141	53	53	53
Normal Eqs.	-6	-25	-25	-25
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
p	**-7**	-2	**-2**	-2
i	-21	**-45**	-45	-45
p,i	39	25	25	25
Normal Eqs.	-18	-56	-56	-56
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
p	**-2**	-2	**-2**	-2
i	94	**26**	26	26
p,i	84	44	44	44
Normal Eqs.	-19	-35	-35	-35
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
p	**-4**	-2	**-2**	-2
i	-60	**-65**	-65	-65
p,i	10	9	9	9
Normal Eqs.	-29	-71	-71	-71

Table H9. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $i : p$ Design, Polytomous Data

	SE(*p*)	*SE*(*i*:*p*)	*SE*(*δ*)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
p	**-2**	-2	**-2**	-2
i	143	**93**	93	93
p,i	89	68	68	68
Normal Eqs.	-10	7	7	7
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
p	**-3**	-1	**-1**	-1
i	-63	**1**	1	1
p,i	9	72	72	72
Normal Eqs.	-35	12	12	12
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
p	**-3**	-1	**-1**	-1
i	27	**103**	103	103
p,i	39	72	72	72
Normal Eqs.	-14	13	13	13
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
p	**-6**	-1	**-1**	-1
i	-77	**1**	1	1
p,i	0	72	72	72
Normal Eqs.	-23	18	18	18

Table H10. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $i:h:p$ Design, Normal Data

	SE(*p*)	*SE*(*h*:*p*)	*SE*(*i*:*h*:*p*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**-1**	-1	0	**-1**	-1
i	-78	223	**51**	221	221
h	-67	**141**	-11	142	142
p,i	4	110	71	108	108
p,h	6	66	61	66	66
i,h	-48	339	73	338	338
p,i,h	14	200	145	198	198
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**-1**	-2	-2	**-2**	-2
i	-35	260	**141**	210	210
h	123	**384**	-30	406	406
p,i	35	139	69	133	133
p,h	39	93	39	98	98
i,h	291	784	134	810	810
p,i,h	106	272	114	274	274

Table H10. Continued

	$SE(p)$	$SE(h{:}p)$	$SE(i{:}h{:}p)$	$SE(\delta)$	$SE(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**-1**	-1	0	**-1**	-1
i	-73	9	**51**	5	5
h	26	**142**	-11	142	142
p,i	7	34	71	33	33
p,h	33	66	61	66	66
i,h	49	196	73	195	195
p,i,h	45	107	145	107	107
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**-29**	-30	-30	**-30**	-30
i	-49	102	**134**	68	68
h	344	**398**	-50	405	405
p,i	1	26	20	23	23
p,h	43	40	0	41	41
i,h	525	613	116	618	618
p,i,h	99	111	53	110	110

Table H11. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $i : h : p$ Design, Dichotomous Data

	SE(*p*)	*SE*(*h*:*p*)	*SE*(*i*:*h*:*p*)	*SE*(*δ*)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p *= 100,* n_i *= 20,* n_h *= 5)*					
p	**3**	0	1	**0**	0
i	-70	226	**-63**	154	154
h	-56	**99**	-73	62	62
p,i	10	128	8	92	92
p,h	12	64	8	45	45
i,h	-25	310	-52	229	229
p,i,h	28	220	19	163	163
Normal Eqs.	-25	-19	-78	-35	-35
VC Structure A, Sample Size Pattern 2 (n_p *= 50,* n_i *= 5,* n_h *= 2)*					
p	**-3**	-3	1	**-3**	-3
i	-13	233	**43**	165	165
h	197	**301**	-51	295	295
p,i	58	140	38	124	124
p,h	70	89	24	85	85
i,h	436	658	45	635	635
p,i,h	185	276	68	257	257
Normal Eqs.	-21	-17	-43	-22	-22

Table H11. Continued

	$SE(p)$	$SE(h{:}p)$	$SE(i{:}h{:}p)$	$SE(\delta)$	$SE(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**3**	1	0	**1**	1
i	-60	41	**-54**	25	25
h	38	**82**	-52	68	68
p,i	17	49	11	43	43
p,h	46	56	20	49	49
i,h	83	162	-32	140	140
p,i,h	71	115	36	102	102
Normal Eqs.	-17	-27	-72	-33	-33
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**-1**	-1	-1	**-1**	-1
i	-3	152	**51**	107	107
h	383	**322**	-43	315	315
p,i	70	105	41	94	94
p,h	135	94	30	90	90
i,h	665	591	54	571	571
p,i,h	268	232	77	217	217
Normal Eqs.	-12	-15	-40	-19	-19

Table H12. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $i:h:p$ Design, Polytomous Data

	SE(*p*)	*SE*(*h*:*p*)	*SE*(*i*:*h*:*p*)	*SE*(δ)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**-2**	-1	0	**-1**	-1
i	-84	73	**68**	66	66
h	-75	**44**	-9	44	44
p,i	1	40	74	38	38
p,h	2	52	62	51	51
i,h	-60	144	80	142	142
p,i,h	7	102	149	100	100
Normal Eqs.	-24	-46	18	-46	-46
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**-5**	-2	-1	**-1**	-1
i	-47	231	**173**	187	187
h	87	**353**	-28	376	376
p,i	24	122	77	117	117
p,h	28	90	41	95	95
i,h	225	721	162	748	748
p,i,h	77	243	124	247	247
Normal Eqs.	-17	-7	15	-7	-7

Table H12. Continued

	SE(*p*)	*SE*(*h*:*p*)	*SE*(*i*:*h*:*p*)	*SE*(δ)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**-1**	-1	1	**-1**	-1
i	-75	1	**69**	-3	-3
h	14	**124**	-11	124	124
p,i	5	29	75	29	29
p,h	28	63	63	63	63
i,h	35	174	81	174	174
p,i,h	39	99	150	98	98
Normal Eqs.	-9	-8	18	-8	-8
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**-1**	-1	0	**-1**	-1
i	-31	109	**174**	79	79
h	332	**392**	-30	402	402
p,i	38	74	76	71	71
p,h	96	96	41	98	98
i,h	508	608	162	617	617
p,i,h	168	190	124	190	190
Normal Eqs.	-3	-2	14	-2	-2

Table H13. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $p \times (i:h)$ Design, Normal Data

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(*δ*)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**-1**	-73	**0**	-59	**0**	**0**	48
i	-71	-31	289	**-2**	139	267	89
h	-20	**-3**	227	-9	-11	229	90
p,i	7	-22	315	13	176	295	102
p,h	33	4	254	6	61	256	103
i,h	16	35	429	59	105	426	190
p,i,h	63	43	465	77	177	462	207
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**0**	-51	**-1**	-43	-1	**-1**	-21
i	-61	69	1772	**43**	**1190**	1327	1019
h	-34	**-7**	32	-6	-5	42	26
p,i	13	87	1779	62	1196	1336	1026
p,h	29	15	91	22	67	97	66
i,h	1275	624	1355	1250	3663	555	423
p,i,h	62	122	769	95	510	593	452

Table H13. Continued

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(δ)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**-1**	-83	**0**	-81	**0**	**0**	-73
i	-67	-41	66	**-2**	139	56	-30
h	125	**-3**	229	-9	-11	229	29
p,i	9	-38	104	2	176	96	-23
p,h	156	0	255	-6	61	255	34
i,h	167	26	303	60	105	302	62
p,i,h	198	29	334	64	177	333	68
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**0**	-75	**-1**	-73	**-1**	**0**	-62
i	-54	59	1241	**54**	1190	855	285
h	13	**-7**	37	-6	-5	42	2
p,i	13	-7	37	-6	-5	865	289
p,h	68	0	94	1	67	97	15
i,h	59	79	512	66	482	358	132
p,i,h	113	87	545	73	510	398	144

Table H14. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $p \times (i:h)$ Design, Dichotomous Data

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(*δ*)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**-15**	-62	**-9**	-36	**-43**	**-25**	-35
i	-59	-28	304	**-12**	-46	168	124
h	3	**-5**	148	-12	-25	80	61
p,i	0	-10	329	21	-19	189	142
p,h	40	8	177	20	-3	103	81
i,h	59	36	375	46	-5	233	187
p,i,h	92	54	414	83	16	261	211
Normal Eqs.	-34	-3	-25	-17	-83	-48	-41
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**-8**	-27	**-2**	-14	**-24**	**-20**	-22
i	-47	61	1603	**17**	388	678	640
h	-5	**-6**	22	-10	-24	-9	-10
p,i	16	105	1614	63	396	687	649
p,h	45	38	82	46	16	33	31
i,h	56	99	688	52	137	279	263
p,i,h	105	164	724	116	159	307	290
Normal Eqs.	-27	-3	-15	-8	-62	-51	-50

Table H14. Continued

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(δ)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**-34**	-78	**-49**	-72	**-74**	**-59**	-60
i	-55	-38	36	**-25**	-56	2	-20
h	108	**-8**	45	-11	-12	20	7
p,i	-14	-33	51	-17	-48	14	-12
p,h	126	-3	60	-4	-7	31	14
i,h	179	24	116	33	3	72	46
p,i,h	198	29	133	41	8	85	54
Normal Eqs.	-42	-6	-60	-38	-90	-70	-42
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**-20**	-63	**-17**	-59	**-57**	**-51**	-58
i	-38	52	1134	**19**	238	351	273
h	56	**-7**	11	-10	-10	-9	-10
p,i	14	63	1142	30	242	357	277
p,h	91	5	58	5	5	10	5
i,h	144	84	520	47	89	159	128
p,i,h	182	101	548	64	101	174	140
Normal Eqs.	-29	-9	-30	-22	-74	-66	-55

Table H15. Magnitudes of Difference Between Parameters and Estimated Standard Errors of Estimated Variance Components for the $p \times (i:h)$ Design, Polytomous Data

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(*δ*)	*SE*(Δ)
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**1**	-94	**1**	-87	**2**	**1**	-79
i	-78	-83	129	**-10**	176	115	-53
h	-40	**-22**	102	-12	-10	102	-12
p,i	5	-82	158	-8	211	146	-47
p,h	21	-22	140	-10	63	140	-8
i,h	-13	-19	217	50	124	214	4
p,i,h	40	-18	253	52	195	250	9
Normal Eqs.	-25	-37	-41	-65	18	-41	-37
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**-4**	-80	**-1**	-77	**1**	**-1**	-58
i	-72	8	1605	**-12**	1341	1199	450
h	-51	**-7**	26	-3	-4	34	1
p,i	3	13	1613	-7	1348	1208	454
p,h	14	-2	85	2	69	91	16
i,h	-24	58	660	45	547	493	180
p,i,h	34	65	692	51	576	529	193
Normal Eqs.	-26	-38	-9	-50	12	-9	-38

Table H15. Continued

	SE(*p*)	*SE*(*h*)	*SE*(*ph*)	*SE*(*i*:*h*)	*SE*(*pi*:*h*)	*SE*(δ)	*SE*(Δ)
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**-1**	-93	**-1**	-83	**0**	**-1**	-88
i	-77	-76	53	**-4**	147	44	-71
h	56	**-23**	204	-9	-10	204	-15
p,i	3	-74	91	-1	184	84	-68
p,h	91	-22	232	-7	61	232	-14
i,h	85	-16	272	56	109	271	-5
p,i,h	119	-15	303	60	181	302	-4
Normal Eqs.	-31	-34	-8	-11	4	-8	-34
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**-2**	-84	**0**	-76	**-2**	**1**	-76
i	-70	-2	1211	**42**	1205	835	131
h	-23	**-7**	36	-5	-7	41	-4
p,i	5	1	1220	46	1213	846	133
p,h	34	-5	94	1	63	98	2
i,h	8	34	499	61	488	349	60
p,i,h	60	38	532	68	516	388	66
Normal Eqs.	-32	-28	-2	-9	1	-2	-29

APPENDIX I: ESTIMATED STANDARD ERRORS OF ESTIMATED ABSOLUTE ERROR VARIANCES USING TONG AND BRENNAN'S (2007) WORKAROUND

Table I1. Estimated Standard Error of Estimated Absolute Error Variance Using Tong and Brennan's (2007) Workaround for the $p \times i$ Design

	Parameter $SE(\Delta)$	Estimate $S\hat{E}(\Delta)$	Pct. Difference
	Normal Data		
A1	0.0032	0.0032	-1
A2	0.0003	0.0003	0
B1	0.0097	0.0096	-1
B2	0.0009	0.0008	-1
	Dichotomous Data		
A1	0.0011	0.0007	-39
A2	0.0002	0.0002	-15
B1	0.0021	0.0010	-55
B2	0.0002	0.0001	-38
	Polytomous Data		
A1	0.0076	0.0065	-15
A2	0.0007	0.0007	-4
B1	0.0128	0.0118	-8
B2	0.0011	0.0011	1

Table I2. Estimated Standard Error of Estimated Absolute Error Variance Using Tong and Brennan's (2007) Workaround for the $p \times i \times h$ Design

	Parameter $SE(\Delta)$	Estimate $S\hat{E}(\Delta)$	Pct. Difference
	Normal Data		
A1	0.0355	0.0427	20
A2	0.0176	0.0169	-4
B1	0.0507	0.0601	19
B2	0.0209	0.0198	-5
	Dichotomous Data		
A1	0.0055	0.0036	-35
A2	0.0027	0.0017	-38
B1	0.0073	0.0042	-42
B2	0.0029	0.0016	-45
	Polytomous Data		
A1	0.0801	0.0663	-17
A2	0.0398	0.0304	-23
B1	0.0903	0.0804	-11
B2	0.0470	0.0336	-29

Table I3. Estimated Standard Error of Estimated Absolute Error Variance Using Tong and Brennan's (2007) Workaround for the $i : p$ Design

	Parameter $SE(\Delta)$	Estimate $S\hat{E}(\Delta)$	Pct. Difference
	Normal Data		
A1	0.0180	0.0179	0
A2	0.0005	0.0005	-1
B1	0.0140	0.0139	0
B2	0.0004	0.0004	-1
	Dichotomous Data		
A1	0.0034	0.0034	-1
A2	0.0002	0.0002	-2
B1	0.0035	0.0035	-2
B2	0.0002	0.0002	-2
	Polytomous Data		
A1	0.0168	0.0165	-2
A2	0.0005	0.0005	-1
B1	0.0123	0.0121	-1
B2	0.0003	0.0003	-1

Table I4. Estimated Standard Error of Estimated Absolute Error Variance Using Tong and Brennan's (2007) Workaround for the $i:h:p$ Design

	Parameter $SE(\Delta)$	Estimate $\hat{SE}(\Delta)$	Pct. Difference
Normal Data			
A1	0.0008	0.0008	-1
A2	0.0160	0.0157	-2
B1	0.0026	0.0026	-1
B2	0.0290	0.0204	-30
Dichotomous Data			
A1	0.0002	0.0002	0
A2	0.0032	0.0031	-3
B1	0.0003	0.0003	1
B2	0.0041	0.0041	-1
Polytomous Data			
A1	0.0015	0.0014	-1
A2	0.0173	0.0171	-1
B1	0.0028	0.0028	-1
B2	0.0294	0.0291	-1

Table I5. Estimated Standard Error of Estimated Absolute Error Variance Using Tong and Brennan's (2007) Workaround for the $p \times (i : h)$ Design

	Parameter $SE(\Delta)$	Estimate $S\hat{E}(\Delta)$	Pct. Difference
	Normal Data		
A1	0.0021	0.0021	-2
A2	0.0018	0.0020	8
B1	0.0088	0.0086	-3
B2	0.0051	0.0056	10
	Dichotomous Data		
A1	0.0003	0.0002	-27
A2	0.0004	0.0003	-22
B1	0.0010	0.0006	-40
B2	0.0009	0.0005	-42
	Polytomous Data		
A1	0.0084	0.0067	-21
A2	0.0037	0.0035	-5
B1	0.0217	0.0168	-22
B2	0.0085	0.0085	-1

APPENDIX J: PERCENTILE CONFIDENCE INTERVAL COVERAGE

Table J1. 90% Percentile Confidence Interval Coverage for the $p \times i$ Design, Normal Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**88.4**	43.7	**87.8**	**87.8**	72.0
i	84.3	**81.2**	98.6	98.6	95.4
p,i	97.9	87.3	87.8	87.8	72.0
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**88.5**	58.2	**89.8**	**89.8**	79.8
i	33.5	**88.2**	89.1	89.1	88.5
p,i	91.3	94.5	89.8	89.8	79.8
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**88.3**	20.4	**87.8**	**87.8**	30.5
i	83.1	**80.8**	98.6	98.6	82.5
p,i	97.8	82.1	87.8	87.8	30.5
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**88.5**	27.5	**89.8**	**89.8**	39.8
i	33.2	**88.3**	89.1	89.1	88.0
p,i	91.3	89.7	89.8	89.8	39.8

Table J2. 90% Percentile Confidence Interval Coverage for the $p \times i$ Design, Dichotomous Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**80.0**	62.7	**70.2**	**70.2**	68.8
i	95.1	**76.4**	84.4	84.4	83.6
p,i	99.5	90.2	70.2	70.2	68.8
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**84.2**	84.4	**84.6**	**84.6**	84.7
i	49.7	**82.2**	56.4	56.4	56.7
p,i	88.9	97.2	84.6	84.6	84.7
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**63.7**	33.4	**42.8**	**42.8**	41.4
i	95.2	**69.0**	83.8	83.8	81.1
p,i	98.3	74.7	42.8	42.8	41.4
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**80.4**	56.0	**72.6**	**72.6**	71.0
i	66.6	**79.2**	76.1	76.1	76.5
p,i	90.6	89.2	72.6	72.6	71.0

Table J3. 90% Percentile Confidence Interval Coverage for the $p \times i$ Design, Polytomous Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**88.8**	0	**89.6**	**89.6**	18.0
i	74.3	**39.4**	98.7	98.7	62.4
p,i	97.0	39.4	89.6	89.6	18.0
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**85.3**	22.5	**89.1**	**89.1**	48.2
i	25.4	**86.8**	88.4	88.4	84.8
p,i	86.9	87.4	89.1	89.1	48.2
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 10)					
p	**88.4**	18.8	**90.5**	**90.5**	25.1
i	76.2	**87.9**	99.8	99.8	88.3
p,i	96.8	88.7	90.5	90.5	25.1
VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 50)					
p	**84.9**	20.5	**89.0**	**89.0**	25.3
i	27.0	**83.6**	89.9	89.9	83.1
p,i	87.0	85.1	89.0	89.0	25.3

Table J4. 90% Percentile Confidence Interval Coverage for the $p \times i \times h$ Design, Normal Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
	VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)								
p	**89.2**	53.1	31.4	**89.8**	**89.9**	70.1	**90.4**	**89.4**	71.3
i	60.4	**75.7**	32.6	84.4	99.1	**70.9**	99.9	98.1	90.3
h	100	66.3	**46.3**	100	100	67.0	100	98.7	97.0
p,i	96.3	86.1	46.8	98.9	100	89.5	99.9	99.9	94.9
p,h	100	89.8	61.3	100	100	89.1	100	100	99.7
i,h	100	96.6	63.8	100	100	82.2	100	100	100
p,i,h	100	99.6	69.4	100	100	93.6	100	100	100
	VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)								
p	**89.5**	58.2	42.7	**89.0**	**89.4**	86.9	**90.0**	**88.9**	73.0
i	43.9	**77.5**	29.0	87.0	88.9	**83.5**	98.7	86.1	70.1
h	99.5	75.8	**56.3**	100	99.9	79.4	100	95.6	91.8
p,i	93.6	90.6	53.3	98.7	98.8	98.8	99.8	98.7	90.4
p,h	100	95.4	74.7	100	100	99.0	100	99.6	99.1
i,h	99.7	98.0	68.4	100	100	95.4	100	99.9	98.8
p,i,h	100	100	79.9	100	100	99.8	100	100	99.8

Table J4. Continued

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
	VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)								
p	**88.8**	15.6	20.1	**90.8**	**89.7**	33.9	**90.4**	**89.2**	31.2
i	73.8	**70.8**	53.9	92.8	99.9	**71.7**	99.9	99.1	81.1
h	99.0	44.9	**45.0**	100	100	73.9	100	80.9	56.3
p,i	97.6	72.6	58.2	99.4	100	75.9	99.9	99.9	86.3
p,h	100	55.3	53.7	100	100	83.7	100	98.3	76.9
i,h	100	88.0	73.9	100	100	86.3	100	100	99.7
p,i,h	100	88.6	77.4	100	100	89.3	100	100	99.9
	VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)								
p	**89.4**	20.7	26.4	**89.9**	**89.5**	45.4	**90.0**	**88.8**	40.0
i	52.7	**77.3**	47.8	89.2	96.8	**84.0**	98.7	94.1	67.1
h	90.9	45.9	**58.0**	100	99.7	86.3	100	85.2	65.1
p,i	94.6	79.2	54.8	99.6	99.7	90.5	99.8	99.4	76.4
p,h	99.6	54.2	65.9	100	100	92.7	100	98.8	80.0
i,h	95.8	87.6	76.5	100	100	96.2	100	99.8	91.4
p,i,h	99.9	89.0	79.4	100	100	98.4	100	100	95.9

Table J5. 90% Percentile Confidence Interval Coverage for the $p \times i \times h$ Design, Dichotmous Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
p	**79.2**	77.4	41.2	**89.1**	**74.8**	88.6	**51.6**	**63.4**	59.0
i	69.3	**75.1**	36.0	84.5	99.6	**68.3**	83.3	94.9	91.4
h	99.1	68.5	**45.6**	100	95.5	55.7	99.9	58.8	57.4
p,i	95.5	95.0	41.2	89.1	74.8	88.6	51.6	63.4	59.0
p,h	99.9	99.1	65.0	100	99.2	94.3	99.9	87.4	86.7
i,h	100	96.9	63.5	100	100	75.6	100	99.8	99.5
p,i,h	100	100	75.3	100	100	97.6	100	99.9	99.9
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
p	**78.2**	82.5	49.2	**89.0**	**74.9**	96.5	**55.9**	**68.7**	65.8
i	53.8	**76.2**	35.8	87.8	90.5	**78.8**	56.4	81.9	74.8
h	95.8	79.0	**54.8**	100	90.5	70.9	99.3	69.0	67.4
p,i	89.6	98.1	49.2	89.0	74.9	96.5	55.9	68.7	65.8
p,h	99.8	99.5	74.7	100	97.4	99.5	99.7	92.6	91.3
i,h	98.8	98.3	67.3	100	99.4	91.4	99.7	96.0	93.9
p,i,h	100	100	82.9	100	99.8	100	99.7	98.8	98.4

Table J5. Continued

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
	VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)								
p	**58.6**	22.0	28.9	**75.4**	**70.6**	44.9	**34.9**	**43.5**	34.9
i	76.1	**59.3**	57.2	79.2	100	**59.9**	77.6	94.9	87.4
h	78.6	44.9	**43.4**	100	82.1	59.4	97.3	72.2	60.9
p,i	90.1	64.2	28.9	75.4	70.6	44.9	34.9	43.5	34.9
p,h	93.7	60.2	55.2	100	95.3	78.2	99.0	89.7	79.7
i,h	99.6	78.8	71.8	100	100	76.1	99.6	98.9	95.6
p,i,h	99.9	81.9	76.9	100	100	84.8	99.7	99.4	97.3
	VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)								
p	**64.8**	30.3	35.4	**74.9**	**70.7**	60.1	**39.9**	**52.7**	45.1
i	70.1	**65.6**	51.5	78.8	98.4	**71.4**	69.1	90.3	81.7
h	71.2	48.5	**55.3**	100	82.1	72.2	95.8	79.2	71.4
p,i	87.3	71.9	35.4	74.9	70.7	60.1	39.9	52.7	45.1
p,h	90.8	65.4	65.9	100	93.2	89.2	97.3	92.0	85.0
i,h	91.4	82.4	73.7	100	99.9	88.2	98.6	98.7	95.0
p,i,h	97.5	85.8	80.4	100	100	95.2	98.9	99.6	97.4

Table J6. 90% Percentile Confidence Interval Coverage for the $p \times i \times h$ Design, Polytomous Data

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)									
p	**88.3**	6.4	0.8	**89.3**	**90.6**	8.2	**90.4**	**89.9**	53.3
i	54.0	**10.4**	7.5	84.6	99.8	**11.9**	99.8	99.3	93.7
h	100	8.7	**6.2**	100	100	10.8	100	99.2	94.5
p,i	95.2	14.4	7.6	99.3	100	34.7	100	100	94.4
p,h	100	13.8	9.0	100	100	34.9	100	100	99.7
i,h	100	16.4	9.6	100	100	21.0	100	100	100
p,i,h	100	36.5	13.9	100	100	52.7	100	100	100
VC Structure A, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)									
p	**88.2**	5.9	0.4	**90.9**	**90.4**	2.5	**89.4**	**90.4**	38.3
i	31.9	**13.8**	1.5	87.7	96.5	**28.3**	99.3	94.7	61.2
h	96.6	24.6	**10.7**	100	100	28.2	100	98.6	82.3
p,i	92.1	21.7	2.8	99.1	99.8	43.4	100	99.7	79.9
p,h	99.7	25.8	12.9	100	100	43.2	100	99.9	96.6
i,h	98.2	33.9	13.2	100	100	33.9	100	100	98.3
p,i,h	99.9	42.2	17.2	100	100	71.7	100	100	100

Table J6. Continued

	$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(h)$	$\sigma^2(pi)$	$\sigma^2(ph)$	$\sigma^2(ih)$	$\sigma^2(pih)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
	VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 5, n_h = 2)								
p	**89.0**	10.1	2.2	**89.2**	**89.1**	0	**90.0**	**90.7**	32.8
i	68.2	**63.3**	33.0	91.6	99.5	**37.2**	99.9	98.1	70.7
h	98.1	30.6	**10.7**	100	99.7	38.1	100	74.9	41.7
p,i	96.4	63.3	33.6	99.3	100	37.2	100	100	72.6
p,h	100	35.1	14.8	100	100	38.1	100	97.3	64.7
i,h	100	76.3	26.2	100	100	38.1	100	100	96.9
p,i,h	100	76.3	33.8	100	100	38.1	100	100	98.7
	VC Structure B, Sample Size Pattern 2 (n_p = 100, n_i = 10, n_h = 3)								
p	**86.3**	19.0	0.5	**90.3**	**88.4**	11.6	**89.6**	**88.3**	24.8
i	39.8	**83.0**	6.0	88.6	91.2	**73.9**	96.3	86.9	47.6
h	82.8	32.4	**14.1**	100	97.3	65.8	100	80.1	36.0
p,i	90.5	83.0	8.0	99.3	99.0	74.1	99.9	98.6	58.9
p,h	98.5	41.4	15.8	100	99.4	75.0	100	98.2	60.8
i,h	89.8	85.0	23.3	100	100	76.6	100	99.0	72.6
p,i,h	98.9	85.3	25.0	100	100	76.6	100	100	84.7

Table J7. 90% Percentile Confidence Interval Coverage for the $i:p$ Design, Normal Data

	$\sigma^2(p)$	$\sigma^2(i:p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
p	**87.1**	89.0	**89.0**	89.0
i	99.4	**98.4**	98.4	98.4
p,i	99.9	89.0	89.0	89.0
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
p	**86.2**	88.9	**88.9**	88.9
i	65.6	**89.6**	89.6	89.6
p,i	95.4	88.9	88.9	88.9
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
p	**86.8**	89.0	**89.0**	89.0
i	95.7	**98.4**	98.4	98.4
p,i	98.2	89.0	89.0	89.0
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
p	**85.3**	88.9	**88.9**	88.9
i	38.1	**89.6**	89.6	89.6
p,i	88.4	88.9	88.9	88.9

Table J8. 90% Percentile Confidence Interval Coverage for the $i : p$ Design, Dichotomous Data

	$\sigma^2(p)$	$\sigma^2(i{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
p	**83.3**	90.2	**90.2**	90.2
i	99.6	**95.0**	95.0	95.0
p,i	100	90.2	90.2	90.2
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
p	**81.7**	89.4	**89.4**	89.4
i	80.8	**63.7**	63.7	63.7
p,i	96.8	89.4	89.4	89.4
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
p	**81.6**	89.4	**89.4**	89.4
i	97.6	**92.7**	92.7	92.7
p,i	99.6	89.4	89.4	89.4
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
p	**80.0**	89.1	**89.1**	89.1
i	47.3	**44.2**	44.2	44.2
p,i	87.8	89.1	89.1	89.1

Table J9. 90% Percentile Confidence Interval Coverage for the $i : p$ Design, Polytomous Data

	$\sigma^2(p)$	$\sigma^2(i{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
p	**86.2**	89.2	**89.2**	89.2
i	99.2	**97.8**	98.0	98.0
p,i	99.7	89.2	89.2	89.2
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
p	**87.0**	88.2	**88.2**	88.2
i	44.5	**88.8**	88.8	88.8
p,i	90.1	88.2	88.2	88.2
VC Structure B, Sample Size Pattern 1 (n_p = 50, n_i = 5)				
p	**85.5**	89.9	**89.4**	89.4
i	93.6	**97.8**	97.8	97.8
p,i	97.5	89.9	89.4	89.4
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 50)				
p	**79.6**	88.8	**88.8**	88.8
i	27.6	**88.7**	88.7	88.7
p,i	82.1	88.8	88.8	88.8

Table J10. 90% Percentile Confidence Interval Coverage for the $i : h : p$ Design, Normal Data

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**88.2**	87.9	89.9	**88.6**	88.6
i	25.6	100	**98.5**	100	100
h	38.3	**99.3**	78.3	99.6	99.6
p,i	87.1	100	99.9	100	100
p,h	90.9	99.5	99.4	99.6	99.6
i,h	58.7	100	99.5	100	100
p,i,h	93.1	100	100	100	100
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**86.9**	86.8	87.3	**86.5**	86.5
i	65.9	99.9	**99.1**	99.9	99.9
h	95.7	**99.7**	48.8	99.9	99.9
p,i	96.7	100	98.8	100	100
p,h	97.4	99.1	97.5	98.9	98.9
i,h	100	100	99.7	100	100
p,i,h	99.8	100	99.9	100	100

Table J10. Continued

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**87.6**	88.0	89.9	**88.2**	88.2
i	34.4	91.8	**98.5**	90.0	90.0
h	92.8	**99.7**	78.3	99.7	99.7
p,i	89.5	95.8	99.9	95.9	95.9
p,h	96.0	99.3	99.4	99.3	99.3
i,h	96.3	100	99.5	100	100
p,i,h	97.9	100	100	100	100
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**88.7**	89.1	88.3	**88.8**	88.8
i	70.4	99.6	**100**	99.3	99.3
h	100	**100**	50.1	100	100
p,i	98.3	99.6	99.2	99.6	99.6
p,h	99.9	99.3	97.7	99.3	99.3
i,h	100	100	100	100	100
p,i,h	100	100	100	100	100

Table J11. 90% Percentile Confidence Interval Coverage for the $i : h : p$ Design, Dichotomous Data

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**86.2**	89.4	90.4	**89.3**	89.3
i	35.6	100	**46.1**	100	100
h	47.5	**98.9**	31.9	96.0	96.0
p,i	87.8	99.9	92.2	99.9	99.9
p,h	89.7	99.1	92.1	97.7	97.7
i,h	74.7	100	56.1	100	100
p,i,h	94.3	100	95.1	100	100
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**83.9**	84.4	89.3	**84.0**	84.0
i	79.4	99.7	**93.1**	99.5	99.5
h	98.0	**99.1**	36.4	98.5	98.5
p,i	98.5	100	97.3	100	100
p,h	99.1	98.0	95.1	96.9	96.9
i,h	100	100	95.3	100	100
p,i,h	100	100	99.6	100	100

Table J11. Continued

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**84.7**	89.1	90.7	**89.1**	89.1
i	47.6	97.2	**55.5**	95.2	95.2
h	92.8	**97.3**	52.3	96.3	96.3
p,i	90.6	97.6	93.5	97.0	97.0
p,h	97.0	98.3	95.3	98.1	98.1
i,h	98.0	100	72.4	100	100
p,i,h	98.8	100	97.6	100	100
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**86.3**	85.2	89.1	**86.7**	86.7
i	83.5	99.5	**93.2**	99.4	99.4
h	99.7	**99.0**	42.5	98.9	98.9
p,i	98.5	99.9	97.9	99.7	99.7
p,h	99.9	98.0	96.2	97.9	97.9
i,h	100	100	96.4	100	100
p,i,h	100	100	99.7	100	100

Table J12. 90% Percentile Confidence Interval Coverage for the $i:h:p$ Design, Polytomous Data

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**85.5**	87.7	89.0	**87.3**	87.3
i	21.1	99.0	**98.9**	98.8	98.8
h	29.7	**93.5**	78.6	93.3	93.3
p,i	86.7	98.1	99.3	98.2	98.2
p,h	86.7	97.4	99.0	97.1	97.1
i,h	44.9	100	99.5	100	100
p,i,h	88.8	99.8	100	99.8	99.8
VC Structure A, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**83.5**	86.6	89.9	**85.7**	85.7
i	60.7	100	**99.8**	100	100
h	94.5	**100**	51.0	100	100
p,i	93.2	100	99.7	99.9	99.9
p,h	94.1	99.0	97.6	99.0	99.0
i,h	100	100	99.8	100	100
p,i,h	99.7	100	99.9	100	100

Table J12. Continued

	$\sigma^2(p)$	$\sigma^2(h{:}p)$	$\sigma^2(i{:}h{:}p)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 100, n_i = 20, n_h = 5)					
p	**90.0**	89.1	87.9	**89.1**	89.1
i	32.2	88.2	**98.8**	86.5	86.5
h	88.9	**98.8**	76.2	98.8	98.8
p,i	90.5	97.0	99.7	97.1	97.1
p,h	96.6	99.1	99.5	99.2	99.2
i,h	95.8	99.8	99.6	99.8	99.8
p,i,h	98.2	99.8	100	99.8	99.8
VC Structure B, Sample Size Pattern 2 (n_p = 50, n_i = 5, n_h = 2)					
p	**86.9**	88.2	89.9	**87.7**	87.7
i	69.6	99.0	**99.8**	98.0	98.0
h	99.7	**99.7**	53.0	99.7	99.7
p,i	96.7	99.7	99.8	99.5	99.5
p,h	99.6	99.2	97.5	99.2	99.2
i,h	100	100	100	100	100
p,i,h	99.9	100	100	100	100

Table J13. 90% Percentile Confidence Interval Coverage for the $p \times (i:h)$ Design, Normal Data

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i:h)$	$\sigma^2(pi:h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**88.8**	31.0	**88.5**	50.5	**90.4**	**88.7**	59.6
i	34.3	67.1	100	**84.8**	100	100	98.7
h	76.2	**68.0**	99.9	78.4	77.3	99.9	97.7
p,i	91.1	74.8	100	91.5	100	100	99.4
p,h	95.8	73.6	100	86.8	98.4	100	99.6
i,h	90.4	89.1	100	97.2	99.9	100	100
p,i,h	98.6	92.2	100	98.8	100	100	100
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**84.5**	54.8	**84.3**	58.3	**85.9**	**85.0**	75.7
i	35.0	92.2	100	**82.4**	99.4	100	100
h	66.6	**76.5**	92.1	75.9	80.2	94.1	91.1
p,i	90.7	95.3	100	94.5	100	100	100
p,h	93.7	88.4	98.7	89.0	98.9	99.3	98.5
i,h	86.0	99.0	100	94.4	100	100	100
p,i,h	98.1	99.6	100	98.6	100	100	100

Table J13. Continued

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i{:}h)$	$\sigma^2(pi{:}h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**83.6**	18.7	**84.2**	23.0	**85.4**	**84.5**	29.1
i	36.3	57.2	97.1	**83.0**	99.9	96.2	69.6
h	97.8	**65.9**	100	75.0	72.2	100	83.6
p,i	88.3	60.6	99.1	84.3	100	98.9	74.6
p,h	99.7	70.0	100	76.6	96.7	100	89.3
i,h	99.2	85.2	100	96.2	99.6	100	97.6
p,i,h	100	86.8	100	96.8	100	100	98.9
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**85.0**	30.8	**83.7**	29.5	**85.9**	**83.8**	44.6
i	38.7	91.3	100	**85.0**	99.4	100	100
h	89.4	**76.2**	91.9	74.5	80.2	93.4	80.0
p,i	92.1	92.6	100	90.9	100	100	100
p,h	98.1	80.6	98.6	80.4	98.9	99.1	88.6
i,h	96.3	99.1	100	94.8	100	100	99.9
p,i,h	99.7	99.2	100	96.4	100	100	100

Table J14. 90% Percentile Confidence Interval Coverage for the $p \times (i : h)$ Design, Dichotomous Data

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i{:}h)$	$\sigma^2(pi{:}h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**80.5**	43.5	**85.6**	70.3	**67.0**	**75.7**	72.7
i	48.1	69.5	100	**80.7**	64.3	100	99.6
h	84.9	**66.3**	99.0	76.6	72.9	96.2	93.9
p,i	87.8	83.0	100	94.4	83.3	100	99.9
p,h	96.9	76.4	99.9	93.1	86.7	99.1	98.0
i,h	97.5	91.0	100	95.6	85.1	100	100
p,i,h	99.6	96.0	100	98.1	92.8	100	100
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**83.9**	76.1	**88.1**	84.0	**77.4**	**79.0**	78.3
i	48.5	90.0	100	**76.5**	100	100	100
h	86.5	**78.2**	91.9	77.2	74.7	84.1	82.6
p,i	91.9	99.0	100	98.4	100	100	100
p,h	97.8	94.8	98.9	97.9	93.5	97.0	96.2
i,h	98.1	99.8	100	94.4	99.9	100	100
p,i,h	99.9	100	100	99.9	99.9	100	100

Table J14. Continued

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i{:}h)$	$\sigma^2(pi{:}h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**69.6**	24.6	**58.6**	33.9	**34.2**	**48.4**	38.8
i	52.3	61.6	96.7	**73.4**	52.3	86.8	79.9
h	98.9	**65.3**	91.9	76.6	77.8	89.2	81.5
p,i	83.0	65.8	98.6	79.2	60.4	92.8	84.3
p,h	99.7	68.5	95.5	80.7	82.1	92.8	86.4
i,h	99.9	84.5	99.4	92.1	86.7	97.9	95.0
p,i,h	100	87.2	99.6	93.5	89.1	98.8	96.8
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**79.2**	45.4	**80.4**	50.3	**52.5**	**57.0**	53.1
i	53.8	90.8	100	**76.4**	99.4	100	100
h	97.4	**73.9**	89.4	75.8	82.2	84.5	80.5
p,i	91.2	95.5	100	88.0	100	100	100
p,h	99.6	83.8	98.2	84.4	89.8	91.2	88.7
i,h	100	99.6	100	90.6	99.0	100	99.9
p,i,h	100	99.9	100	94.3	99.2	100	100

Table J15. 90% Percentile Confidence Interval Coverage for the $p \times (i:h)$ Design, Polytomous Data

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i:h)$	$\sigma^2(pi:h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure A, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**87.2**	1.9	**87.8**	19.5	**90.5**	**88.3**	60.7
i	26.8	21.7	99.8	**80.4**	100	99.4	95.3
h	61.8	**8.8**	97.9	81.4	80.5	97.8	84.7
p,i	89.0	26.5	99.9	80.3	100	99.9	95.4
p,h	93.5	11.6	100	78.8	98.7	100	95.7
i,h	81.3	38.4	100	94.5	99.9	100	100
p,i,h	96.8	44.6	100	94.4	100	100	100
VC Structure A, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**84.1**	21.4	**90.4**	31.1	**89.3**	**90.5**	66.6
i	25.9	62.6	100	**68.9**	100	100	99.8
h	55.1	**25.7**	94.6	70.3	83.5	95.8	83.9
p,i	88.1	70.7	100	72.1	100	100	99.9
p,h	92.9	56.2	99.6	70.3	99.1	99.6	96.6
i,h	78.4	74.7	100	72.2	100	100	100
p,i,h	97.8	94.2	100	72.2	100	100	100

Table J15. Continued

	$\sigma^2(p)$	$\sigma^2(h)$	$\sigma^2(ph)$	$\sigma^2(i{:}h)$	$\sigma^2(pi{:}h)$	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
VC Structure B, Sample Size Pattern 1 (n_p = 200, n_i = 15, n_h = 5)							
p	**89.8**	0.1	**88.8**	21.6	**89.2**	**88.9**	1.6
i	28.1	5.0	98.1	**87.9**	100	97.2	21.0
h	95.4	**25.4**	100	78.2	79.6	100	41.7
p,i	91.7	5.7	99.9	89.5	100	99.9	29.4
p,h	99.5	25.9	100	80.7	98.5	100	60.2
i,h	98.9	38.3	100	98.3	99.9	100	82.4
p,i,h	99.8	41.1	100	99.0	100	100	91.0
VC Structure B, Sample Size Pattern 2 (n_p =100, n_i = 3, n_h = 10)							
p	**87.1**	16.5	**89.8**	31.1	**91.0**	**89.3**	25.6
i	30.0	80.6	100	**82.8**	100	100	97.6
h	77.4	**59.4**	95.4	86.6	85.1	95.2	64.1
p,i	89.7	85.9	100	88.3	100	100	98.2
p,h	96.1	62.3	99.5	89.5	99.3	99.4	74.3
i,h	90.9	97.6	100	98.1	100	100	99.9
p,i,h	99.0	98.8	100	98.4	100	100	99.9

REFERENCES

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychological Association (1985). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

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

Brennan, R. L. (2006). Unbiased estimates of variance components with bootstrap procedures: Detailed results (CASMA Research Report No. 21). Iowa City: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

Brennan R. L. (2007). Unbiased estimates of variance components with bootstrap procedures. *Educational and Psychological Measurement*, *67*(5), 784-803.

Brennan, R. L., Harris, D. J., & Hanson, B. A. (1987). *The bootstrap and other procedures for examining the variability of estimated variance components in testing contexts* (ACT Research Report 87-7). Iowa City, IA: ACT, Inc.

Burch, V. C., Norman, G. R., Schmidt, H. G., & van der Vleuten, C. P. M. (2008). Are specialist certification examinations a reliable measure of physician competence? *Advances in Health Science Education, 13*, 521-533.

Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine, 19*(9), 1141-1164.

Chafouleas, S. M., Christ, T. J., & Riley-Tillman, T. C. (2009). Generalizability of scaling gradients on direct behavior ratings. *Educational and Psychological Measurement, 69*(1), 157-173.

Christophersen, K., Helseth, S., & Lund, T. (2008). A generalizability study of the Norwegian version of KINDL in a sample of healthy adolescents. Quality of Life Research, 17, 87-93.

Clauser, B. E., Harik, P., Margolis, M. J., McManus, I. C., Mollon, J., Chis, L., & Williams, S. (2009). An empirical examination of the impact of group discussion and examinee performance information on judgments made in the Angoff standard-setting procedure. *Applied Measurement in Education 22*, 1-21.

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

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. London: Chapman & Hall/CRC.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105-146). New York: American Council on Education and Macmillan.

Feng, W. (2002). *Applicability of the jackknife procedure for estimating standard errors of variance component estimates in selected random effects G study designs.* Unpublished doctoral dissertation, University of Iowa.

Follesdal, H., & Hagtvet, K. A. (2009). Emotional intelligence: The MSCEIT from the perspective of generalizability theory. *Intelligence 37,* 94-105.

Gagnon, R., Charlin, B., Lambert, C., Carrière, B., & Van der Vleuten, C. (2009). Script concordance testing: More cases or more questions? *Advances in Health Science Education, 14,* 367-375.

Gao, X., & Brennan, R.L. (2001). Variability of estimated variance components and related statistics in a performance assessment. *Applied Measurement in Education 14*(2), 191-203.

Harik, P., Clauser, B.E., Grabovsky, I., Nungester, R.J., Swanson, D., & Nandakumar, R. (2009). An examination of rater drift within a generalizability theory framework. *Journal of Educational Measurement 46*(1), 43-58.

Heitman, R. J., Kovaleski, J. E., Pugh, S. F. (2009). Application of generalizability theory in estimating the reliability of ankle-complex laxity measurement. *Journal of Athletic Training, 44*(1), 48-52.

Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments? A generalizability theory approach. *Assessing Writing, 13,* 201-218.

Lakes, K. D., & Hoyt, W. T. (2008). What sources contribute to variance in observer ratings? Using generalizability theory to assess construct validity of psychological measures. *Infant and Child Development, 17*, 269-284.

Lane, S., Liu, M., Ankenmann, R.D., & Stone, C.A. (1996). Generalizability and validity of a mathematics performance assessment. *Journal of Educational Measurement, 33*(1), 71-92.

Lane, S., Stone, C. A., Ankenmann, R. D., & Liu, M. (1994). Reliability and validity of a mathematics performance assessment. *International Journal of Educational Research, 21*(3), 247-266.

Lee, D. A., & Frisbie, D. A. (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets. *Applied Measurement in Education, 12*(3), 237-255.

Lee, Y., & Kantor, R. (2007). Evaluating prototype tasks and alternative rating schemes for a new ESL writing test through g-theory. *International Journal of Testing, 7*(4), 353-385.

Lee, G., & Lewis, D. M. (2008). A generalizability theory approach to standard error estimates for bookmark standard settings. *Educational and Psychological Measurement, 68*(4), 603-620.

Luecht, R. M., & Smith, P. L. (1989). *The effects of bootstrapping strategies on the estimation of variance components.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). New York: American Council on Education and Macmillan.

Moscoso, S. C., Tello, F. P. H., & Lopez, J. L, L. (2006). Using generalizability theory to assess the validity of the evaluation process. *Quality and Quantity, 40*, 315-329.

Murphy, D. J., Bruce, D. A., Mercer, S. W., & Eva, K. W. (2009). The reliability of workplace-based assessment in postgraduate medical education and training: A national evaluation in general practice in the United Kingdom. *Advances in Health Science Education 14*, 219-232.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Othman, A. R. (1995). *Examining task sampling variability in science performance assessments*. Unpublished Doctoral Dissertation, University of California, Santa Barbara.

Searle, S. R. (1971). *Linear Models*. New York, NY: Wiley.

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance Components.* New York, NY: Wiley.

Shao, J., & Tu, D. (1995). *The jackknife and bootstrap*. New York, NY: Springer-Verlag.

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement, 30*(3), 215-232.

Solano-Flores, G., & Li, M. (2009). Generalizability of cognitive interview-based measures across cultural groups. *Educational Measurement: Issues and Practice 28*(2), 9-18.

Tong, Y., & Brennan, R. L. (2006). *Bootstrap techniques for estimating variability in generalizability theory* (CASMA Research Report No. 15). Iowa City: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

Tong, Y., & Brennan, R. L. (2007). Bootstrap estimates of standard errors in generalizability theory. *Educational and Psychological Measurement, 67*(5), 804-817.

Turner, A. A., Lozano-Nieto, A., & Bouffard, M. (2003). Generalizability of bioimpedance spectroscopy measures. *Measurement in Physical Education and Exercise Science, 7*(3), 129-148.

Wiley, E. W. (2001). *Bootstrap strategies for variance component estimation: Theoretical and empirical results.* Unpublished Dissertation, Stanford University.

Yin, P., & Sconing, J. (2008). Estimating standard errors of cut scores for item rating and mapmark procedures: A generalizability theory approach. *Educational and Psychological Measurement, 68*(1), 25-41.

Yin, Y., & Shavelson, R. J. (2008). *Application of generalizability theory to concept map assessment research.* Applied Measurement in Education, 21, 273-291.