

Article

Extending Applications of Generalizability Theory-Based Bifactor Model Designs

Walter P. Vispoel ^{1,*}, Hyeryung Lee ¹, Tingting Chen ¹ and Hyeri Hong ²

¹ Department of Psychological and Quantitative Foundations, University of Iowa, Iowa City, IA 52242, USA; hyeryung-lee@uiowa.edu (H.L.); tingting-chen@uiowa.edu (T.C.)

² Department of Curriculum and Instruction, California State University, Fresno, CA 93740, USA; hyerihong@mail.fresnostate.edu

* Correspondence: walter-vispoel@uiowa.edu

Abstract: In recent years, researchers have described how to analyze generalizability theory (GT) based univariate, multivariate, and bifactor designs using structural equation models. However, within GT studies of bifactor models, variance components have been limited to those reflecting relative differences in scores for norm-referencing purposes, with only limited guidance provided for estimating key indices when making changes to measurement procedures. In this article, we demonstrate how to derive variance components for multi-facet GT-based bifactor model designs that represent both relative and absolute differences in scores for norm- or criterion-referencing purposes using scores from selected scales within the recently expanded form of the Big Five Inventory (BFI-2). We further develop and apply prophecy formulas for determining how changes in numbers of items, numbers of occasions, and universes of generalization affect a wide variety of indices instrumental in determining the best ways to change measurement procedures for specific purposes. These indices include coefficients representing score generalizability and dependability; scale viability and added value; and proportions of observed score variance attributable to general factor effects, group factor effects, and individual sources of measurement error. To enable readers to apply these techniques, we provide detailed formulas, code in R, and sample data for conducting all demonstrated analyses.

Keywords: generalizability theory; bifactor model; structural equation modeling; psychometrics; R programming; Big Five Factor Inventory; confidence intervals; prophecy formulas; score optimization; subscale added value



Citation: Vispoel, W.P.; Lee, H.; Chen, T.; Hong, H. Extending Applications of Generalizability Theory-Based Bifactor Model Designs. *Psych* **2023**, *5*, 545–575. <https://doi.org/10.3390/psych5020036>

Academic Editor: Alexander P. Christensen

Received: 29 April 2023

Revised: 25 May 2023

Accepted: 6 June 2023

Published: 13 June 2023

Corrected: 28 May 2024



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since its inception in the early 1960's [1] generalizability theory (GT) has provided an enduring framework for conceptualizing, evaluating, and improving the dependability of scores yielded by both objectively and subjectively scored measures within numerous disciplines. For example, when conducting a *PsycNet* database search at the time of writing using the key words “generalizability theory”, we recorded 2474 hits, with nearly half of them (1226) appearing in the research literature since 2012. Advantages of GT over previous measurement models include unambiguous definitions of the domains to which results are generalized, indices to reflect the extent to which results can be generalized to those domains, and estimation of how generalizability and dependability of scores might change when altering measurement procedures. Domains to which GT analyses have been applied within the last decade include medicine and health sciences [2–4], education [5–7], psychology (e.g., [8–10]), athletic training [11], management [12], communication [13], and many others.

Along with these recent applications of GT, we have seen an even greater surge in interest in the use of bifactor models [14,15], as reflected in the 1614 hits we recorded using the key words “bifactor model” in a parallel *PsycNet* database search over the same period (2012 to present) compared to only 27 hits before then. Bifactor models allow for

partitioning of reliable variance in scores into general and independent group factor effects to provide greater insights into the overall dimensionality of scores at composite and subscale levels and to gauge possible improvements gained when reporting subscale in addition to composite scores. Within the last five years alone, bifactor models have been used to represent mental abilities [16,17], motor skills [18], social skills [19], emotional intelligence [20], personality [21–25], psychological well-being [26], attention deficient hyperactivity disorders [27–29], and numerous other areas of functioning.

Until recently, applications of GT and bifactor models seldom overlapped due in part to GT designs typically being represented in ANOVA models and bifactor designs within factor analytic models. However, Vispoel, Lee, Xu, and Hong ([24,25]; also see [30]) demonstrated that GT and bifactor designs could be integrated together into structural equation models (SEMs) to allow for partitioning of universe score variance into general and group factor effects and measurement error into multiple sources. In this article, we extend the work of Vispoel and colleagues into GT-based bifactor models to allow for the derivation of consistency and agreement indices reflecting both relative and absolute differences in scores and demonstrate explicitly how changes in measurement procedures might affect bifactor model-based indices of generalizability, dependability, measurement error, scale viability, and subscale added value.

2. Background

2.1. GT-Based Bifactor Structural Equation Modeling

To illustrate the benefits and versatility of integrated GT and bifactor model analyses, we will consider a two-facet design that takes measurement error due to both item and occasion score differences into account. In GT, this would typically represent a *persons* \times *items* \times *occasions* random-effects design with items and occasions serving as facets corresponding to the universes to which results are generalized. The GT-based SEM for this design is depicted in Figure 1 for open-mindedness personality domain composite and subscale scores (aesthetic sensitivity, creative imagination, and intellectual curiosity) from the recently updated form of the Big Five Inventory (BFI-2, ref. [31]). The SEM has separate orthogonal factors for the 12-item composite scale representing the open-mindedness domain, for each of its nested 4-item subscales, for each individual item, and for each individual occasion.

The personality characteristics represented within the SEM are considered fixed because results are not generalized beyond those constructs, whereas items and occasions within corresponding scales are viewed as randomly sampled or exchangeable with those from broader domains of possible items and occasions. The open-mindedness composite factor is linked to all items on both occasions; separate factors for each subscale are linked only to items within that subscale on both occasions; separate factors for each occasion are linked to all items administered on that occasion; separate factors for each item are linked to all occasions; and uniquenesses are linked to each item on each occasion. Uniquenesses and loadings for the general (composite) factor, group (subscale) factors, and occasion factors are set as equal across occasions and equal within but not across subscales. Item variances also are set as equal within but not across subscales. Under these equality constraints, fifteen parameters (five per subscale) are estimated including three factor loadings (general, group, and occasion), one item variance, and one uniqueness for each subscale.

Once these parameters are estimated, they can be placed into the equations shown in Table 1 to compute variance components for general factor, group factor, and measurement error effects that subsequently can be inserted into additional formulas to compute indices of generalizability; dependability; proportional contributions of general factor, group factor, and measurement error effects; and scale viability and added value for the original GT design or ones altered to estimate effects of changes made to the measurement procedure (see Table 2). In applications of GT, derivation of variance components is part of what is called a calibration or *generalizability study*. Before estimating variance components for such a study, objects of measurement, universes of generalization, and admissible observations

would have been defined. In the present example, persons are the objects of measurement, and universes of generalization and admissible observations would include all possible or interchangeable items and occasions to measure the constructs open-mindedness, aesthetic sensitivity, creative imagination, and intellectual curiosity within the response structure of the BFI-2.

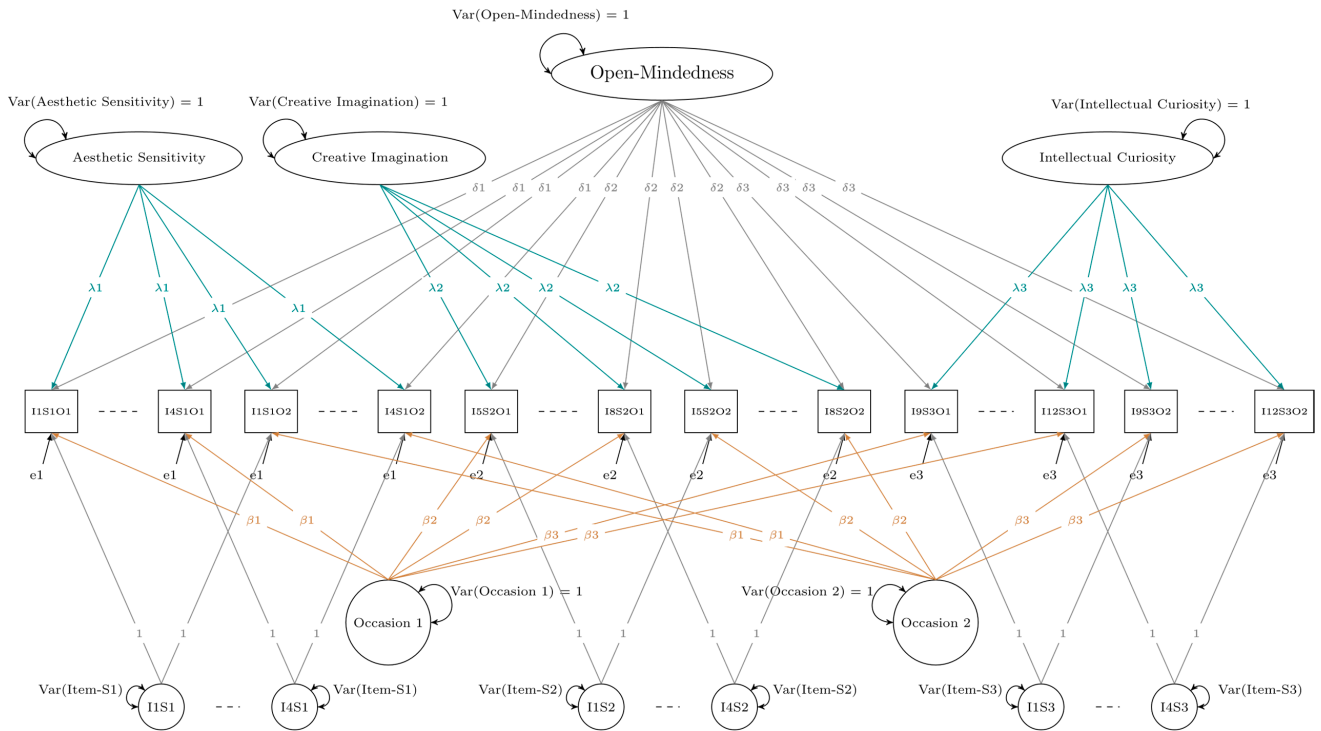


Figure 1. GT *persons* × *items* × *occasions* design bifactor structural equation model for open-mindedness composite and subscale scores (I = item, S = subscale, and O = occasion).

Table 1. Formulas for estimating item-level variance components for composite and subscale scores.

Composite	Subscale
$\hat{\sigma}_{general(C)}^2 = \left(\sum_{j=1}^{n_j} \left(\frac{n_{i(j)}}{n_i} \hat{\delta}_j \right) \right)^2$	$\hat{\sigma}_{general(j)}^2 = \hat{\delta}_j^2$
$\hat{\sigma}_{group(C)}^2 = \sum_{j=1}^{n_j} \left(\frac{n_{i(j)}}{n_i} \hat{\lambda}_j \right)^2$	$\hat{\sigma}_{group(j)}^2 = \hat{\lambda}_j^2$
$\hat{\sigma}_{pi(C)}^2 = \sum_{j=1}^{n_j} \left(\left(\frac{n_{i(j)}}{n_i} \right)^2 \hat{\sigma}_{pi(j)}^2 \right)$	$\hat{\sigma}_{pi(j)}^2 = \hat{\sigma}_{pi(j)}^2$
$\hat{\sigma}_{po(C)}^2 = \left(\sum_{j=1}^{n_j} \left(\frac{n_{i(j)}}{n_i} \hat{\beta}_j \right) \right)^2$	$\hat{\sigma}_{po(j)}^2 = \hat{\beta}_j^2$
$\hat{\sigma}_{pio,e(C)}^2 = \sum_{j=1}^{n_j} \left(\left(\frac{n_{i(j)}}{n_i} \right)^2 \hat{\sigma}_{pio,e(j)}^2 \right)$	$\hat{\sigma}_{pio,e(j)}^2 = \hat{\sigma}_{pio,e(j)}^2$
$\hat{\sigma}_{i(C)}^2 = \sum_{j=1}^{n_j} \left(\left(\frac{n_{i(j)}}{n_i} \right)^2 \frac{1}{n_{i(j)}-1} \sum_{i=1}^{n_{i(j)}} (item\ factor\ mean_i)^2 \right)$	$\hat{\sigma}_{i(j)}^2 = \frac{1}{n_{i(j)}-1} \sum_{i=1}^{n_{i(j)}} (item\ factor\ mean_i)^2$
$\hat{\sigma}_{o(C)}^2 = \sum_{j=1}^{n_j} \left(\left(\frac{n_{i(j)}}{n_i} \right)^2 \frac{1}{n_o-1} \sum_{i=1}^{n_o} (occasion\ factor\ mean_i)^2 \right)$	$\hat{\sigma}_{o(j)}^2 = \frac{1}{n_o-1} \sum_{i=1}^{n_o} (occasion\ factor\ mean_i)^2$
$\hat{\sigma}_{io(C)}^2 = \sum_{j=1}^{n_j} \left(\left(\frac{n_{i(j)}}{n_i} \right)^2 \frac{1}{(n_{i(j)}-1) \times (n_o-1)} \sum_{i=1}^{n_{i(j)} \times n_o} (intercept_i)^2 \right)$	$\hat{\sigma}_{io(j)}^2 = \frac{1}{(n_{i(j)}-1) \times (n_o-1)} \sum_{i=1}^{n_{i(j)} \times n_o} (intercept_i)^2$

Note. n_j : number of subscales, n_i : number of items in the composite scale, n_o : number of occasions, $n_{i(j)}$: number of items in the j th subscale.

Table 2. Prophecy formulas for key GT-based indices within *persons* × *items* × *occasions* designs.

Formula
$G \text{ coefficient} = \frac{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2)}{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2) + \left(\frac{\hat{\sigma}_{(pi)}^2}{n'_i} + \frac{\hat{\sigma}_{(po)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o} \right)}$
$\text{Proportion of general factor variance} = \frac{\hat{\sigma}_{(general)}^2}{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2) + \left(\frac{\hat{\sigma}_{(pi)}^2}{n'_i} + \frac{\hat{\sigma}_{(po)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o} \right)}$
$\text{Proportion of group factor variance} = \frac{\hat{\sigma}_{(group)}^2}{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2) + \left(\frac{\hat{\sigma}_{(pi)}^2}{n'_i} + \frac{\hat{\sigma}_{(po)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o} \right)}$
$\text{Global D coefficient} = \frac{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2)}{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2) + \left(\frac{\hat{\sigma}_{(pi)}^2 + \hat{\sigma}_{(i)}^2}{n'_i} + \frac{\hat{\sigma}_{(po)}^2 + \hat{\sigma}_{(o)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2 + \hat{\sigma}_{(io)}^2}{n'_i n'_o} \right)}$
$\text{Cut-score-specific D coefficient} = \frac{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2) + [(\bar{Y} - \text{Cut Score})^2 - \hat{\sigma}_{\bar{Y}}^2]}{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2) + [(\bar{Y} - \text{Cut Score})^2 - \hat{\sigma}_{\bar{Y}}^2] + \left(\frac{\hat{\sigma}_{(pi)}^2 + \hat{\sigma}_{(i)}^2}{n'_i} + \frac{\hat{\sigma}_{(po)}^2 + \hat{\sigma}_{(o)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2 + \hat{\sigma}_{(io)}^2}{n'_i n'_o} \right)}$
$\text{where } \hat{\sigma}_{\bar{Y}}^2 = \frac{\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2}{n'_p} + \frac{\hat{\sigma}_{(pi)}^2}{n'_p n'_i} + \frac{\hat{\sigma}_{(po)}^2}{n'_p n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_p n'_i n'_o} + \frac{\hat{\sigma}_{(i)}^2}{n'_i} + \frac{\hat{\sigma}_{(o)}^2}{n'_o} + \frac{\hat{\sigma}_{(io)}^2}{n'_i n'_o}$
$\text{Proportion of specific-factor error variance} = \frac{\frac{\hat{\sigma}_{(pi)}^2}{n'_i}}{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2) + \left(\frac{\hat{\sigma}_{(pi)}^2}{n'_i} + \frac{\hat{\sigma}_{(po)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o} \right)}$
$\text{Proportion of transient error variance} = \frac{\frac{\hat{\sigma}_{(po)}^2}{n'_o}}{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2) + \left(\frac{\hat{\sigma}_{(pi)}^2}{n'_i} + \frac{\hat{\sigma}_{(po)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o} \right)}$
$\text{Proportion of random-response error variance} = \frac{\frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o}}{(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2) + \left(\frac{\hat{\sigma}_{(pi)}^2}{n'_i} + \frac{\hat{\sigma}_{(po)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o} \right)}$
$\text{Value-added ratio} = \frac{\bar{G} \text{ coefficient}_{(subscale_j)}}{r_{U_{subscale_j} U_{composite}}^2 * G \text{ coefficient}_{(composite)}}, \text{ where } U = \text{universe score.}$

Note. Item-level variance components for composites and subscales from Table 1 are used within these formulas. Primes appear over *n*s in the equations to signify that they can be changed in decision studies.

Once variance components are derived from a generalizability study, they can be used in an application or *decision study* to estimate indices of generalizability, dependability, and measurement error for the original GT design or ones altered for possible changes made to the measurement procedure. For objectively scored self-report measures such as the BFI-2, common changes would include deriving the indices mentioned above for different numbers of items and/or occasions and limiting universes of generalization to just items or just occasions. In subsequent sections of this article, we will demonstrate how such changes affect indices of score consistency and measurement error but further extend those applications to include effects of the same changes on proportions of general and group factor variance and indices of scale viability and added value.

2.2. Indices of Generalizability, Dependability, Measurement Error, Viability, and Added Value

Generalizability and related coefficients. The most common estimates of score consistency reported in GT analyses are called *generalizability* (*G* or E_p^2) *coefficients* because they represent the extent to which results can be generalized to the targeted domain(s) or universes(s) of interest. These coefficients parallel conventional alpha, split-half, equivalent form, and test-retest reliability coefficients in that they represent relative differences in scores used for norm-referencing purposes such as rank ordering. Within the present bifactor design, a generalizability coefficient for aggregated item scores for persons would

represent the proportion of relative observed score variance accounted for by the effects shown in Equation (1).

$$\begin{aligned} & \text{G (or omega total) coefficient for bifactor designs} \\ &= \frac{\text{Universe score variance}}{\text{Sum of universe score and relative measurement error variances}} \\ &= \frac{\text{Sum of general factor and group factor variances}}{\text{Sum of general factor, group factor, and all relative error variances}} \end{aligned} \quad (1)$$

Within typical GT designs, a G coefficient would represent an estimate of the proportion of relative observed score variance within the domain(s) of interest accounted for by universe scores. Within a GT-bifactor design, universe scores are represented by the sum of general and group factor effects. In applications of bifactor models, G coefficients in Equation (1) would be labeled as *omega total* coefficients at composite and subscale levels. If the numerator of Equation (1) is replaced with just general or just group factor effects, we would create indices analogous to *omega hierarchical* coefficients often reported for bifactor models in the research literature (see, e.g., refs. [32–34]). *Omega hierarchical total* coefficients represent composite scores and include just general factor effects in the numerator of Equation 1, whereas *omega hierarchical subscale* coefficients represent subscale scores and include just group factor effects in the numerator. For present purposes, and to avoid confusion, we will usually describe these coefficients as representing proportions of general or group factor effects at either composite or subscale levels as shown in Equations (2) and (3).

$$\begin{aligned} & \text{Proportion of general factor variance for GT bifactor designs} \\ &= \frac{\text{General factor variance}}{\text{Sum of general factor, group factor, and all relative error variances}} \end{aligned} \quad (2)$$

$$\begin{aligned} & \text{Proportion of group factor variance(s) for GT bifactor designs} \\ &= \frac{\text{Group factor variance(s)}}{\text{Sum of general factor, group factor, and all relative error variances}} \end{aligned} \quad (3)$$

Sources of measurement error. When applying GT to measures of psychological traits such as those represented here, three primary sources of measurement error can affect scores: specific factor, transient, and random response. Specific-factor errors represent person-specific effects on scores unrelated to the targeted construct(s) that endure over occasions such as interpretations and understandings of words within items and response options. Transient errors represent unrelated effects on scores that are pervasive within an occasion but not across occasions. These temporary within-occasion effects relate to a respondent's disposition, mindset, and physiological condition; his or her reactions to administration and environmental factors; and other consistent entities that might affect behavior within the assessment setting that are unrelated to the constructs(s) being measured. Random-response errors reflect additional fleeting "within-occasion noise" effects that follow no systematic pattern (e.g., distractions, momentary lapses in attention, fluctuations in moods, changes in motivation, etc.; see, e.g., refs. [35–37]). In frameworks such as latent state-trait theory, specific-factor and transient error would, respectively, be described as *method* and *state* effects (see, e.g., refs. [38–40]).

Equations (4)–(6) can be used to estimate proportions of measurement error within GT-bifactor designs. They resemble Equations (2) and (3) but with the variance for the targeted source of measurement error represented in the numerator of the equation.

$$\begin{aligned} & \text{Proportion of specific – factor error variance for GT bifactor designs} \\ &= \frac{\text{Specific–factor error variance}}{\text{Sum of general factor, group factor, and all relative error variances}} \end{aligned} \quad (4)$$

$$\begin{aligned} & \text{Proportion of transient error variance for GT bifactor designs} \\ &= \frac{\text{Transient error variance}}{\text{Sum of general factor, group factor, and all relative error variances}} \end{aligned} \quad (5)$$

$$\begin{aligned} & \text{Proportion of random – response error variance for GT bifactor designs} \\ &= \frac{\text{Random–response error variance}}{\text{Sum of general factor, group factor, and all relative error variances}} \end{aligned} \quad (6)$$

Dependability coefficients. All equations described until this point reflect proportions of relative variance in scores catered to norm-referenced uses that do not depend on the absolute levels of item or occasion mean scores. However, when absolute levels of scores are used directly for screening, selection, classification, or domain referencing purposes, mean differences in item and/or occasion scores selected from the universe(s) of generalization pertinent in those applications could affect the magnitude of observed scores and thereby directly impact those decisions. In GT, two general types of *dependability* (D or Φ) *coefficients* are used to take absolute differences in scores into account: global and cut-score specific [41,42].

Kane and Brennan [42] characterized global D coefficients as representing the contribution of the assessment procedure to the overall dependability of scores when making criterion-referenced decisions and cut-score specific D coefficients as reflecting the contribution of the assessment procedure to the decision made from the cut score over what would be expected by chance agreement (p. 110). Both types of indices can vary from 0 to 1, with higher values representing greater dependability. When deriving indices for dependability in making such decisions with the present measures, additional variance components are needed that quantify differences in item and occasion mean scores. We will soon show that these components are then added to the denominators of the equations for dependability coefficients to broaden the definition of observed score variance and overall error to include mean differences in item and occasion scores.

Until very recently, SEMs for doing GT analyses were limited almost exclusively to those already discussed for deriving variance components reflecting only relative inter-person differences in scores. However, Jorgensen [43] demonstrated that additional variance components representing differences in item and occasion mean scores could be obtained within univariate designs by imposing effect coding [44] and related constraints on factor loadings, factor means, and intercepts. When implementing these procedures here, we constrained general, group, and occasion factor variances to equal one and restricted intercepts, item factor means, and occasion factor means each to the sum of zero. Under these conditions, Equations (7)–(9) can be used to derive variance components for items, occasions, and their interaction for each subscale within the present design.

$$\hat{\sigma}_i^2 = \frac{1}{n_i - 1} \sum_1^{n_i} (\text{Item factor mean}_i)^2, \quad (7)$$

$$\hat{\sigma}_o^2 = \frac{1}{n_o - 1} \sum_1^{n_o} (\text{Occasion factor mean}_o)^2, \quad (8)$$

$$\hat{\sigma}_{io}^2 = \frac{1}{(n_i \times n_o) - 1} \sum_1^{n_i \times n_o} (\text{Intercept}_{io})^2, \quad (9)$$

where n_i = number of items and n_o = number of occasions.

Corresponding variance components for the composite score can be obtained from the variance components for its nested subscales using Equations (10)–(12).

$$\hat{\sigma}_{i(C)}^2 = \sum_{j=1}^{n_j} \left(\frac{n_{i(j)}}{n_I} \right)^2 \frac{1}{n_{i(j)} - 1} \sum_{i=1}^{n_{i(j)}} (\text{item factor mean}_i)^2, \quad (10)$$

$$\hat{\sigma}_{o(C)}^2 = \sum_{j=1}^{n_j} \left(\frac{n_{i(j)}}{n_I} \right)^2 \frac{1}{n_o - 1} \sum_{i=1}^{n_o} (\text{occasion factor mean}_i)^2, \quad (11)$$

$$\hat{\sigma}_{io(C)}^2 = \sum_{j=1}^{n_j} \left(\frac{n_{i(j)}}{n_I} \right)^2 \frac{1}{(n_{i(j)} - 1) \times (n_o - 1)} \sum_{i=1}^{n_{i(j)} \times n_o} (\text{intercept}_i)^2, \quad (12)$$

where n_j = number of subscales, n_o : number of occasions, $n_{i(j)}$ = number of items per subscale, and n_I = total number of items in the composite scale.

Once all relevant variance components are derived, they can be inserted into Equations (13) and (14) to derive global and cut-score specific D coefficients for observed scores within the present design.

$$\begin{aligned} & \text{Global D coefficient for GT bifactor designs} \\ & = \frac{\text{Universe score variance}}{\text{Sum of universe score and error variances for relative and absolute differences in scores}'} \end{aligned} \quad (13)$$

$$\begin{aligned} & \text{Cut – score specific D coefficient for GT bifactor designs} \\ & = \frac{\text{Universe score variance} + (\text{Grand Mean} - \text{Cut-score})^2}{\text{Universe score variance} + (\text{Grand Mean} - \text{Cut-score})^2 + \text{sum of error variances for relative and absolute differences in scores}'} \end{aligned} \quad (14)$$

More extended versions of Equations (13) and (14) with relevant variance components and adjustments for bias appear in Table 2 and are illustrated further in our online Supplemental Material.

Scale viability and added value. A wide variety of procedures have been discussed in the research literature for assessing scale viability and added value when reporting subscale in addition to composite scores. We discuss two general methods here because they are widely used, can be extended to universe score estimation in GT designs, and be re-estimated for changes made to a measurement procedure. The first procedure involves estimation of the proportion of combined general and group factor variances (i.e., universe score variance here) accounted for by general factor effects alone (see Equation (15)). In applications of bifactor models, this index is called *explained common variance* (ECV; refs. [24,25,32–34,45]). Replacing the numerator of Equations (15) with group factor variance(s) would yield a similar index representing the proportion of combined general and group factor variances accounted for by group factor effects alone (see Equation (16)). We will refer to this coefficient as *explained unique variance* (EUV; ref. [25]). Finally, a ratio can be created to represent relative proportions of common and unique explained variance by dividing ECV by EUV as shown in Equation (17). The higher this ratio is, the more redundant subscale scores are with composite scores.

$$\text{Explained common variance (ECV)} = \frac{\text{General factor variance}}{\text{General factor variance} + \text{Group factor variance(s)}} \quad (15)$$

$$\text{Explained unique variance (EUV)} = \frac{\text{Group factor variance(s)}}{\text{General factor variance} + \text{Group factor variance(s)}} \quad (16)$$

$$\text{Common to unique explained variance ratio} = \frac{\text{Explained common variance}}{\text{Explained unique variance}} = \frac{\text{ECV}}{\text{EUV}} \quad (17)$$

When considered by themselves, the viability of composite and subscale scores would be best supported by high values for G and global D coefficients coupled, respectively, with high ECV coefficients for composites and high EUV coefficients for subscales. Reporting both composite and subscale scores would be best supported by high values for G and global D coefficients and more of a balance in values for ECV and EUV coefficients [26].

The second procedure we describe was developed by Haberman ([46]; additionally, see [47,48]) to determine whether a subscale's observed scores better estimate the subscale's true scores than do the composite's observed scores. Vispoel, Lee, Hong, and Chen [49] later adapted Haberman's procedure to GT designs by substituting universe score for true score estimation. Haberman's [46] original procedure required the calculation of indices representing proportional reductions in mean squared error (PRMSE) when estimating the subscale's true scores based on observed scores from the subscale and composite. A subscale would demonstrate added value if its PRMSE estimate exceeded that for its associated composite. To simplify this procedure further, Feinberg and Wainer [50] recommended forming the *value-added ratio* (VAR) shown in Equation (18) in which the

PRMSE for the subscale is divided by the PRMSE for the composite. A subscale's added value is increasingly supported as its VAR deviates upwardly from one.

$$\text{Value – Added Ratio (VAR)} = \frac{\text{PRMSE}(\text{subscale})}{\text{PRMSE}(\text{composite})} \quad (18)$$

The PRMSE index for a subscale reduces to its reliability coefficient (conventional or GT-based), and the corresponding PRMSE index for its composite can be estimated using Equation (19).

$$\text{PRMSE}_{(\text{composite})} = r_{U_{\text{subscale}}; U_{\text{composite}}}^2 * G \text{ coefficient}_{(\text{composite})}, \text{ where } U = \text{universe score.} \quad (19)$$

2.3. Confidence Intervals

Applications of GT are based on three primary assumptions: (a) the universe(s) of generalization is/are clearly defined, (b) facet conditions are experimentally independent, and (c) scores are expressed on equal interval metrics [1] (p. 145). Consistent with ANOVA procedures that form the foundation for GT analyses, facet conditions (items, occasions, raters, etc.) are typically treated as being unordered [1]. As previously noted, facet conditions included in the generalizability study also are considered randomly sampled from or exchangeable with others within the broader universe(s) from which they are drawn. However, because no explicit assumptions are made about the content of the universe or statistical properties of scores, fit indices for the overall GT-SEM are not required. Instead, Monte Carlo confidence intervals can be built around estimates of variance components, G coefficients, D coefficients, and proportions of measurement error to evaluate their trustworthiness. An advantage of doing GT analyses using the *lavaan* SEM package in R [51,52] is that results can be linked to the *semTools* package [53] to build such intervals to any desired degree of confidence. Accordingly, we provide 95% confidence intervals for all relevant indices when reporting results for BFI-2 open-mindedness composite and subscales later in this article.

2.4. Changing Measurement Procedures

One of the key attributes of GT is the inclusion of methods to estimate generalizability and dependability indices for changes made to a measurement procedure. Typical changes to objectively scored measures, such as those used here within the *persons* × *items* × *occasions* random-effects design, are to alter numbers of items or occasions and/or limit universes of generalization to just items or just occasions. Effects of changes to numbers of items and occasions on G and D coefficients are estimated simply by substituting those values in the equations used to compute those coefficients (see Table 2). The same substitutions can be made to other formulas presented here for proportions of general factor, group factor, and measurement error effects as shown in that table as well. Determining values for these indices in more restricted universes would entail treating the measurement error index for the excluded facet as universe score variance and limiting variance components for absolute differences in scores to only those facets retained in the universe(s) of generalization. We present formulas representing key indices for GT bifactor designs restricted to universes of just items (i.e., *persons* × *items* designs) in Table 3 and for designs restricted to universes of just occasions (i.e., *persons* × *occasions* designs) in Table 4. In sections to follow, we demonstrate a wide variety of these and other changes to measurement procedures within multi-facet GT-bifactor designs.

Table 3. Prophecy formulas for key GT-based indices within restricted *persons* × *items* designs.

Formula
$G \text{ coefficient} = \frac{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(po)}^2}{n'_o} \right)}{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(po)}^2}{n'_o} \right) + \left(\frac{\hat{\sigma}_{(pi)}^2}{n'_i} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o} \right)}$
$\text{Global D coefficient} = \frac{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(po)}^2}{n'_o} \right)}{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(po)}^2}{n'_o} \right) + \left(\frac{\hat{\sigma}_{(pi)}^2 + \hat{\sigma}_{(i)}^2}{n'_i} + \frac{\hat{\sigma}_{(pio,e)}^2 + \hat{\sigma}_{(io)}^2}{n'_i n'_o} \right)}$
$\text{Cut-score-specific D coefficient} = \frac{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 \right) + \left[(\bar{Y} - \text{Cut Score})^2 - \hat{\sigma}_{\bar{Y}}^2 \right]}{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 \right) + \left[(\bar{Y} - \text{Cut Score})^2 - \hat{\sigma}_{\bar{Y}}^2 \right] + \left(\frac{\hat{\sigma}_{(pi)}^2 + \hat{\sigma}_{(i)}^2}{n'_i} + \frac{\hat{\sigma}_{(po)}^2 + \hat{\sigma}_{(o)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2 + \hat{\sigma}_{(io)}^2}{n'_i n'_o} \right)}$
$\text{where } \hat{\sigma}_{\bar{Y}}^2 = \frac{\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2}{n'_p} + \frac{\hat{\sigma}_{(pi)}^2}{n'_p n'_i} + \frac{\hat{\sigma}_{(po)}^2}{n'_p n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_p n'_i n'_o} + \frac{\hat{\sigma}_{(i)}^2}{n'_i} + \frac{\hat{\sigma}_{(io)}^2}{n'_i n'_o}$
$\text{Total error} = \frac{\frac{\hat{\sigma}_{(pi)}^2}{n'_i} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o}}{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(po)}^2}{n'_o} \right) + \left(\frac{\hat{\sigma}_{(pi)}^2}{n'_i} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o} \right)}$
$\text{Value-added ratio} = \frac{G \text{ coefficient}_{(subscale_j)}}{r_{U_{subscale_j} U_{composite}}^2 * G \text{ coefficient}_{(composite)}}, \text{ where } U = \text{universe score.}$

Note. Item-level variance components for composites and subscales from Table 1 are used within these formulas. Primes appear over *n*s in the equations to signify that they can be changed in decision studies.

Table 4. Prophecy formulas for key GT-based indices within restricted *persons* × *occasions* designs.

Formula
$G \text{ coefficient} = \frac{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(pi)}^2}{n'_i} \right)}{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(pi)}^2}{n'_i} \right) + \left(\frac{\hat{\sigma}_{(po)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o} \right)}$
$\text{Global D coefficient} = \frac{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(pi)}^2}{n'_i} \right)}{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(pi)}^2}{n'_i} \right) + \left(\frac{\hat{\sigma}_{(po)}^2 + \hat{\sigma}_{(o)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2 + \hat{\sigma}_{(io)}^2}{n'_i n'_o} \right)}$
$\text{Cut-score-specific D coefficient} = \frac{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(pi)}^2}{n'_i} \right) + \left[(\bar{Y} - \text{Cut Score})^2 - \hat{\sigma}_{\bar{Y}}^2 \right]}{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(pi)}^2}{n'_i} \right) + \left[(\bar{Y} - \text{Cut Score})^2 - \hat{\sigma}_{\bar{Y}}^2 \right] + \left(\frac{\hat{\sigma}_{(po)}^2 + \hat{\sigma}_{(o)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2 + \hat{\sigma}_{(io)}^2}{n'_i n'_o} \right)}$
$\text{where } \hat{\sigma}_{\bar{Y}}^2 = \frac{\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2}{n'_p} + \frac{\hat{\sigma}_{(pi)}^2}{n'_p n'_i} + \frac{\hat{\sigma}_{(po)}^2}{n'_p n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_p n'_i n'_o} + \frac{\hat{\sigma}_{(o)}^2}{n'_o} + \frac{\hat{\sigma}_{(io)}^2}{n'_i n'_o}$
$\text{Total error} = \frac{\frac{\hat{\sigma}_{(po)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o}}{\left(\hat{\sigma}_{(general)}^2 + \hat{\sigma}_{(group)}^2 + \frac{\hat{\sigma}_{(pi)}^2}{n'_i} \right) + \left(\frac{\hat{\sigma}_{(po)}^2}{n'_o} + \frac{\hat{\sigma}_{(pio,e)}^2}{n'_i n'_o} \right)}$
$\text{Value-added ratio} = \frac{G \text{ coefficient}_{(subscale_j)}}{r_{U_{subscale_j} U_{composite}}^2 * G \text{ coefficient}_{(composite)}}, \text{ where } U = \text{universe score.}$

Note. Item-level variance components for composites and subscales from Table 1 are used within these formulas. Primes appear over *n*s in the equations to signify that they can be changed in decision studies.

3. Purpose

In the remainder of this article, we illustrate how GT can be applied to multi-facet bifactor model designs to derive variance components for general factor, group factor, and measurement error effects and show how these indices can be used to assess the generalizability, dependability, viability, and added value of scale scores for data at hand and for

possible changes made to the measurement procedure. We first describe the respondent sample and measures used, then the analyses run, and finally the results obtained for partitioning of variance, G coefficients, D coefficients, and scale viability/added value indices within numerous GT bifactor designs.

4. Methods

4.1. Sample, Procedures, and Measures

Our sample consisted of 389 students from a large Midwestern University (71.72% female, 70.95% Caucasian, mean age = 20.38) who completed the BFI-2 [31] on the Qualtrics online platform on two occasions a week apart. Students were compensated by receiving extra credit points within the classes from which they were recruited. The study (ID# 200809738, 8 August 2022) was approved by the presiding Institutional Review Board, and all respondents provided informed consent before participating. The study was not preregistered, and questions about access to the data should be directed to the first author.

BFI-2: The BFI-2 [31] is a recently expanded version of the Big Five Inventory (BFI; [54]). When creating the BFI-2, Soto and John sought to retain the focus, efficiency, and clarity of the BFI but improve it by more accurately representing the hierarchical structure of traits nested within each global personality domain, balancing the bandwidth and fidelity of scores within all scales, and reducing the influence of acquiescence by content balancing all domain and subdomain/facet scales for negative and positive wording. We chose the domain open-mindedness and its nested subdomain facet subscales to illustrate applications of GT bifactor designs here but the same techniques can be applied to other personality composite and subscale scores within the BFI-2 or those from any other instrument that assesses hierarchically structured constructs (see, e.g., refs. [24,25]).

As previously noted, the open-mindedness domain composite scale has 12 items that are subdivided into three nested 4-item subscales representing the personality subdomain facets: aesthetic sensitivity, creative imagination, and intellectual curiosity. Items for the composite and each subscale are equally balanced for positive and negative phrasing and answered along a 5-point Likert-style response metric (1 = disagree strongly, 2 = disagree a little, 3 = neutral, no opinion, 4 = agree a little, and 5 = agree strongly). Using data collected from internet ($n = 1000$) and college student ($n = 470$) samples, Soto and John [31] reported respective alpha reliability estimates of 0.84 and 0.85 for open-mindedness, 0.67 and 0.73 for aesthetic sensitivity, 0.76 and 0.83 for creative imagination, and 0.74 and 0.77 for intellectual curiosity. Corresponding 8-week test-retest coefficients for these same scales for a subset of students ($n = 110$) from their college sample, respectively, equaled 0.76, 0.78, 0.67, and 0.67. Evidence provided by the authors in support of the validity of open-mindedness composite and subscale scores within both samples included logically consistent patterns of discriminant and convergent validity coefficients with other measures of personality and related constructs, and adequate fits for confirmatory correlated multifactor models for subscale scores when acquiescence effects were controlled.

4.2. Analyses

Analyses reported here entailed derivation of descriptive statistics (means, standard deviations) and conventional reliability coefficients (alpha, test-retest) for open-mindedness composite and subscale scores, followed by variance components; G coefficients; D coefficients; proportions of variance attributable to general factor, group factor, and individual measurement error effects; and indices of scale viability and added value. Nine complete *persons* \times *items* \times *occasions* random-effects designs were analyzed that had varying numbers of items and occasions to evaluate their effects on key indices. These designs included: (1) 4 items per subscale and 1 occasion (baseline), (2) 4 items per subscale and 2 occasions, (3) 4 items per subscale and 3 occasions, (4) 8 items per subscale and 1 occasion, (5) 8 items per subscale and 2 occasions, (6) 8 items per subscale and 3 occasions, (7) 12 items per subscale and 1 occasion, (8) 12 items per subscale and 2 occasions, and (9) 12 items per subscale and 3 occasions. To demonstrate parallel effects when restricting universes to just

items or just occasions, we analyzed item-only (i.e., *persons* \times *items*) designs based on 4, 8, and 12 items per subscale, and occasion-only (i.e., *persons* \times *occasions*) designs based on 1, 2, and 3 occasions. Results were derived using the *psych* [55], *lavaan* [51,52], and *semTools* [53] packages in R. In keeping with ANOVA applications of GT, we used unweighted least squares (ULS) parameter estimates in all SEMs we analyzed.

5. Results

5.1. Descriptive Statistics and Conventional Reliability Estimates

Table 5 includes means, standard deviations, and conventional reliability estimates (alpha, test–retest) for BFI-2 open-mindedness composite and subscale scores. Across scales and occasions, means on the item-score metric fall well above the scale midpoint value of 3.0, ranging from 3.633 to 3.844. These results reflect positive overall levels of endorsement for the measured traits within this sample. Standard deviations on the same metric range from 0.679 to 0.923, with open-mindedness and aesthetic sensitivity, respectively, showing the least and most variability in average item scores within both occasions. As would be expected, alpha reliability estimates are higher for the 12-item open-mindedness composite (0.837 and 0.855) than for its nested 4-item subscales ($M_s = 0.709$ and 0.729), and a similar pattern is evident for 1-week test–retest coefficients (0.856 for the composite and $M = 0.754$ for subscales).

Table 5. Means, standard deviations, and reliability estimates for BFI-2 open-mindedness composite and subscale scores.

Occasion/Index	Composite/Subscale				
	Open-Mindedness	Aesthetic Sensitivity	Creative Imagination	Intellectual Curiosity	Subscale Average
Number of Items	12	4	4	4	4
Time 1					
Mean: Scale (Item)	44.483 (3.707)	14.576 (3.644)	15.375 (3.844)	14.532 (3.633)	14.828 (3.707)
SD: Scale (Item)	8.145 (0.679)	3.693 (0.923)	3.005 (0.751)	3.230 (0.808)	3.309 (0.827)
Alpha	0.837	0.730	0.671	0.725	0.709
Time 2					
Mean: Scale (Item)	44.290 (3.691)	14.553 (3.638)	15.185 (3.796)	14.553 (3.638)	14.763 (3.691)
SD: Scale (Item)	8.212 (0.684)	3.719 (0.930)	3.036 (0.759)	3.071 (0.768)	3.275 (0.819)
Alpha	0.855	0.769	0.696	0.723	0.729
Test-retest	0.856	0.828	0.793	0.759	0.794

5.2. GT Bifactor Designs including Both Item and Occasion Effects

Variance components. In Table 6, we present variance components for the *persons* \times *items* \times *occasions* random-effects GT bifactor design expressed on the item-score metric for the BFI-2 open-mindedness, aesthetic sensitivity, creative imagination, and intellectual curiosity scales. Confidence intervals for all variance components fail to capture zero except those for *o* and *io* within each scale. These results replicate findings from previous GT studies of BFI-2 scores (see, e.g., ref. [56]) and make sense because we did not expect occasion means or relative differences in the magnitude or order of item score means to vary much over the 1-week gap in administrations of the current trait-based measures.

Partitioning of variance. In Table 7, we report proportions of universe score (i.e., G or ω total coefficients), general factor, group factor, specific-factor error, transient error, and random-response error variance for GT bifactor designs varying in number of items per subscale and number of occasions. The first design (Design 1), with number of items per subscale equaling 4 and number of occasions equaling 1, reflects the typical situation in which the BFI-2 is administered in its original form on one occasion but with GT techniques used to account for multiple sources of measurement error. This model serves as a baseline for determining effects when numbers of items and/or occasions are increased.

Table 6. Variance components for BFI-2 open-mindedness composite and subscale scores for full persons \times items \times occasions designs.

Variance Component	Composite/Subscale				
	Open-Mindedness	Aesthetic Sensitivity	Creative Imagination	Intellectual Curiosity	Subscale Average
$\hat{\sigma}_{(general)}^2$	0.323 (0.314, 0.333)	0.355 (0.326, 0.386)	0.274 (0.254, 0.296)	0.343 (0.316, 0.371)	0.324
$\hat{\sigma}_{(group)}^2$	0.043 (0.039, 0.049)	0.256 (0.224, 0.289)	0.116 (0.089, 0.147)	0.015 (0.000, 0.062)	0.129
$\hat{\sigma}_{(pi)}^2$	0.126 (0.116, 0.137)	0.436 (0.381, 0.491)	0.328 (0.272, 0.384)	0.373 (0.317, 0.428)	0.379
$\hat{\sigma}_{(po)}^2$	0.035 (0.024, 0.048)	0.021 (0.005, 0.048)	0.056 (0.027, 0.095)	0.033 (0.011, 0.065)	0.036
$\hat{\sigma}_{(pio,e)}^2$	0.127 (0.115, 0.140)	0.435 (0.370, 0.500)	0.361 (0.291, 0.431)	0.348 (0.281, 0.415)	0.381
$\hat{\sigma}_{(i)}^2$	0.020 (0.017, 0.025)	0.074 (0.054, 0.099)	0.014 (0.007, 0.026)	0.094 (0.071, 0.121)	0.061
$\hat{\sigma}_{(o)}^2$	0.001 (0.000, 0.014)	0.003 (0.000, 0.042)	0.003 (0.000, 0.042)	0.003 (0.000, 0.042)	0.003
$\hat{\sigma}_{(io)}^2$	0.000 (0.000, 0.001)	0.000 (0.000, 0.004)	0.001 (0.000, 0.006)	0.000 (0.000, 0.005)	0.001

Note. p = person, i = item, o = occasion, and e = other residual error. Values within parentheses in the body of the table represent 95% confidence interval limits.

Results in Table 7 reveal that the baseline design (Design 1) yields G coefficients for all scales lower than the conventional (alpha and test-retest) reliability indices described in the previous section, which would be expected given that the G coefficients take both item and occasion effects into account. As was the case with conventional reliability indices, the G coefficient for open-mindedness composite scores (0.789) is higher than the G coefficients for its nested subscale scores ($M = 0.659$). For all scales, each source of measurement error accounts for noteworthy proportions of observed score variance ranging from 0.068 to 0.163 for specific factor, from 0.024 to 0.090 for transient, and from 0.068 to 0.152 for random response. Confidence intervals for all proportions of measurement error fail to capture zero, thereby reflecting trustworthy effects. Proportions of general factor effects exceed proportions of group factor effects for all scales, with aesthetic sensitivity showing the best balance of general (0.418) and group (0.301) factor effects, and intellectual curiosity showing the worst (0.601 vs. 0.027).

The remaining designs in Table 7 represent effects of doubling or tripling numbers of items and/or occasions. In general, G coefficients, proportions of general factor effects, and proportions of group factor effects increase, whereas overall proportions of measurement error decrease when increasing either numbers of items or occasions. Across all designs, confidence intervals for general and group factor effects fail to capture zero except group effects for intellectual curiosity. In relation to measurement error, the most noticeable effects of increasing items are to lower specific-factor and random-response error, and the most noticeable effects of increasing occasions are to lower transient and random-response error. In all cases, confidence intervals for each source of measurement error do not capture zero, again highlighting trustworthy estimates and the importance of taking each source of error into account.

In Figure 2, we provide prophecy graphs for the aesthetic sensitivity subscale representing G coefficients; proportions of general factor, group factor, random-response error, specific-factor error, and transient error variance; and global D coefficients for number of items ranging from 4 to 12 and number of occasions ranging from 1 to 3. We present parallel graphs for the other scales analyzed here in our online Supplemental Material. Figure 2 shows that, as numbers of items and/or occasions increase, the magnitude of G coefficients, global D coefficients, proportions of general factor variance, and proportions of group factor variance increase, whereas proportions of random-response error variance decrease. However, the magnitudes of these changes steadily diminish with the same progressive incremental changes in numbers of items or occasions. Proportions of specific-factor error decrease most noticeably with increases in numbers of items, and proportions of transient error do so with increases in numbers of occasions.

Table 7. Partitioning of G coefficient denominator variance for BFI-2 open-mindedness composite and subscale scores within *persons* × *items* × *occasions* full designs.

Design/Scale	G (US)	Gen	Grp	Index (CI) SFE	TE	RRE	TREI
Design 1: <i>i</i>(s) = 4, <i>o</i> = 1							
Open-Mindedness	0.789 (0.764, 0.811)	0.696 (0.670, 0.718)	0.093 (0.084, 0.105)	0.068 (0.062, 0.074)	0.075 (0.051, 0.102)	0.068 (0.061, 0.075)	0.211 (0.189, 0.236)
Aesthetic Sensitivity	0.719 (0.692, 0.740)	0.418 (0.378, 0.456)	0.301 (0.266, 0.335)	0.128 (0.111, 0.145)	0.024 (0.006, 0.056)	0.128 (0.109, 0.147)	0.281 (0.260, 0.308)
Creative Imagination	0.632 (0.583, 0.673)	0.444 (0.402, 0.483)	0.188 (0.145, 0.234)	0.133 (0.109, 0.156)	0.090 (0.043, 0.149)	0.146 (0.117, 0.174)	0.368 (0.327, 0.417)
Intellectual Curiosity	0.628 (0.585, 0.667)	0.601 (0.531, 0.646)	0.027 (0.000, 0.103)	0.163 (0.136, 0.187)	0.057 (0.019, 0.110)	0.152 (0.121, 0.180)	0.372 (0.333, 0.415)
Subscale Average	0.659	0.487	0.172	0.141	0.057	0.142	0.341
Design 2: <i>i</i>(s) = 4, <i>o</i> = 2							
Open-Mindedness	0.849 (0.834, 0.864)	0.750 (0.730, 0.766)	0.100 (0.090, 0.113)	0.073 (0.067, 0.079)	0.040 (0.027, 0.056)	0.037 (0.033, 0.040)	0.151 (0.136, 0.166)
Aesthetic Sensitivity	0.779 (0.758, 0.796)	0.453 (0.413, 0.492)	0.326 (0.288, 0.363)	0.139 (0.121, 0.156)	0.013 (0.003, 0.031)	0.069 (0.059, 0.080)	0.221 (0.204, 0.242)
Creative Imagination	0.716 (0.677, 0.751)	0.503 (0.461, 0.543)	0.213 (0.166, 0.263)	0.150 (0.124, 0.176)	0.051 (0.024, 0.087)	0.083 (0.066, 0.099)	0.284 (0.249, 0.323)
Intellectual Curiosity	0.701 (0.667, 0.735)	0.671 (0.597, 0.714)	0.030 (0.000, 0.115)	0.182 (0.153, 0.208)	0.032 (0.011, 0.063)	0.085 (0.067, 0.101)	0.299 (0.265, 0.333)
Subscale Average	0.732	0.542	0.190	0.157	0.032	0.079	0.268
Design 3: <i>i</i>(s) = 4, <i>o</i> = 3							
Open-Mindedness	0.872 (0.860, 0.883)	0.769 (0.752, 0.783)	0.102 (0.093, 0.116)	0.075 (0.069, 0.081)	0.028 (0.019, 0.038)	0.025 (0.023, 0.028)	0.128 (0.117, 0.140)
Aesthetic Sensitivity	0.801 (0.781, 0.818)	0.466 (0.425, 0.505)	0.335 (0.296, 0.374)	0.143 (0.125, 0.161)	0.009 (0.002, 0.021)	0.048 (0.040, 0.055)	0.199 (0.182, 0.219)
Creative Imagination	0.750 (0.714, 0.782)	0.526 (0.483, 0.567)	0.223 (0.175, 0.275)	0.157 (0.130, 0.184)	0.035 (0.017, 0.061)	0.058 (0.046, 0.069)	0.250 (0.218, 0.286)
Intellectual Curiosity	0.729 (0.698, 0.762)	0.698 (0.622, 0.741)	0.031 (0.000, 0.119)	0.190 (0.159, 0.216)	0.022 (0.007, 0.044)	0.059 (0.047, 0.070)	0.271 (0.238, 0.302)
Subscale Average	0.760	0.563	0.197	0.163	0.022	0.055	0.240
Design 4: <i>i</i>(s) = 8, <i>o</i> = 1							
Open-Mindedness	0.846 (0.819, 0.870)	0.747 (0.718, 0.771)	0.099 (0.090, 0.112)	0.036 (0.033, 0.040)	0.081 (0.055, 0.109)	0.037 (0.033, 0.040)	0.154 (0.130, 0.181)
Aesthetic Sensitivity	0.825 (0.793, 0.846)	0.480 (0.432, 0.524)	0.345 (0.307, 0.383)	0.074 (0.064, 0.083)	0.028 (0.006, 0.064)	0.073 (0.062, 0.084)	0.175 (0.154, 0.207)
Creative Imagination	0.734 (0.675, 0.783)	0.515 (0.464, 0.564)	0.219 (0.168, 0.271)	0.077 (0.063, 0.091)	0.104 (0.051, 0.172)	0.085 (0.067, 0.102)	0.266 (0.217, 0.325)
Intellectual Curiosity	0.745 (0.692, 0.789)	0.713 (0.623, 0.769)	0.032 (0.000, 0.121)	0.097 (0.080, 0.112)	0.068 (0.023, 0.129)	0.090 (0.071, 0.108)	0.255 (0.211, 0.308)
Subscale Average	0.768	0.569	0.199	0.082	0.067	0.083	0.232
Design 5: <i>i</i>(s) = 8, <i>o</i> = 2							
Open-Mindedness	0.899 (0.883, 0.913)	0.793 (0.772, 0.810)	0.106 (0.096, 0.119)	0.039 (0.035, 0.042)	0.043 (0.029, 0.059)	0.019 (0.018, 0.021)	0.101 (0.087, 0.117)
Aesthetic Sensitivity	0.869 (0.849, 0.884)	0.506 (0.460, 0.549)	0.363 (0.322, 0.404)	0.078 (0.067, 0.088)	0.015 (0.003, 0.034)	0.039 (0.033, 0.045)	0.131 (0.116, 0.151)
Creative Imagination	0.811 (0.769, 0.845)	0.569 (0.520, 0.616)	0.241 (0.189, 0.296)	0.085 (0.069, 0.100)	0.058 (0.027, 0.098)	0.047 (0.037, 0.056)	0.189 (0.155, 0.231)
Intellectual Curiosity	0.809 (0.772, 0.840)	0.774 (0.683, 0.822)	0.035 (0.000, 0.132)	0.105 (0.087, 0.121)	0.037 (0.012, 0.073)	0.049 (0.039, 0.058)	0.191 (0.160, 0.228)
Subscale Average	0.830	0.616	0.213	0.089	0.036	0.045	0.170
Design 6: <i>i</i>(s) = 8, <i>o</i> = 3							
Open-Mindedness	0.918 (0.906, 0.928)	0.810 (0.792, 0.825)	0.108 (0.097, 0.122)	0.040 (0.036, 0.043)	0.029 (0.020, 0.040)	0.013 (0.012, 0.015)	0.082 (0.072, 0.094)
Aesthetic Sensitivity	0.885 (0.868, 0.898)	0.515 (0.470, 0.559)	0.370 (0.328, 0.412)	0.079 (0.068, 0.089)	0.010 (0.002, 0.023)	0.026 (0.022, 0.030)	0.115 (0.102, 0.132)
Creative Imagination	0.840 (0.806, 0.868)	0.590 (0.540, 0.637)	0.250 (0.196, 0.306)	0.088 (0.072, 0.104)	0.040 (0.019, 0.069)	0.032 (0.026, 0.039)	0.160 (0.132, 0.194)
Intellectual Curiosity	0.833 (0.803, 0.860)	0.797 (0.704, 0.842)	0.036 (0.000, 0.136)	0.108 (0.090, 0.125)	0.025 (0.008, 0.050)	0.034 (0.027, 0.040)	0.167 (0.140, 0.197)
Subscale Average	0.852	0.634	0.219	0.092	0.025	0.031	0.148

Table 7. Cont.

Design/Scale	G (US)	Gen	Grp	Index (CI) SFE	TE	RRE	TREI
Design 7: $i(s) = 12, o = 1$							
Open-Mindedness	0.867 (0.839, 0.892)	0.765 (0.736, 0.791)	0.102 (0.092, 0.115)	0.025 (0.023, 0.027)	0.083 (0.057, 0.112)	0.025 (0.022, 0.028)	0.133 (0.108, 0.161)
Aesthetic Sensitivity	0.867 (0.833, 0.889)	0.505 (0.454, 0.552)	0.363 (0.323, 0.402)	0.052 (0.044, 0.059)	0.030 (0.007, 0.067)	0.051 (0.043, 0.059)	0.133 (0.111, 0.167)
Creative Imagination	0.776 (0.711, 0.828)	0.545 (0.488, 0.597)	0.231 (0.178, 0.286)	0.054 (0.044, 0.064)	0.110 (0.054, 0.181)	0.060 (0.047, 0.072)	0.224 (0.172, 0.289)
Intellectual Curiosity	0.795 (0.736, 0.841)	0.760 (0.661, 0.821)	0.034 (0.000, 0.129)	0.069 (0.057, 0.079)	0.072 (0.025, 0.138)	0.064 (0.050, 0.077)	0.205 (0.159, 0.264)
Subscale Average	0.813	0.603	0.209	0.058	0.071	0.058	0.187
Design 8: $i(s) = 12, o = 2$							
Open-Mindedness	0.917 (0.900, 0.931)	0.809 (0.788, 0.826)	0.108 (0.097, 0.122)	0.026 (0.024, 0.029)	0.044 (0.030, 0.060)	0.013 (0.012, 0.015)	0.083 (0.069, 0.100)
Aesthetic Sensitivity	0.904 (0.883, 0.918)	0.526 (0.478, 0.572)	0.378 (0.336, 0.420)	0.054 (0.046, 0.061)	0.015 (0.004, 0.036)	0.027 (0.023, 0.031)	0.096 (0.082, 0.117)
Creative Imagination	0.848 (0.805, 0.882)	0.595 (0.543, 0.645)	0.253 (0.198, 0.309)	0.059 (0.048, 0.070)	0.060 (0.028, 0.102)	0.033 (0.026, 0.039)	0.152 (0.118, 0.195)
Intellectual Curiosity	0.853 (0.815, 0.883)	0.816 (0.718, 0.866)	0.037 (0.000, 0.138)	0.074 (0.061, 0.085)	0.039 (0.013, 0.076)	0.034 (0.027, 0.041)	0.147 (0.117, 0.185)
Subscale Average	0.868	0.646	0.222	0.062	0.038	0.031	0.132
Design 9: $i(s) = 12, o = 3$							
Open-Mindedness	0.934 (0.923, 0.945)	0.825 (0.806, 0.839)	0.110 (0.099, 0.124)	0.027 (0.024, 0.029)	0.030 (0.020, 0.041)	0.009 (0.008, 0.010)	0.066 (0.055, 0.077)
Aesthetic Sensitivity	0.917 (0.901, 0.928)	0.533 (0.486, 0.579)	0.384 (0.340, 0.427)	0.055 (0.047, 0.062)	0.010 (0.002, 0.024)	0.018 (0.015, 0.021)	0.083 (0.072, 0.099)
Creative Imagination	0.875 (0.842, 0.901)	0.614 (0.562, 0.664)	0.261 (0.205, 0.318)	0.061 (0.050, 0.073)	0.041 (0.019, 0.071)	0.022 (0.018, 0.027)	0.125 (0.099, 0.158)
Intellectual Curiosity	0.874 (0.845, 0.898)	0.836 (0.738, 0.883)	0.038 (0.000, 0.141)	0.076 (0.062, 0.088)	0.026 (0.009, 0.053)	0.024 (0.019, 0.028)	0.126 (0.102, 0.155)
Subscale Average	0.889	0.661	0.227	0.064	0.026	0.021	0.111

Note. $i(s)$ = items per subscale, o = occasion(s), CI = 95% confidence interval limits, G = generalizability coefficient, US = proportion of universe score variance, Gen = proportion of general factor variance, Grp = proportion of group factor variance, SFE = proportion of specific-factor error variance, TE = proportion of transient error variance, RRE = proportion of random-response error variance, and TREI = total proportion of relative error variance.

Graphs such as those shown in Figure 2 would enable developers and users of the present scales to estimate how many items and/or occasions would be needed to reach a targeted level for any of the indices represented. For example, if a G coefficient of at least 0.800 is desired, then the present 4-item aesthetic sensitivity subscale would need to have the results pooled over at least three occasions ($G = 0.801$), or alternatively have three items added (i.e., include seven items in total) if the scale is administered on only one occasion ($G = 0.808$; see Figure 2).

Global D coefficients in Table 8 and Figure 2 that take relative and absolute components of measurement error into account, although somewhat lower in magnitude than G coefficients, show the same basic pattern of differences across designs in that they are higher for composite than for subscale scores, improve with increases in either number of items or occasions, and improve the most with increases in both. In general, differences in item and occasion mean scores account for relatively small proportions (0.011 to 0.044) of the overall variance in open-mindedness composite and subscale scores accounted for by universe scores and all components of error represented in the denominator of the global D coefficient formula (see Table 3).

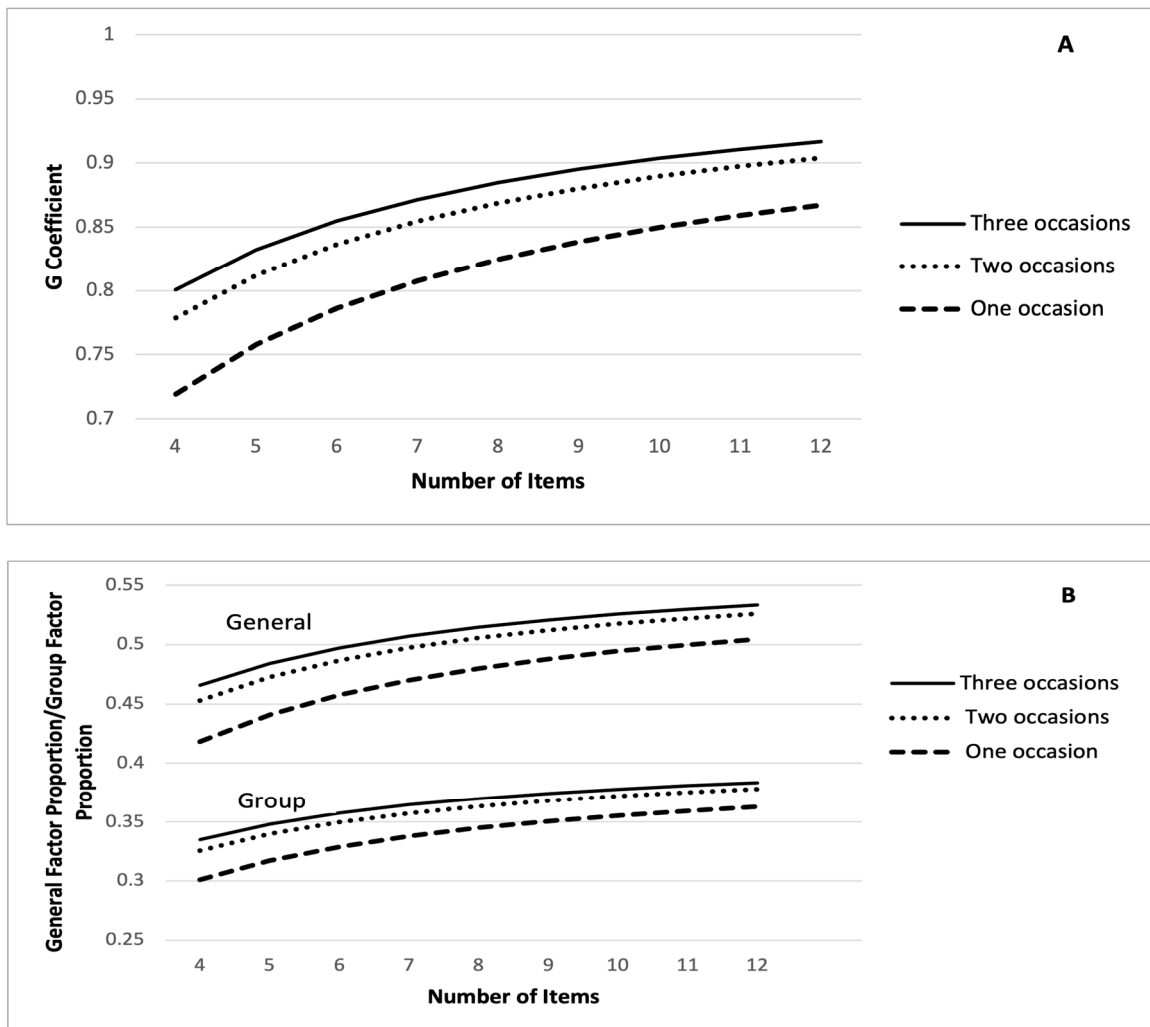


Figure 2. Cont.

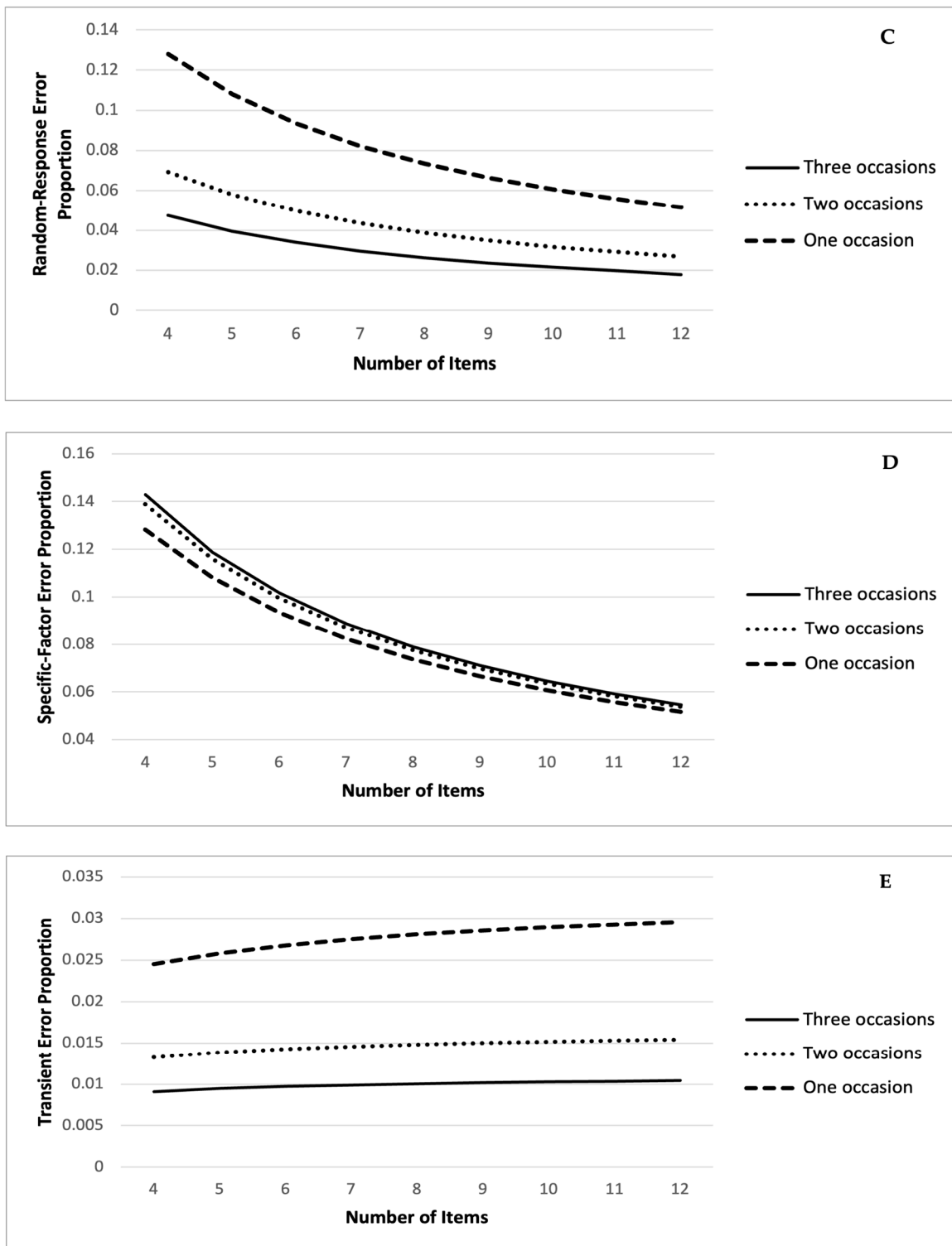


Figure 2. Cont.

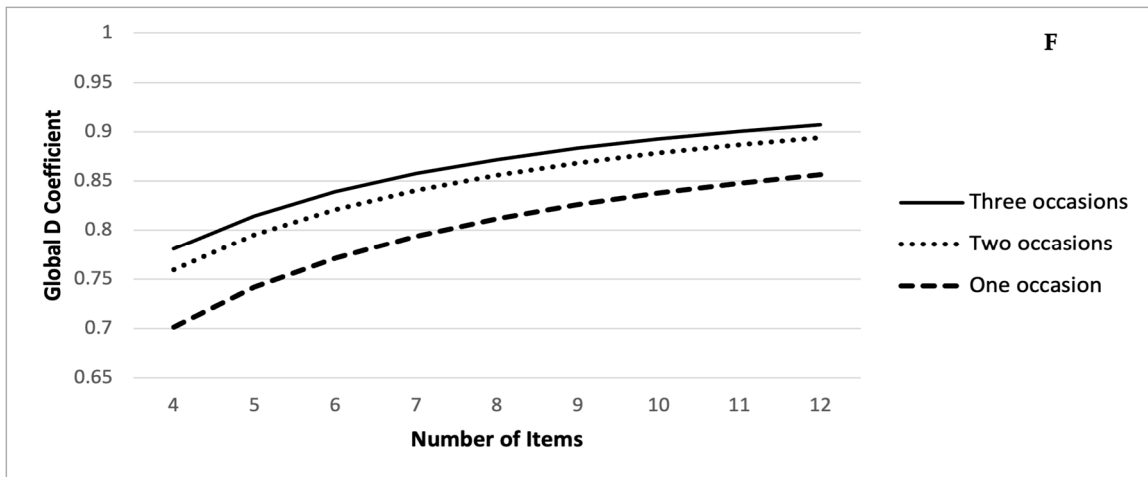


Figure 2. Panels representing aesthetic sensitivity prophecy graphs for G coefficients, general and group factor effects, measurement error effects, and global D coefficients. Panels (A–F): aesthetic sensitivity scale prophecy graphs for G coefficients (A), general and group factors effects (B), random-response error (C), specific-factor error (D), transient error (E) and global D coefficients (F). Within the graph for specific-factor error (D), relative proportions of such error increase as occasions increase due to increases in relative proportions of universe score variance and reductions in relative proportions of other sources of measurement error. Within the graph for transient error (E), relative proportions of such error increase as items increase for the same reasons.

In Figure 3, we provide cut-score specific D coefficients for open-mindedness composite and subscale scores for the baseline design. In the equations in Table 2, cut scores are expressed as average item scores but can be converted to total scores by multiplying them by the number of items in the scale (i.e., 12 for the composite and 4 for subscales), as shown in corresponding horizontal axes in Figure 3. The figure reveals that cut-score dependability increases as scores move further and further away from the means of the scales. Although not depicted here, the same relationships would hold within the other designs and, like G and global D coefficients, cut-score specific dependability coefficients would improve with increases in numbers of items, occasions, or both.

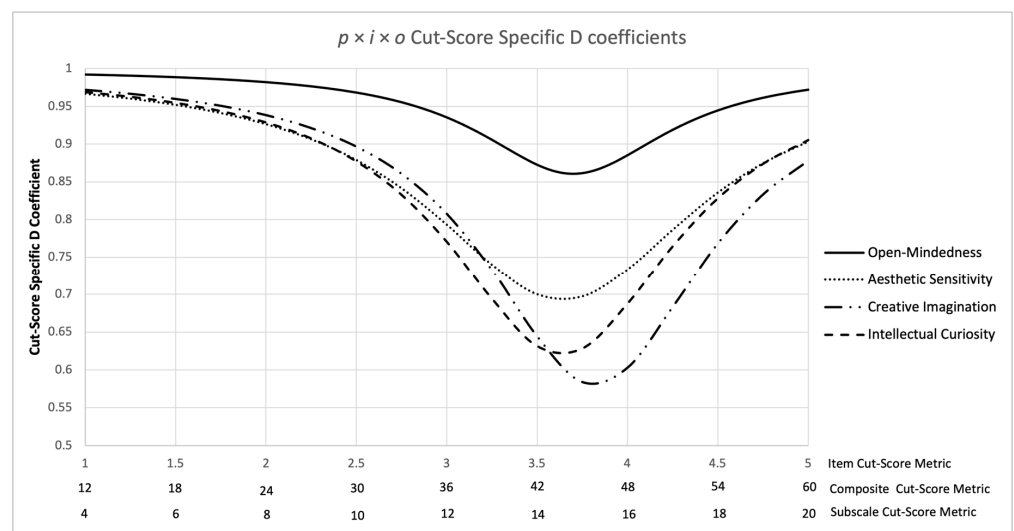


Figure 3. Baseline design cut-score specific D coefficients for open-mindedness composite and subscale scores.

Table 8. Partitioning of global D coefficient denominator variance for BFI-2 open-mindedness composite and subscale scores for *persons* × *items* × *occasions* full designs.

Design/Scale	Index (CI)							
	Global D (US)	Gen	Grp	TRE	I	O	IO	Overall MDs
Design 1: $i(s) = 4, o = 1$								
Open-Mindedness	0.778 (0.747, 0.798)	0.687 (0.656, 0.707)	0.091 (0.082, 0.103)	0.209 (0.186, 0.232)	0.011 (0.009, 0.013)	0.002 (0.000, 0.029)	0.000 (0.000, 0.001)	0.013 (0.010, 0.040)
Aesthetic Sensitivity	0.701 (0.662, 0.720)	0.408 (0.365, 0.443)	0.293 (0.256, 0.326)	0.274 (0.251, 0.299)	0.021 (0.016, 0.028)	0.003 (0.000, 0.046)	0.000 (0.000, 0.001)	0.025 (0.018, 0.067)
Creative Imagination	0.625 (0.566, 0.663)	0.439 (0.391, 0.475)	0.186 (0.142, 0.230)	0.364 (0.318, 0.410)	0.006 (0.003, 0.010)	0.005 (0.000, 0.063)	0.000 (0.000, 0.002)	0.010 (0.005, 0.069)
Intellectual Curiosity	0.600 (0.547, 0.636)	0.574 (0.500, 0.615)	0.026 (0.000, 0.098)	0.356 (0.312, 0.394)	0.039 (0.029, 0.050)	0.005 (0.000, 0.065)	0.000 (0.000, 0.002)	0.044 (0.033, 0.103)
Subscale Average	0.642	0.474	0.169	0.331	0.022	0.004	0.000	0.026
Design 2: $i(s) = 4, o = 2$								
Open-Mindedness	0.839 (0.819, 0.852)	0.740 (0.718, 0.755)	0.099 (0.089, 0.111)	0.149 (0.134, 0.164)	0.012 (0.010, 0.014)	0.001 (0.000, 0.016)	0.000 (0.000, 0.000)	0.013 (0.011, 0.028)
Aesthetic Sensitivity	0.759 (0.731, 0.776)	0.442 (0.400, 0.478)	0.318 (0.279, 0.353)	0.216 (0.198, 0.236)	0.023 (0.017, 0.030)	0.002 (0.000, 0.025)	0.000 (0.000, 0.001)	0.025 (0.019, 0.049)
Creative Imagination	0.710 (0.663, 0.742)	0.498 (0.453, 0.536)	0.211 (0.163, 0.260)	0.281 (0.245, 0.319)	0.006 (0.003, 0.012)	0.003 (0.000, 0.037)	0.000 (0.000, 0.001)	0.009 (0.005, 0.044)
Intellectual Curiosity	0.668 (0.628, 0.700)	0.640 (0.565, 0.679)	0.029 (0.000, 0.109)	0.285 (0.250, 0.317)	0.044 (0.033, 0.055)	0.003 (0.000, 0.037)	0.000 (0.000, 0.001)	0.046 (0.036, 0.082)
Subscale Average	0.712	0.526	0.186	0.261	0.024	0.002	0.000	0.027
Design 3: $i(s) = 4, o = 3$								
Open-Mindedness	0.861 (0.846, 0.871)	0.760 (0.740, 0.773)	0.101 (0.091, 0.114)	0.126 (0.115, 0.138)	0.012 (0.010, 0.014)	0.001 (0.000, 0.011)	0.000 (0.000, 0.000)	0.013 (0.011, 0.023)
Aesthetic Sensitivity	0.781 (0.757, 0.797)	0.454 (0.413, 0.491)	0.327 (0.288, 0.364)	0.194 (0.177, 0.213)	0.024 (0.017, 0.031)	0.001 (0.000, 0.018)	0.000 (0.000, 0.000)	0.025 (0.019, 0.043)
Creative Imagination	0.743 (0.703, 0.773)	0.522 (0.476, 0.560)	0.221 (0.172, 0.272)	0.248 (0.215, 0.283)	0.007 (0.003, 0.012)	0.002 (0.000, 0.026)	0.000 (0.000, 0.001)	0.008 (0.005, 0.034)
Intellectual Curiosity	0.695 (0.659, 0.727)	0.665 (0.590, 0.705)	0.030 (0.000, 0.113)	0.258 (0.225, 0.287)	0.045 (0.034, 0.057)	0.002 (0.000, 0.026)	0.000 (0.000, 0.001)	0.047 (0.037, 0.073)
Subscale Average	0.740	0.547	0.193	0.234	0.025	0.002	0.000	0.027
Design 4: $i(s) = 8, o = 1$								
Open-Mindedness	0.839 (0.804, 0.861)	0.741 (0.706, 0.763)	0.099 (0.089, 0.111)	0.153 (0.128, 0.179)	0.006 (0.005, 0.007)	0.002 (0.000, 0.031)	0.000 (0.000, 0.000)	0.008 (0.006, 0.037)
Aesthetic Sensitivity	0.812 (0.762, 0.832)	0.472 (0.420, 0.513)	0.339 (0.298, 0.375)	0.172 (0.149, 0.203)	0.012 (0.009, 0.016)	0.004 (0.000, 0.053)	0.000 (0.000, 0.001)	0.016 (0.011, 0.065)
Creative Imagination	0.728 (0.654, 0.773)	0.511 (0.450, 0.555)	0.217 (0.165, 0.267)	0.264 (0.211, 0.320)	0.003 (0.002, 0.006)	0.005 (0.000, 0.073)	0.000 (0.000, 0.001)	0.009 (0.003, 0.076)
Intellectual Curiosity	0.723 (0.653, 0.763)	0.692 (0.594, 0.743)	0.031 (0.000, 0.116)	0.247 (0.201, 0.297)	0.024 (0.018, 0.030)	0.006 (0.000, 0.078)	0.000 (0.000, 0.001)	0.029 (0.021, 0.100)
Subscale Average	0.754	0.558	0.196	0.228	0.013	0.005	0.000	0.018
Design 5: $i(s) = 8, o = 2$								
Open-Mindedness	0.892 (0.872, 0.905)	0.787 (0.763, 0.803)	0.105 (0.095, 0.118)	0.100 (0.086, 0.116)	0.006 (0.005, 0.007)	0.001 (0.000, 0.017)	0.000 (0.000, 0.000)	0.007 (0.006, 0.023)
Aesthetic Sensitivity	0.856 (0.826, 0.870)	0.498 (0.450, 0.539)	0.358 (0.315, 0.397)	0.129 (0.114, 0.149)	0.013 (0.009, 0.017)	0.002 (0.000, 0.029)	0.000 (0.000, 0.000)	0.015 (0.011, 0.042)
Creative Imagination	0.805 (0.754, 0.837)	0.565 (0.511, 0.609)	0.240 (0.186, 0.293)	0.188 (0.153, 0.229)	0.004 (0.002, 0.007)	0.003 (0.000, 0.042)	0.000 (0.000, 0.001)	0.007 (0.003, 0.046)
Intellectual Curiosity	0.786 (0.738, 0.815)	0.752 (0.658, 0.796)	0.034 (0.000, 0.128)	0.185 (0.154, 0.220)	0.026 (0.019, 0.033)	0.003 (0.000, 0.044)	0.000 (0.000, 0.001)	0.029 (0.022, 0.069)
Subscale Average	0.816	0.605	0.211	0.168	0.014	0.003	0.000	0.017
Design 6: $i(s) = 8, o = 3$								
Open-Mindedness	0.912 (0.896, 0.921)	0.804 (0.784, 0.818)	0.107 (0.097, 0.121)	0.081 (0.071, 0.093)	0.006 (0.005, 0.008)	0.001 (0.000, 0.011)	0.000 (0.000, 0.000)	0.007 (0.006, 0.018)
Aesthetic Sensitivity	0.872 (0.849, 0.884)	0.507 (0.461, 0.549)	0.365 (0.322, 0.405)	0.114 (0.101, 0.129)	0.013 (0.010, 0.017)	0.001 (0.000, 0.019)	0.000 (0.000, 0.000)	0.015 (0.011, 0.033)
Creative Imagination	0.835 (0.794, 0.861)	0.586 (0.534, 0.631)	0.249 (0.194, 0.303)	0.159 (0.131, 0.192)	0.004 (0.002, 0.007)	0.002 (0.000, 0.029)	0.000 (0.000, 0.000)	0.006 (0.003, 0.033)
Intellectual Curiosity	0.809 (0.772, 0.835)	0.774 (0.681, 0.817)	0.035 (0.000, 0.131)	0.162 (0.136, 0.191)	0.026 (0.020, 0.034)	0.002 (0.000, 0.030)	0.000 (0.000, 0.000)	0.029 (0.022, 0.057)
Subscale Average	0.839	0.623	0.216	0.145	0.014	0.002	0.000	0.016

Table 8. Cont.

Design/Scale	Index (CI)							
	Global D (US)	Gen	Grp	TRE	I	O	IO	Overall MDs
Design 7: $i(s) = 12, o = 1$								
Open-Mindedness	0.862 (0.825, 0.885)	0.761 (0.724, 0.784)	0.101 (0.091, 0.114)	0.132 (0.106, 0.159)	0.004 (0.003, 0.005)	0.002 (0.000, 0.032)	0.000 (0.000, 0.000)	0.006 (0.004, 0.036)
Aesthetic Sensitivity	0.856 (0.802, 0.877)	0.498 (0.442, 0.542)	0.358 (0.314, 0.395)	0.131 (0.108, 0.164)	0.009 (0.006, 0.011)	0.004 (0.000, 0.056)	0.000 (0.000, 0.001)	0.013 (0.008, 0.064)
Creative Imagination	0.770 (0.688, 0.818)	0.540 (0.474, 0.589)	0.229 (0.174, 0.282)	0.222 (0.167, 0.285)	0.002 (0.001, 0.004)	0.006 (0.000, 0.077)	0.000 (0.000, 0.001)	0.008 (0.002, 0.079)
Intellectual Curiosity	0.776 (0.697, 0.819)	0.743 (0.632, 0.798)	0.034 (0.000, 0.125)	0.201 (0.153, 0.256)	0.017 (0.012, 0.021)	0.006 (0.000, 0.083)	0.000 (0.000, 0.001)	0.023 (0.015, 0.099)
Subscale Average	0.801	0.594	0.207	0.185	0.009	0.005	0.000	0.015
Design 8: $i(s) = 12, o = 2$								
Open-Mindedness	0.912 (0.890, 0.925)	0.805 (0.780, 0.821)	0.107 (0.097, 0.121)	0.083 (0.068, 0.099)	0.004 (0.003, 0.005)	0.001 (0.000, 0.017)	0.000 (0.000, 0.000)	0.005 (0.004, 0.021)
Aesthetic Sensitivity	0.894 (0.862, 0.907)	0.520 (0.470, 0.564)	0.374 (0.330, 0.414)	0.095 (0.081, 0.115)	0.009 (0.007, 0.012)	0.002 (0.000, 0.030)	0.000 (0.000, 0.000)	0.011 (0.008, 0.039)
Creative Imagination	0.843 (0.789, 0.875)	0.592 (0.534, 0.639)	0.251 (0.195, 0.306)	0.151 (0.117, 0.193)	0.002 (0.001, 0.005)	0.003 (0.000, 0.043)	0.000 (0.000, 0.000)	0.006 (0.002, 0.046)
Intellectual Curiosity	0.835 (0.784, 0.863)	0.798 (0.697, 0.845)	0.036 (0.000, 0.134)	0.144 (0.114, 0.180)	0.018 (0.014, 0.023)	0.003 (0.000, 0.046)	0.000 (0.000, 0.000)	0.022 (0.015, 0.064)
Subscale Average	0.857	0.637	0.220	0.130	0.010	0.003	0.000	0.013
Design 9: $i(s) = 12, o = 3$								
Open-Mindedness	0.930 (0.915, 0.939)	0.820 (0.800, 0.834)	0.109 (0.099, 0.123)	0.065 (0.055, 0.077)	0.004 (0.004, 0.005)	0.001 (0.000, 0.012)	0.000 (0.000, 0.000)	0.005 (0.004, 0.016)
Aesthetic Sensitivity	0.907 (0.884, 0.918)	0.528 (0.480, 0.571)	0.379 (0.335, 0.421)	0.082 (0.071, 0.098)	0.009 (0.007, 0.012)	0.001 (0.000, 0.020)	0.000 (0.000, 0.000)	0.011 (0.008, 0.030)
Creative Imagination	0.871 (0.830, 0.895)	0.611 (0.556, 0.658)	0.259 (0.203, 0.315)	0.124 (0.098, 0.157)	0.003 (0.001, 0.005)	0.002 (0.000, 0.030)	0.000 (0.000, 0.000)	0.005 (0.002, 0.033)
Intellectual Curiosity	0.856 (0.818, 0.879)	0.819 (0.720, 0.863)	0.037 (0.000, 0.137)	0.123 (0.099, 0.151)	0.019 (0.014, 0.024)	0.002 (0.000, 0.032)	0.000 (0.000, 0.000)	0.021 (0.016, 0.051)
Subscale Average	0.878	0.653	0.225	0.110	0.010	0.002	0.000	0.012

Note. $i(s)$ = items per subscale, o = occasion(s), CI = 95% confidence interval limits, Global D = global dependability coefficient, US = proportion of universe score variance, TRE = total proportion of relative measurement error, I = proportion of item mean effects, O = proportion of occasion mean effects, IO = proportion of item by occasion mean interaction effects, Overall MD = proportion of overall item and occasion mean difference effects.

Scale viability and added value. In Table 9, we provide ECV, EUV, ECV/EUV, and VAR indices for the same designs within Tables 7 and 8. ECV exceed EUV indices and ECV/EUV ratios exceed 1.000 for all scales, and these relationships remain consistent with changes in numbers of items and/or occasions. The results in Table 9 reveal that the general construct open-mindedness accounts for the majority of universe score variance for all subscales, with its effects being from 1.391 to 22.170 times larger than the independent unique effects of its subdomain constructs: aesthetic sensitivity, creative imagination, and intellectual curiosity. Consistent with results previously presented, aesthetic sensitivity overlaps the least with the general factor, and intellectual curiosity overlaps the most.

VARs for designs with the number of items equaling 4, 8, and 12 and the number of occasions equaling 1, 2, and 3 are shown in Table 9 for all scales. Prophecy graphs showing VARs for all subscales such as those shown in Figure 2 for all in-between values for the number of items are provided in our online Supplemental Material. In line with the ECV/EUV ratios just reported, VAR values in Table 9 for the baseline design (Design 1) support added value (i.e., confidence interval lower limits exceed 1.000) for just the aesthetic sensitivity subscale. In most cases, VARs increase with added items and/or occasions but typically to a diminishing degree with progressively similar incremental changes. For the aesthetic sensitivity subscale, lower confidence interval limits for VARs exceed 1.000 in all designs. For the creative imagination subscale, confidence interval lower limits exceed 1.000 with 4 items across 3 occasions, 8 items across 1, 2, or 3 occasions, and 12 items across 1, 2, or 3 occasions. Finally, the intellectual curiosity subscale only yields confidence interval lower limits above 1.00 with 12 items across 2 or 3 occasions. Overall, these results underscore the benefits of altered GT-bifactor designs not only in gauging possible improvements in subscale score generalizability and dependability but also in isolating specific conditions that would support added value for any given subscale.

5.3. GT Bifactor Designs including Just Item and Just Occasion Effects

Partitioning of variance. In Table 10, we illustrate the partitioning of observed score variance represented in the denominators of G and global D coefficients when the universe of generalization is restricted to just items (i.e., *persons* × *items* designs) or just occasions (i.e., *persons* × *occasions* designs). To be consistent with the two-facet designs already discussed, we report results for number of items within subscales equaling 4, 8, and 12 within the *persons* × *items* design and number of occasions equaling 1, 2, and 3 within the *persons* × *occasions* design.

In general, patterns of relationships shown in Table 10 for changes to either items or occasions within the restricted designs mirror those for the previous designs except that G and global D coefficients are higher, and overall proportions of measurement error are lower. This occurs because transient error is treated as universe score variance in the *persons* × *items* design, and specific-factor error is treated as universe score variance in the *persons* × *occasions* design. As before, score consistency indices are higher for composites than for subscales and increase in diminishing magnitude with the same progressive increments in numbers of items or occasions.

Scale viability and added value. Results for scale viability and added value for the restricted designs in Table 11 again show that general factor effects exceed group factor effects for all scales and that ECV/EUV ratios are lowest for aesthetic sensitivity and highest for intellectual curiosity. Added value is supported (lower confidence interval limits exceed 1.000) for aesthetic sensitivity and creative imagination in all designs shown and for intellectual curiosity in all designs except within the *persons* × *items* design with 4 items per subscale. Overall, these results demonstrate that subscale added value depends both on the construct being measured and the specific source(s) of measurement error being modeled.

Table 9. Scale viability and added-value indices for BFI-2 open-mindedness composite and subscale scores for *persons* × *items* × *occasions* full designs.

Design/Scale	Index (CI)					
	ECV	EUV	ECV/EUV	PRMSE(s)	PRMSE(c)	VAR
Design 1: $i(s) = 4, o = 1$						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.531, 8.412)			
Aesthetic Sensitivity	0.582 (0.534, 0.629)	0.418 (0.371, 0.466)	1.391 (1.145, 1.699)	0.719 (0.692, 0.740)	0.634 (0.607, 0.658)	1.135 (1.094, 1.178)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.784, 3.206)	0.632 (0.583, 0.673)	0.624 (0.594, 0.652)	1.012 (0.928, 1.096)
Intellectual Curiosity	0.957 (0.839, 1.000)	0.043 (0.000, 0.161)	22.170 (5.229, 2026.368)	0.628 (0.585, 0.667)	0.687 (0.648, 0.712)	0.914 (0.855, 1.000)
Subscale Average	0.747	0.253	8.639	0.659	0.648	1.020
Design 2: $i(s) = 4, o = 2$						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.534, 8.417)			
Aesthetic Sensitivity	0.582 (0.534, 0.629)	0.418 (0.371, 0.466)	1.391 (1.144, 1.696)	0.779 (0.758, 0.796)	0.683 (0.660, 0.703)	1.140 (1.106, 1.177)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.785, 3.200)	0.716 (0.677, 0.751)	0.672 (0.645, 0.697)	1.065 (0.995, 1.137)
Intellectual Curiosity	0.957 (0.840, 1.000)	0.043 (0.000, 0.160)	22.170 (5.231, 2028.344)	0.701 (0.667, 0.735)	0.740 (0.702, 0.761)	0.948 (0.899, 1.028)
Subscale Average	0.747	0.253	8.639	0.732	0.698	1.051
Design 3: $i(s) = 4, o = 3$						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.532, 8.421)			
Aesthetic Sensitivity	0.582 (0.534, 0.630)	0.418 (0.370, 0.466)	1.391 (1.145, 1.702)	0.801 (0.781, 0.818)	0.701 (0.679, 0.720)	1.142 (1.110, 1.179)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.783, 3.209)	0.750 (0.714, 0.782)	0.690 (0.664, 0.714)	1.086 (1.021, 1.155)
Intellectual Curiosity	0.957 (0.840, 1.000)	0.043 (0.000, 0.160)	22.170 (5.248, 2038.343)	0.729 (0.698, 0.762)	0.759 (0.722, 0.779)	0.961 (0.915, 1.039)
Subscale Average	0.747	0.253	8.639	0.760	0.717	1.063
Design 4: $i(s) = 8, o = 1$						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.531, 8.420)			
Aesthetic Sensitivity	0.582 (0.534, 0.629)	0.418 (0.371, 0.466)	1.391 (1.145, 1.697)	0.825 (0.793, 0.846)	0.680 (0.650, 0.706)	1.213 (1.172, 1.256)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.786, 3.205)	0.734 (0.675, 0.783)	0.670 (0.637, 0.700)	1.096 (1.002, 1.186)
Intellectual Curiosity	0.957 (0.840, 0.999)	0.043 (0.001, 0.160)	22.170 (5.250, 1964.877)	0.745 (0.692, 0.789)	0.737 (0.694, 0.765)	1.011 (0.944, 1.097)
Subscale Average	0.747	0.253	8.639	0.768	0.696	1.107
Design 5: $i(s) = 8, o = 2$						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.531, 8.415)			
Aesthetic Sensitivity	0.582 (0.533, 0.629)	0.418 (0.371, 0.467)	1.391 (1.143, 1.697)	0.869 (0.849, 0.884)	0.723 (0.698, 0.744)	1.203 (1.171, 1.238)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.784, 3.201)	0.811 (0.769, 0.845)	0.712 (0.683, 0.738)	1.139 (1.069, 1.209)
Intellectual Curiosity	0.957 (0.839, 1.000)	0.043 (0.000, 0.161)	22.170 (5.218, 2044.612)	0.809 (0.772, 0.840)	0.783 (0.743, 0.805)	1.034 (0.985, 1.108)
Subscale Average	0.747	0.253	8.639	0.830	0.739	1.125
Design 6: $i(s) = 8, o = 3$						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.533, 8.420)			
Aesthetic Sensitivity	0.582 (0.533, 0.630)	0.418 (0.370, 0.467)	1.391 (1.143, 1.700)	0.885 (0.868, 0.898)	0.738 (0.715, 0.758)	1.199 (1.169, 1.233)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.785, 3.207)	0.840 (0.806, 0.868)	0.727 (0.699, 0.751)	1.156 (1.094, 1.219)
Intellectual Curiosity	0.957 (0.839, 1.000)	0.043 (0.000, 0.161)	22.170 (5.214, 2033.926)	0.833 (0.803, 0.860)	0.799 (0.759, 0.820)	1.042 (1.000, 1.114)
Subscale Average	0.747	0.253	8.639	0.852	0.755	1.132

Table 9. Cont.

Design/Scale	Index (CI)					
	ECV	EUV	ECV/EUV	PRMSE(s)	PRMSE(c)	VAR
Design 7: $i(s) = 12, o = 1$						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.527, 8.420)			
Aesthetic Sensitivity	0.582 (0.534, 0.630)	0.418 (0.370, 0.466)	1.391 (1.144, 1.699)	0.867 (0.833, 0.889)	0.697 (0.666, 0.724)	1.244 (1.203, 1.287)
Creative Imagination	0.702 (0.640, 0.762)	0.298 (0.238, 0.360)	2.357 (1.781, 3.206)	0.776 (0.711, 0.828)	0.687 (0.652, 0.718)	1.130 (1.032, 1.223)
Intellectual Curiosity	0.957 (0.839, 1.000)	0.043 (0.000, 0.161)	22.170 (5.220, 2101.003)	0.795 (0.736, 0.841)	0.755 (0.711, 0.784)	1.052 (0.981, 1.138)
Subscale Average	0.747	0.253	8.639	0.813	0.713	1.142
Design 8: $i(s) = 12, o = 2$						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.529, 8.419)			
Aesthetic Sensitivity	0.582 (0.534, 0.629)	0.418 (0.371, 0.466)	1.391 (1.144, 1.698)	0.904 (0.883, 0.918)	0.737 (0.712, 0.759)	1.227 (1.196, 1.261)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.783, 3.205)	0.848 (0.805, 0.882)	0.726 (0.696, 0.752)	1.168 (1.098, 1.238)
Intellectual Curiosity	0.957 (0.840, 1.000)	0.043 (0.000, 0.160)	22.170 (5.236, 2032.403)	0.853 (0.815, 0.883)	0.798 (0.757, 0.821)	1.068 (1.020, 1.141)
Subscale Average	0.747	0.253	8.639	0.868	0.754	1.155
Design 9: $i(s) = 12, o = 3$						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.535, 8.414)			
Aesthetic Sensitivity	0.582 (0.534, 0.630)	0.418 (0.370, 0.466)	1.391 (1.144, 1.699)	0.917 (0.901, 0.928)	0.751 (0.728, 0.771)	1.221 (1.192, 1.254)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.783, 3.202)	0.875 (0.842, 0.901)	0.740 (0.712, 0.765)	1.183 (1.123, 1.245)
Intellectual Curiosity	0.957 (0.840, 1.000)	0.043 (0.000, 0.160)	22.170 (5.262, 2048.940)	0.874 (0.845, 0.898)	0.814 (0.773, 0.835)	1.074 (1.034, 1.144)
Subscale Average	0.747	0.253	8.639	0.889	0.768	1.159

Note. $i(s)$ = items per subscale, o = occasion(s), CI = 95% confidence interval limits, ECV = explained common variance, EUV = explained unique variance, PRMSE(c) = proportional reduction in mean squared error for composite score, PRMSE(s) = proportional reduction in mean squared error for subscale score, and VAR = value-added ratio. Values for ECV, EUV, and ECV/EUV are the same across designs because loadings for general and group factors remain constant within the prophecy formulas and the number of items is the same across subscales.

Table 10. Partitioning of G and global D coefficient variance for BFI-2 open-mindedness composite and subscale scores within restricted designs.

Design/Scale	Index (CI)				
	G Coefficient Denominator Partitioning		Global D Coefficient Denominator Partitioning		
	US (G)	TReI (G)	US (G-D)	TReI (G-D)	MD (G-D)
Persons × Items					
Design 1: $i(s) = 4$					
Open-Mindedness	0.864 (0.859, 0.869)	0.136 (0.131, 0.141)	0.854 (0.849, 0.860)	0.135 (0.129, 0.140)	0.011 (0.009, 0.013)
Aesthetic Sensitivity	0.744 (0.730, 0.759)	0.256 (0.241, 0.270)	0.728 (0.713, 0.743)	0.251 (0.236, 0.264)	0.021 (0.016, 0.029)
Creative Imagination	0.722 (0.700, 0.745)	0.278 (0.255, 0.300)	0.717 (0.695, 0.740)	0.277 (0.254, 0.298)	0.006 (0.003, 0.011)
Intellectual Curiosity	0.685 (0.665, 0.713)	0.315 (0.287, 0.335)	0.657 (0.637, 0.686)	0.303 (0.276, 0.321)	0.040 (0.030, 0.051)
Subscale Average	0.717	0.283	0.701	0.277	0.022

Table 10. Cont.

Design/Scale	Index (CI)				
	G Coefficient Denominator Partitioning		Global D Coefficient Denominator Partitioning		
	US (G)	TREI (G)	US (G-D)	TREI (G-D)	MD (G-D)
Design 2: $i(s) = 8$					
Open-Mindedness	0.927 (0.924, 0.930)	0.073 (0.070, 0.076)	0.921 (0.918, 0.925)	0.073 (0.070, 0.076)	0.006 (0.005, 0.007)
Aesthetic Sensitivity	0.853 (0.844, 0.863)	0.147 (0.137, 0.156)	0.842 (0.832, 0.852)	0.145 (0.136, 0.154)	0.012 (0.009, 0.017)
Creative Imagination	0.838 (0.824, 0.854)	0.162 (0.146, 0.176)	0.835 (0.820, 0.851)	0.161 (0.146, 0.175)	0.003 (0.002, 0.007)
Intellectual Curiosity	0.813 (0.799, 0.832)	0.187 (0.168, 0.201)	0.793 (0.778, 0.814)	0.183 (0.164, 0.196)	0.024 (0.018, 0.031)
Subscale Average	0.835	0.165	0.824	0.163	0.013
Design 3: $i(s) = 12$					
Open-Mindedness	0.950 (0.948, 0.952)	0.050 (0.048, 0.052)	0.946 (0.944, 0.948)	0.050 (0.048, 0.052)	0.004 (0.003, 0.005)
Aesthetic Sensitivity	0.897 (0.890, 0.904)	0.103 (0.096, 0.110)	0.889 (0.882, 0.897)	0.102 (0.095, 0.109)	0.009 (0.007, 0.012)
Creative Imagination	0.886 (0.875, 0.897)	0.114 (0.103, 0.125)	0.884 (0.873, 0.895)	0.114 (0.102, 0.124)	0.002 (0.001, 0.005)
Intellectual Curiosity	0.867 (0.856, 0.882)	0.133 (0.118, 0.144)	0.852 (0.840, 0.868)	0.131 (0.116, 0.141)	0.017 (0.013, 0.022)
Subscale Average	0.883	0.117	0.875	0.116	0.009
Persons \times Occasions					
Design 1: $o = 1$					
Open-Mindedness	0.856 (0.832, 0.878)	0.144 (0.122, 0.168)	0.862 (0.838, 0.867)	0.143 (0.121, 0.167)	0.002 (0.000, 0.030)
Aesthetic Sensitivity	0.847 (0.819, 0.869)	0.153 (0.131, 0.181)	0.741 (0.707, 0.755)	0.152 (0.129, 0.179)	0.003 (0.000, 0.048)
Creative Imagination	0.764 (0.717, 0.804)	0.236 (0.196, 0.283)	0.718 (0.672, 0.740)	0.235 (0.192, 0.280)	0.005 (0.000, 0.064)
Intellectual Curiosity	0.791 (0.747, 0.828)	0.209 (0.172, 0.253)	0.681 (0.636, 0.708)	0.208 (0.168, 0.250)	0.005 (0.000, 0.069)
Subscale Average	0.801	0.199	0.713	0.198	0.004
Design 2: $o = 2$					
Open-Mindedness	0.923 (0.908, 0.935)	0.077 (0.065, 0.092)	0.929 (0.910, 0.945)	0.077 (0.065, 0.092)	0.001 (0.000, 0.016)
Aesthetic Sensitivity	0.917 (0.901, 0.930)	0.083 (0.070, 0.099)	0.790 (0.768, 0.803)	0.082 (0.069, 0.099)	0.002 (0.000, 0.026)
Creative Imagination	0.866 (0.835, 0.891)	0.134 (0.109, 0.165)	0.765 (0.733, 0.783)	0.133 (0.108, 0.164)	0.003 (0.000, 0.037)
Intellectual Curiosity	0.883 (0.855, 0.906)	0.117 (0.094, 0.145)	0.731 (0.700, 0.754)	0.117 (0.093, 0.144)	0.003 (0.000, 0.039)
Subscale Average	0.889	0.111	0.762	0.111	0.003
Design 3: $o = 3$					
Open-Mindedness	0.947 (0.937, 0.956)	0.053 (0.044, 0.063)	0.954 (0.934, 0.974)	0.053 (0.044, 0.063)	0.001 (0.000, 0.011)
Aesthetic Sensitivity	0.943 (0.931, 0.952)	0.057 (0.048, 0.069)	0.809 (0.789, 0.822)	0.057 (0.048, 0.068)	0.001 (0.000, 0.018)
Creative Imagination	0.907 (0.884, 0.925)	0.093 (0.075, 0.116)	0.784 (0.756, 0.804)	0.093 (0.075, 0.116)	0.002 (0.000, 0.026)
Intellectual Curiosity	0.919 (0.898, 0.935)	0.081 (0.065, 0.102)	0.750 (0.723, 0.775)	0.081 (0.064, 0.101)	0.002 (0.000, 0.027)
Subscale Average	0.923	0.077	0.781	0.077	0.002

Note. $i(s)$ = items per subscale, o = occasion(s), CI = 95% confidence interval limits, US = proportion of universe score variance, G = generalizability coefficient, TREI = total proportion of relative measurement error, G-D = global dependability coefficient, MD = proportion of variance due to item or occasion mean differences.

Table 11. Scale viability and added-value indices for BFI-2 open-mindedness composite and subscale scores within restricted designs.

Design/Scale	Index (CI)					
	ECV	EUV	ECV/EUV	PRMSE(s)	PRMSE(c)	VAR
Persons × Items						
Design 1: i(s) = 4						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.532, 8.421)			
Aesthetic Sensitivity	0.582 (0.533, 0.630)	0.418 (0.370, 0.467)	1.391 (1.143, 1.701)	0.744 (0.730, 0.759)	0.633 (0.606, 0.657)	1.175 (1.122, 1.240)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.786, 3.202)	0.722 (0.700, 0.745)	0.608 (0.577, 0.636)	1.186 (1.114, 1.275)
Intellectual Curiosity	0.957 (0.839, 1.000)	0.043 (0.000, 0.161)	22.170 (5.218, 2050.789)	0.685 (0.665, 0.713)	0.670 (0.629, 0.701)	1.021 (0.960, 1.124)
Subscale Average	0.747	0.253	8.639	0.717	0.637	1.127
Design 2: i(s) = 8						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.532, 8.419)			
Aesthetic Sensitivity	0.582 (0.533, 0.630)	0.418 (0.370, 0.467)	1.391 (1.143, 1.699)	0.853 (0.844, 0.863)	0.679 (0.650, 0.706)	1.255 (1.203, 1.320)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.785, 3.202)	0.838 (0.824, 0.854)	0.653 (0.619, 0.683)	1.284 (1.215, 1.368)
Intellectual Curiosity	0.957 (0.840, 1.000)	0.043 (0.000, 0.160)	22.170 (5.251, 2043.285)	0.813 (0.799, 0.832)	0.720 (0.674, 0.753)	1.130 (1.070, 1.227)
Subscale Average	0.747	0.253	8.639	0.835	0.684	1.223
Design 3: i(s) = 12						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.531, 8.422)			
Aesthetic Sensitivity	0.582 (0.534, 0.630)	0.418 (0.370, 0.466)	1.391 (1.144, 1.699)	0.897 (0.890, 0.904)	0.696 (0.666, 0.724)	1.288 (1.236, 1.352)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.787, 3.205)	0.886 (0.875, 0.897)	0.669 (0.634, 0.701)	1.324 (1.256, 1.407)
Intellectual Curiosity	0.957 (0.840, 0.999)	0.043 (0.001, 0.160)	22.170 (5.231, 1986.216)	0.867 (0.856, 0.882)	0.738 (0.691, 0.772)	1.175 (1.116, 1.270)
Subscale Average	0.747	0.253	8.639	0.883	0.701	1.262
Persons × Occasions						
Design 1: o = 1						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.532, 8.423)			
Aesthetic Sensitivity	0.582 (0.533, 0.630)	0.418 (0.370, 0.467)	1.391 (1.143, 1.699)	0.847 (0.819, 0.869)	0.634 (0.608, 0.656)	1.337 (1.293, 1.384)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.783, 3.205)	0.764 (0.717, 0.804)	0.603 (0.576, 0.628)	1.267 (1.187, 1.348)
Intellectual Curiosity	0.957 (0.839, 1.000)	0.043 (0.000, 0.161)	22.170 (5.223, 2041.584)	0.791 (0.747, 0.828)	0.650 (0.618, 0.674)	1.217 (1.154, 1.294)
Subscale Average	0.747	0.253	8.639	0.801	0.629	1.274
Design 2: o = 2						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.536, 8.422)			
Aesthetic Sensitivity	0.582 (0.533, 0.630)	0.418 (0.370, 0.467)	1.391 (1.144, 1.701)	0.917 (0.901, 0.930)	0.683 (0.662, 0.702)	1.343 (1.310, 1.380)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.786, 3.203)	0.866 (0.835, 0.891)	0.650 (0.626, 0.671)	1.333 (1.276, 1.393)
Intellectual Curiosity	0.957 (0.840, 1.000)	0.043 (0.000, 0.160)	22.170 (5.265, 2027.910)	0.883 (0.855, 0.906)	0.700 (0.671, 0.720)	1.262 (1.218, 1.322)
Subscale Average	0.747	0.253	8.639	0.889	0.678	1.313

Table 11. Cont.

Design/Scale	Index (CI)					
	ECV	EUV	ECV/EUV	PRMSE(s)	PRMSE(c)	VAR
Design 3: $o = 3$						
Open-Mindedness	0.882 (0.867, 0.894)	0.118 (0.106, 0.133)	7.509 (6.529, 8.418)			
Aesthetic Sensitivity	0.582 (0.534, 0.629)	0.418 (0.371, 0.466)	1.391 (1.145, 1.698)	0.943 (0.931, 0.952)	0.701 (0.681, 0.719)	1.346 (1.315, 1.380)
Creative Imagination	0.702 (0.641, 0.762)	0.298 (0.238, 0.359)	2.357 (1.784, 3.209)	0.907 (0.884, 0.925)	0.667 (0.644, 0.687)	1.360 (1.311, 1.412)
Intellectual Curiosity	0.957 (0.840, 1.000)	0.043 (0.000, 0.160)	22.170 (5.240, 2012.269)	0.919 (0.898, 0.935)	0.718 (0.690, 0.737)	1.279 (1.242, 1.335)
Subscale Average	0.747	0.253	8.639	0.923	0.695	1.328

Note. $i(s)$ = items per subscale, o = occasion(s), CI = 95% confidence interval limits, ECV = explained common variance, EUV = explained unique variance, PRMSE(c) = proportional reduction in mean squared error for composite score, PRMSE(s) = proportional reduction in mean squared error for subscale score, and VAR = value-added ratio. Values for ECV, EUV, and ECV/EUV are the same across designs because loadings for general and group factors remain constant within the prophecy formulas and the number of items is the same across subscales.

6. Discussion

6.1. Overview

Over the last decade or so, applications of GT and bifactor modeling have truly proliferated, but only recently have those frameworks been integrated to take advantage of what both have to offer [24,25]; also see [49]. Previous work by Vispoel and colleagues into the merger of these frameworks allowed for partitioning of universe score variance into general and group factor effects and measurement error into multiple sources (specific factor, transient, and random response). Their research further revealed that GT-bifactor designs essentially subsumed univariate GT analyses at subscale levels and multivariate GT analyses at composite levels when representing universe scores as a combination of general and group factor effects. However, score partitioning within the models they illustrated was limited to relative differences in scores, and techniques for estimating the effects of changes to measurement procedures were not covered in depth.

Our purpose in the analyses described in this article was to expand upon the work of Vispoel and colleagues to derive indices of score dependability, in addition to generalizability, and demonstrate techniques for estimating the effects of altering measurement procedures on a wide variety of key indices that included G coefficients, D coefficients, omega total and hierarchical coefficients, proportions of measurement error, and indices of scale viability and added value. These techniques were further expanded to produce confidence intervals surrounding estimates of those parameters to gauge their trustworthiness. Collectively, the results underscored the practical value of the demonstrated techniques for evaluating and improving measurement procedures and their potential for becoming standard techniques routinely applied to GT-bifactor designs.

6.2. Relative Differences in Scores and Effects of Measurement Error

Until very recently, uses of SEMs in performing GT analyses were limited almost exclusively to deriving variance components for estimating indices reflecting relative inter-person differences in scores. This is understandable because GT was developed primarily to derive such indices, and they align well with conventional indices of reliability such as alpha, split-half, equivalent form, and test-retest coefficients that also reflect such differences. The main benefits of G coefficients over conventional reliability indices are that they quantify the extent to which results can be generalized to specific assessment domains, clearly identify those domains, and account for multiple sources of measurement error. Separation of sources of measurement error within G coefficients further allows for estimation of how each source of error affects the generalizability of the results to the targeted assessment domains and identifies facets in the design (items and occasions here) that most contribute to measurement error. This information, in turn, can be used to determine the best ways to improve score consistency, which is typically to increase the number of conditions for facets that most contribute to measurement error. Such decisions can be made even more precisely by creating prophecy graphs for G coefficients such as those shown in Figure 2 to determine combinations of facet conditions that best achieve desired levels of generalizability and then choosing the combination that is most reasonable to implement in practice.

As in most studies of psychological traits using self-report measures, the present results highlighted the importance of taking the effects of specific-factor, transient, and random-response measurement error each into account (see, e.g., [24,25,30,35,37,40,57–61]). When estimating reliability using conventional single-occasion indices such as alpha and split-half coefficients, transient (state) error is confounded with universe score effects and specific-factor (method) error is confounded with random-response error. Similarly, when estimating reliability using test-retest coefficients, specific-factor error is confounded with universe score effects and transient error is confounded with random-response error. Including both items and occasions as universes of generalization in a GT design allows for the separation of specific-factor, transient, and random-response errors to create coefficients for generalizing results simultaneously across both items and occasions or just one or the

other. As shown here, but rarely in other studies, prophecy graphs can be created for each individual source of measurement error to determine the extent to which such errors can be reduced by increasing numbers of items and/or occasions. The graphs depicted here emphasized that specific-factor error is best reduced by increasing items, transient error by increasing occasions, and random-response error by either or both.

When integrating bifactor models into GT designs, universe score variance is further partitioned into general and group factor effects and prophecy graphs also can be created to show how their magnitudes change when altering numbers of items and/or occasions. Because general and group effects are part of universe score variance, proportions of such effects will typically increase when assessments are expanded to include additional items or occasions. As noted previously, proportions of general factor variance for composite scores are called *omega hierarchical total* coefficients in bifactor models, and proportions of group factor variance for subscale scores are called *omega hierarchical subscale* coefficients. As demonstrated in prophecy graphs here, GT-bifactor designs also provide a mechanism for estimating how either of these coefficients might change when altering numbers of items and/or occasions.

6.3. Absolute Differences in Item and Occasion Mean Scores

When Marcoulides [62] and Raykov and Marcoulides [63] first described how SEMs could be used to perform univariate GT analyses, estimated variance components were confined to ones reflecting relative inter-person differences in scores. More recently, Jorgensen [43] demonstrated how those same designs could be used to derive variance components for absolute differences in GT facet condition mean scores by placing additional constraints on factor loadings, means, and intercepts. Vispoel, Lee, Hong, and Chen [49] and Vispoel, Lee, and Hong [56] later extended Jorgensen's techniques to multivariate designs, and we further expanded them here to encompass GT-bifactor designs. Including variance components for differences in GT facet condition means is important when deriving indices of dependability for criterion referenced decisions because differences in means for randomly selected items and/or occasions could affect the absolute magnitude of observed scores used to make those decisions. As a result, the denominator of D coefficients takes both relative inter-person and mean differences in item and occasion scores into account when representing overall dependability of scores and levels of agreement in score location when making decisions based on individual cut scores.

In the baseline analyses reported here, differences in item and occasion mean scores accounted for relatively small proportions (0.011 to 0.044) of the collective effects of universe scores and overall error, with item mean differences accounting for most of that. In fact, for all scales, confidence intervals captured zero for o and io variance components but not for i variance components. When representing the dependability of individual cut scores shown in Figure 3, differences in mean item and occasion scores were included, but the formula for those cut scores (see Equation (14)) indicates that the impact of absolute mean differences in scores would continue to diminish as cut scores move further and further away from the mean of scale. Although illustrated only for the baseline design here, graphs of cut-score specific D coefficients such as those shown in Figure 3 also can be adjusted for any changes made to the measurement procedure. As was the case with G and global D coefficients, cut-score specific D coefficients would generally exceed those for the baseline model when numbers of items and/or occasions are increased.

6.4. Scale Viability and Added Value

Another novel and important contribution of this study was to extend previous research by estimating indices of scale viability and subscale added value for possible changes made to a measurement procedure. In the baseline model, general factor effects exceeded group factor effects for both composite and subscale scores with general factor effects being 22.170 times greater than the group effects for intellectual curiosity but only 1.391 times greater for aesthetic sensitivity. Given the sizable contribution of the group factor

effects to universe score variance for aesthetic sensitivity and the negligible contribution for intellectual curiosity, aesthetic sensitivity scores satisfied the criterion for added value with confidence interval lower limits exceeding 1.000 in the baseline and subsequent expanded facet condition designs, but intellectual curiosity did so only when tripling numbers of items and pooling results over 2 or 3 occasions.

6.5. Restricting Universes of Generalization

In addition to showing the effects of increases in numbers of items and/or occasions on key indices within the original two-facet GT-bifactor design, we derived the same indices when restricting universes to just items or just occasions. When measuring psychological traits, we would typically want to generalize results across both items and occasions to properly account for specific-factor, transient, and random-response measurement errors. However, when measuring constructs expected to change noticeably from one occasion to the next, the universe of generalization might be restricted to only items. Similarly, if we were only interested in using scores for predictions facilitated by scores collectively representing trait and method effects, then the universe of generalization might be restricted to just occasions.

The present results showed that reliability was higher, overall measurement error was lower, and scale added value improved when such restrictions were made. These indices are perfectly legitimate to interpret if they are understood to only represent the restricted universe. However, such indices would be inappropriate and potentially misleading to report when decisions entail inferences of generalization to universes beyond those being considered. Misinterpretations of this nature are common when using conventional single occasion (e.g., alpha) or test–retest reliability coefficients to represent the overall consistency of scores for psychological traits because they treat either transient or specific-factor effects as true score variance.

7. Summary and Future Extensions

Our goals in this article were to demonstrate how GT-bifactor designs can be extended to derive variance components needed to estimate dependability coefficients when using scores for criterion-referencing purposes and to determine how key indices are affected by changes made to measurement procedures. Estimated indices included G coefficients; D coefficients; proportions of observed score variance accounted for by general factor, group factor, and measurement error effects; common to unique explained variance ratios; and subscale value-added ratios. We also built Monte Carlo-based confidence intervals around those estimates to evaluate their trustworthiness. Scale viability and added-value indices best supported reporting of aesthetic sensitivity and creative imagination subscales in addition to composite scores. The analyses further demonstrated that psychometric properties of scores from all scales could be improved with increases to numbers of items or pooling results over more than one occasion.

In future research, these same analyses can be applied to cognitive, behavioral, psychomotor, and other affective domains, expanded to subjectively scored instruments, and extended to include additional measurement facets. Analytical techniques illustrated here also can be applied using procedures that account for randomly missing data [64], produce conditional standard errors of measurement for individual scores [65], vary uniquenesses and factor loadings to allow for partitioning of the observed score variance for individual items and occasions [25,30,40,66], and control for scale coarseness effects characteristic of binary and ordinal data [24,30,43,56,59,61,67]. We provide sample data and guidelines for conducting the analyses illustrated here using the *lavaan* and *semTools* packages in R within our online Supplemental Material and hope that these techniques prove useful to readers when constructing, evaluating, and revising measures that can be suitably represented within GT-bifactor model frameworks.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/psych5020036/s1>. Figure S1: Prophecy graphs for Composite Score: Open-mindedness; Figure S2: Prophecy Graphs for Subscale Score: Aesthetic Sensitivity; Figure S3: Prophecy Graphs for Subscale Score: Creative Imagination; Figure S4: Prophecy Graphs for Subscale Score: Intellectual Curiosity. Code and sample data also are provided for running all demonstrated analyses using R.

Author Contributions: Conceptualization, W.P.V., H.L., T.C. and H.H.; methodology, W.P.V. and H.L.; formal analysis, H.L., H.H. and W.P.V.; investigation, W.P.V.; resources, W.P.V.; software, H.L. and H.H.; validation, W.P.V., H.L., T.C. and H.H.; data curation, W.P.V. and H.L.; writing—original draft preparation, W.P.V., T.C. and H.L.; writing—review and editing, W.P.V., T.C. and H.L.; visualization, W.P.V., H.L., T.C. and H.H.; supervision, W.P.V.; project administration, W.P.V.; funding acquisition, W.P.V.; creation of online supplement, H.L., W.P.V. and H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This project received no external funding but did receive internal research assistant support from the Iowa Measurement Research Foundation (Grant number: 520-14-2581-00000-88395100-5045-000-92045-20-0000).

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of Iowa (ID# 200809738).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: This study was not preregistered. A dataset with 200 randomly selected cases from the one analyzed in this article is provided as Supplemental Material.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cronbach, L.J.; Rajaratnam, N.; Gleser, G.C. Theory of generalizability: A liberalization of reliability theory. *Br. J. Stat. Psychol.* **1963**, *16*, 137–163. [[CrossRef](#)]
2. Andersen, S.A.W.; Nayahangan, L.J.; Park, Y.S.; Konge, L. Use of generalizability theory for exploring reliability of and sources of variance in assessment of technical skills: A systematic review and meta-analysis. *Acad. Med.* **2021**, *96*, 1609–1619. [[CrossRef](#)] [[PubMed](#)]
3. Anderson, T.N.; Lau, J.N.; Shi, R.; Sapp, R.W.; Aalami, L.R.; Lee, E.W.; Tekian, A.; Park, Y.S. The utility of peers and trained raters in technical skill-based assessments a generalizability theory study. *J. Surg. Educ.* **2022**, *79*, 206–215. [[CrossRef](#)] [[PubMed](#)]
4. Kreiter, C.; Zaidi, N.B. Generalizability theory’s role in validity research: Innovative applications in health science education. *Health Prof. Educ.* **2020**, *6*, 282–290. [[CrossRef](#)]
5. Chen, D.; Hebert, M.; Wilson, J. Examining human and automated ratings of elementary students’ writing quality: A multivariate generalizability theory application. *Am. Educ. Res. J.* **2022**, *59*, 1122–1156. [[CrossRef](#)]
6. Lightburn, S.; Medvedev, O.N.; Henning, M.A.; Chen, Y. Investigating how students approach learning using generalizability theory. *High. Educ. Res. Dev.* **2021**, *41*, 1618–1632. [[CrossRef](#)]
7. Shin, J. Investigating and optimizing score dependability of a local ITA speaking test across language groups: A generalizability theory approach. *Lang. Test.* **2022**, *39*, 313–337. [[CrossRef](#)]
8. Kumar, S.S.; Merkin, A.G.; Numbers, K.; Sachdev, P.S.; Brodaty, H.; Kochan, N.A.; Trollor, J.N.; Mahon, S.; Medvedev, O. A novel approach to investigate depression symptoms in the aging population using generalizability theory. *Psychol. Assess.* **2022**, *34*, 684–696. [[CrossRef](#)]
9. Moore, L.J.; Freeman, P.; Hase, A.; Solomon-Moore, E.; Arnold, R. How consistent are challenge and threat evaluations? A generalizability analysis. *Front. Psychol.* **2019**, *10*, 1778. [[CrossRef](#)]
10. Truong, Q.C.; Krägeloh, C.U.; Siegert, R.J.; Landon, J.; Medvedev, O.N. Applying Generalizability theory to differentiate between trait and state in the Five Facet Mindfulness Questionnaire (FFMQ). *Mindfulness* **2020**, *11*, 953–963. [[CrossRef](#)]
11. Lafave, M.R.; Butterwick, D.J. A generalizability theory study of athletic taping using the technical skill assessment instrument. *J. Athl. Train.* **2014**, *49*, 368–372. [[CrossRef](#)] [[PubMed](#)]
12. LoPilato, A.C.; Carter, N.T.; Wang, M. Updating generalizability theory in management research: Bayesian estimation of variance components. *J. Manag.* **2015**, *41*, 692–717. [[CrossRef](#)]
13. Ford, A.L.B.; Johnson, L.D. The use of generalizability theory to inform sampling of educator language used with preschoolers with autism spectrum disorder. *J. Speech Lang. Hear. Res.* **2021**, *64*, 1748–1757. [[CrossRef](#)]
14. Holzinger, K.J.; Harman, H.H. Comparison of two factorial analyses. *Psychometrika* **1938**, *3*, 45–60. [[CrossRef](#)]
15. Holzinger, K.J.; Swineford, F. The bi-factor method. *Psychometrika* **1937**, *2*, 41–54. [[CrossRef](#)]

16. Cucina, J.; Byle, K. The bifactor model fits better than the higher order model in more than 90% of comparisons for mental abilities test batteries. *J. Intell.* **2017**, *5*, 27. [[CrossRef](#)]
17. Feraco, T.; Cona, G. Differentiation of general and specific abilities in intelligence. A bifactor study of age and gender differentiation in 8- to 19-year-olds. *Intelligence* **2022**, *94*, 101669. [[CrossRef](#)]
18. Garn, A.C.; Webster, E.K. Bifactor structure and model reliability of the Test of Gross Motor Development—3rd edition. *J. Sci. Med. Sport.* **2021**, *24*, 255–283. [[CrossRef](#)]
19. Panayiotou, M.; Santos, J.; Black, L.; Humphrey, N. Exploring the dimensionality of the Social Skills Improvement System using exploratory graph analysis and bifactor-(S-1) modeling. *Assessment* **2022**, *29*, 257–271. [[CrossRef](#)]
20. Blasco-Belled, A.; Rogoza, R.; Torrelles-Nadal, C.; Alsinet, C. Emotional intelligence structure and its relationship with life satisfaction and happiness: New findings from the bifactor model. *J. Happiness Stud.* **2020**, *21*, 2031–2049. [[CrossRef](#)]
21. Anglim, J.; Morse, G.; De Vries, R.E.; MacCann, C.; Marty, A. Comparing job applicants to non-applicants using an item-level bifactor model on the Hexaco Personality Inventory. *Eur. J. Pers.* **2017**, *31*, 669–684. [[CrossRef](#)]
22. Biderman, M.D.; McAbee, S.T.; Chen, Z.J.; Hendy, N.T. Assessing the evaluative content of personality questionnaires using bifactor models. *J. Pers. Assess.* **2018**, *100*, 375–388. [[CrossRef](#)]
23. Hörz-Sagstetter, S.; Volkert, J.; Rentrop, M.; Benecke, C.; Gremaud-Heitz, D.J.; Unterrainer, H.-F.; Schauenburg, H.; Seidler, D.; Buchheim, A.; Doering, S.; et al. A bifactor model of personality organization. *J. Pers. Assess.* **2021**, *103*, 149–160. [[CrossRef](#)] [[PubMed](#)]
24. Vispoel, W.P.; Lee, H.; Xu, G.; Hong, H. Integrating bifactor models into a generalizability theory structural equation modeling framework. *J. Exp. Educ.* **2022**. [[CrossRef](#)]
25. Vispoel, W.P.; Lee, H.; Xu, G.; Hong, H. Expanding bifactor models of psychological traits to account for multiple sources of measurement error. *Psychol. Assess.* **2022**, *32*, 1093–1111. [[CrossRef](#)]
26. Longo, Y.; Jovanović, V.; Sampaio de Carvalho, J.; Karaš, D. The general factor of well-being: Multinational evidence using bifactor ESEM on the Mental Health Continuum-Short Form. *Assessment* **2020**, *27*, 596–606. [[CrossRef](#)]
27. Burns, G.L.; Geiser, C.; Servera, M.; Becker, S.P.; Beauchaine, T.P. Application of the bifactor S-1 model to multisource ratings of ADHD/ODD symptoms: An appropriate bifactor model for symptom ratings. *J. Abnorm. Child Psych.* **2020**, *48*, 881–894. [[CrossRef](#)]
28. Gomez, R.; Vance, A.; Gomez, R.M. Validity of the ADHD bifactor model in general community samples of adolescents and adults, and a clinic-referred sample of children and adolescents. *J. Atten. Disord.* **2018**, *22*, 1307–1319. [[CrossRef](#)]
29. Willoughby, M.T.; Fabiano, G.A.; Schatz, N.K.; Vujnovic, R.K.; Morris, K.L. Bifactor models of attention deficit/hyperactivity symptomatology in adolescents: Criterion validity and implications for clinical practice. *Assessment* **2019**, *26*, 799–810. [[CrossRef](#)]
30. Vispoel, W.P.; Hong, H.; Lee, H. Benefits of doing generalizability theory analyses within structural equation modeling frameworks: Illustrations using the Rosenberg Self-Esteem Scale [Teacher’s corner]. *Struct. Equ. Model.* **2023**. [[CrossRef](#)]
31. Soto, C.J.; John, O.P. The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *J. Pers. Soc. Psychol.* **2017**, *113*, 117–143. [[CrossRef](#)]
32. Reise, S.P.; Bonifay, W.E.; Haviland, M.G. Scoring and modeling psychological measures in the presence of multidimensionality. *J. Pers. Assess.* **2013**, *95*, 129–140. [[CrossRef](#)]
33. Rodriguez, A.; Reise, S.P.; Haviland, M.G. Applying bifactor statistical indices in the evaluation of psychological measures. *J. Pers. Assess.* **2016**, *98*, 223–237. [[CrossRef](#)]
34. Rodriguez, A.; Reise, S.P.; Haviland, M.G. Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychol. Methods* **2016**, *21*, 137–150. [[CrossRef](#)] [[PubMed](#)]
35. Le, H.; Schmidt, F.L.; Putka, D.J. The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. *Organ. Res. Methods* **2009**, *12*, 165–200. [[CrossRef](#)]
36. Thorndike, R.L. Reliability. In *Educational Measurement*; Lindquist, E.F., Ed.; American Council on Education: Washington, DC, USA, 1951; pp. 560–620.
37. Schmidt, F.L.; Le, H.; Ilies, R. Beyond alpha: An empirical investigation of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychol. Methods* **2003**, *8*, 206–224. [[CrossRef](#)]
38. Geiser, C.; Lockhart, G. A comparison of four approaches to account for method effects in latent state-trait analyses. *Psychol. Methods* **2012**, *17*, 255–283. [[CrossRef](#)]
39. Steyer, R.; Ferring, D.; Schmitt, M.J. States and traits in psychological assessment. *Eur. J. Psychol. Assess.* **1992**, *8*, 79–98. [[CrossRef](#)]
40. Vispoel, W.P.; Xu, G.; Schneider, W.S. Interrelationships between latent state-trait theory and generalizability theory in a structural equation modeling framework. *Psychol. Methods* **2022**, *27*, 773–803. [[CrossRef](#)]
41. Brennan, R.L.; Kane, M.T. An index of dependability for mastery tests. *J. Educ. Meas.* **1977**, *14*, 277–289. [[CrossRef](#)]
42. Kane, M.T.; Brennan, R.L. Agreement coefficients as indices of dependability for domain-referenced tests. *Appl. Psychol. Meas.* **1980**, *4*, 105–126. [[CrossRef](#)]
43. Jorgensen, T.D. How to estimate absolute-error components in structural equation models of generalizability theory. *Psych* **2021**, *3*, 113–133. [[CrossRef](#)]
44. Little, T.D.; Siegers, D.W.; Card, A. A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Struct. Equ. Modeling* **2006**, *13*, 59–72. [[CrossRef](#)]
45. Reise, S.P. The rediscovery of bifactor measurement models. *Multivar. Behav. Res.* **2012**, *47*, 667–696. [[CrossRef](#)]

46. Haberman, S.J. When can subscores have value? *J. Educ. Behav. Stat.* **2008**, *33*, 204–229. [[CrossRef](#)]
47. Haberman, S.J.; Sinharay, S. Reporting of subscores using multidimensional item response theory. *Psychometrika* **2010**, *75*, 209–227. [[CrossRef](#)]
48. Sinharay, S. Added value of subscores and hypothesis testing. *J. Educ. Behav. Stat.* **2019**, *44*, 25–44. [[CrossRef](#)]
49. Vispoel, W.P.; Lee, H.; Hong, H.; Chen, T. Applying Multivariate Generalizability Theory to Psychological Assessments. *Psychol. Methods* **2022**. *submitted*.
50. Feinberg, R.A.; Wainer, H. A simple equation to predict a subscore's value. *Educ. Meas.* **2014**, *33*, 55–56. [[CrossRef](#)]
51. Rosseel, Y. lavaan: An R package for structural equation modeling. *J. Stat. Softw.* **2012**, *48*, 1–36. [[CrossRef](#)]
52. Rosseel, Y.; Jorgensen, T.D.; Rockwood, N. Package 'Lavaan'. R Package Version (0.6–15). 2023. Available online: <https://cran.r-project.org/web/packages/lavaan/lavaan.pdf> (accessed on 27 April 2023).
53. Jorgensen, T.D.; Pornprasertmanit, S.; Schoemann, A.M.; Rosseel, Y. semTools: Useful Tools for Structural Equation Modeling. R Package Version 0.5–6. 2022. Available online: <https://CRAN.R-project.org/package=semTools> (accessed on 9 February 2023).
54. John, O.P.; Donahue, E.M.; Kentle, R.L. *The Big Five Inventory—Versions 4a and 54*; University of California, Berkeley, Institute of Personality and Social Research: Berkeley, CA, USA, 1991.
55. Revelle, W. Psych: Procedures for Psychological, Psychometric, and Personality Research. R Package Version (2.3.3). 2023. Available online: <https://cran.r-project.org/web/packages/psych/index.html> (accessed on 27 April 2023).
56. Vispoel, W.P.; Lee, H.; Hong, H. Analyzing multivariate generalizability theory designs within structural equation modeling frameworks. *Struct. Equ. Model.* **2023**, *in press*.
57. Morris, C.A. Optimal Methods for Disattenuating Correlation Coefficients under Realistic Measurement Conditions with Single-Form, Self-Report Instruments (Publication No. 27668419). Ph.D. Thesis, University of Iowa, Iowa City, IA, USA, 2020.
58. Reeve, C.L.; Heggstad, E.D.; George, E. Estimation of transient error in cognitive ability scales. *Int. J. Select. Assess.* **2005**, *13*, 316–332. [[CrossRef](#)]
59. Vispoel, W.P.; Morris, C.A.; Kilinc, M. Applications of generalizability theory and their relations to classical test theory and structural equation modeling. *Psychol. Methods* **2018**, *23*, 1–26. [[CrossRef](#)] [[PubMed](#)]
60. Vispoel, W.P.; Morris, C.A.; Kilinc, M. Practical applications of generalizability theory for designing, evaluating, and improving psychological assessments. *J. Pers. Assess.* **2018**, *100*, 53–67. [[CrossRef](#)]
61. Vispoel, W.P.; Morris, C.A.; Kilinc, M. Using generalizability theory with continuous latent response variables. *Psychol. Methods* **2019**, *24*, 153–178. [[CrossRef](#)] [[PubMed](#)]
62. Marcoulides, G.A. Estimating variance components in generalizability theory: The covariance structure analysis approach [Teacher's corner]. *Struct. Equ. Modeling* **1996**, *3*, 290–299. [[CrossRef](#)]
63. Raykov, T.; Marcoulides, G.A. Estimation of generalizability coefficients via a structural equation modeling approach to scale reliability evaluation. *Int. J. Test.* **2006**, *6*, 81–95. [[CrossRef](#)]
64. Enders, C.K.; Bandalos, D.L. The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Struct. Equ. Model.* **2001**, *8*, 430–457. [[CrossRef](#)]
65. Huebner, A.; Skar, G.B. Conditional standard error of measurement: Classical test theory, generalizability theory and many-facet Rasch measurement with applications to writing assessment. *Pract. Assess. Res. Eval.* **2021**, *26*, 1–20.
66. Vispoel, W.P.; Xu, G.; Kilinc, M. Expanding G-theory models to incorporate congeneric relationships: Illustrations using the Big Five Inventory. *J. Pers. Assess.* **2021**, *103*, 429–442. [[CrossRef](#)]
67. Ark, T.K. Ordinal Generalizability Theory Using an Underlying Latent Variable Framework. Ph.D. Thesis, University of British Columbia, Vancouver, BC, Canada, 2015. Available online: <https://open.library.ubc.ca/soa/cIRcle/collections/ubctheses/24/items/1.0166304> (accessed on 9 February 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.