# ERRATA

*Psychometric Methods: Theory into Practice*
Larry R. Price

Errors were made in Equations 3.5a and 3.5b, Figure 3.2, equations and text on pages 276–280, and Table 9.1. Versions of the relevant pages that include the corrected material follow.

Readers are encouraged to email the publisher at *errata@guilford.com* to report any further errata.

**Equation 3.5a.** Semipartial correlation coefficient

$$r_{YX_1 \cdot X_2} = \frac{r_{YX_2} - r_{YX_1} r_{X_1 X_2}}{\sqrt{1 - r^2_{X_1 X_2}}}$$

- $r_{YX_1 \cdot X_2}$ = semipartial correlation coeffecient.
- $r_{YX_2}$ = correlation between criterion $Y$ and predictor $X_2$.
- $r_{X_1 X_2}$ = correlation between predictor $X_1$ and predictor $X_2$.
- $r^2_{YX_2}$ = square of the correlation between criterion $Y$ and predictor $X_2$.
- $r^2_{X_1 X_2}$ = square of the correlation between $X_1$ and predictor $X_2$.
- $r^2_{XY}$ = coefficient of determination or proportion of variance accounted for in $Y$ by $X$.
- $r^2_{YX_1 \cdot X_2}$ = coefficient of determination or proportion of variance accounted for in $Y$ by $X_1$ after controlling for $X_2$.

*Note*. The variable following the multiplication dot (·) is the variable being "partialed."

($Y$) accounted for by language development ($X_1$) after the effect of graphic identification ($X_2$) is partialed or controlled.

Applying the correlation coefficients from our example data, we have the result in Equation 3.5b. Note that the result below agrees with Equation 3.4b. Therefore, we have illustrated a second way to arrive at the same conclusion but the semipartial correlation provides a slightly different way to isolate or understand the unique and nonunique relationships among the predictor variables in relation to the criterion.

Figure 3.2 provides a Venn diagram depicting the results of our analysis in Equation 3.5b.

**Equation 3.5b.** Semipartial correlation coefficient with example data

$$r_{YX_1 \bullet X_2} = \frac{r_{YX_2} - r_{YX_1} r_{X_1 X_2}}{\sqrt{1 - r^2_{X_1 X_2}}} = \frac{.428 - .799(.392)}{\sqrt{1 - .153}} = \frac{.1148}{.920} = .125 \Rightarrow r^2_{YX_1 \bullet X_2} = .0156$$

and

$$(.0156)(100) = 1.56\%$$

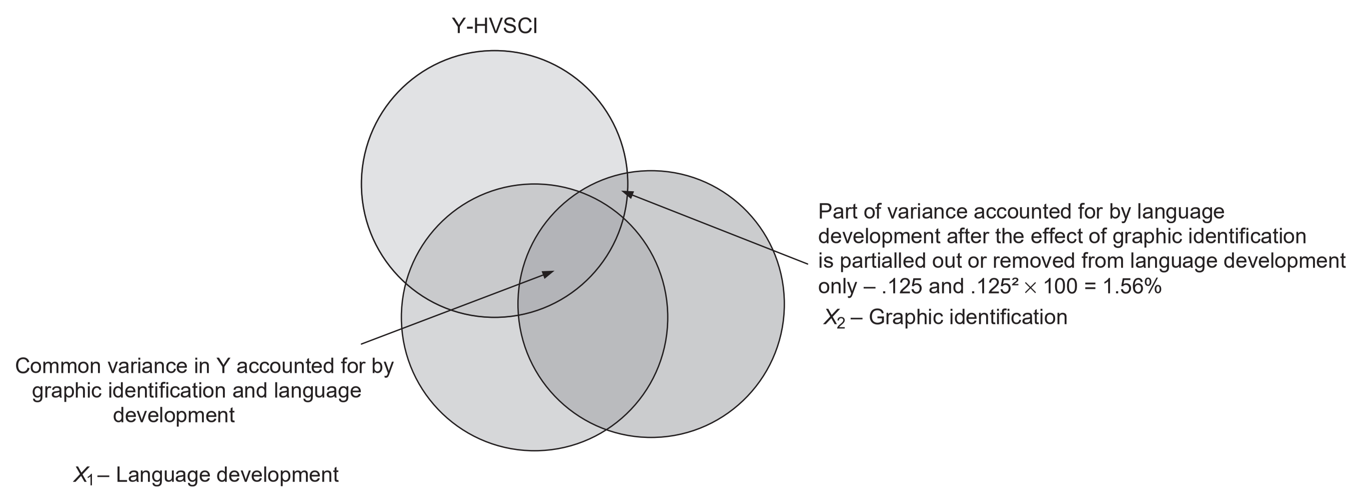


**FIGURE 3.2.** Venn diagram illustrating the semipartial correlation. The circles represent percentages (e.g., each circle represents 100% of each variable). This allows for conversion of correlation coeffecients into the proportion of variance metric, $r^2$. The $r^2$ metric can then be converted to percentages to aid interpretation.

**TABLE 8.8a. Repeated Measures ANOVA Output for the Person × Rater Design**

Tests of Within-Subjects Effects

Measure: MEASURE_1

| Source | | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| raters | Sphericity Assumed | 5.233 | 2 | 2.617 | . | . |
| raters * persons (Residual) | Sphericity Assumed | 20.100 | 38 | .529 | . | . |

*Note*. Parts of the output have been omitted for ease of interpretation.

**TABLE 8.8b. Repeated Measures ANOVA Output for the Person × Rater Design**

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|
| Intercept | 1118.017 | 1 | 1118.017 | . | . |
| persons | 95.650 | 19 | 5.034 | . | . |
| Error | .000 | 0 | . | | |

Next, the variance components are calculated using mean squares from the ANOVA results. The variance component estimate for persons is provided in Equation 8.9, and the estimate for raters is provided in Equation 8.10.

The variance component estimate for error or the residual is provided in Equation 8.11.

To illustrate how the generalizability coefficient obtained in our G-study can be used within a D-study, let's assume that the raters used in our G-study are *representative of the raters in the universe of generalization*. Under this assumption, our best estimate is the

**Equation 8.9.** Variance component for persons

$$\hat{\sigma}_p^2 = \frac{MS_{\text{persons}} - MS_{\text{residual}}}{n_{\text{raters}}} = \frac{5.03 - .53}{3} = \frac{4.5}{3} = 1.5$$

**Equation 8.10.** Variance component for raters

$$\hat{\sigma}_{\text{raters}}^2 = \frac{MS_{\text{raters}} - MS_{\text{residual}}}{n_p} = \frac{2.62 - .53}{20} = \frac{2.09}{20} = .1045$$

**Equation 8.11.** Proportion of variance for residual

$$\hat{\sigma}_e^2 = \text{MS}_{\text{residual}} = .53$$

average observed score variance for all the raters in the universe. The average score variance is captured in the sum of $\sigma_p^2 + \sigma_e^2$. Because we are willing to assume that our raters are representative of the universe of raters we can estimate the coefficient of generalizability in Equation 8.12 from our sample data. *An important point here is that raters are not usually randomly sampled from all possible raters in the universe of generalization*, leading to one difficulty with this design.

The value of .89 indicates that the raters are highly reliable in their ratings. Using this information, we can plan a D-study in a way that ensures that rater reliability will be adequate by changing the number of raters. For example, if the number of raters is reduced to two in the D-study, the variance component for the residual changes to .27. Using the new variance component for the residual in Equation 8.13 yields a generalizability coefficient of .85 (which is still acceptably high).

Next, we turn to the proportion of variance as illustrated in Equation 8.14 as a way to understand the magnitude of the effects.

In G theory studies, the proportion of variance provides a measure of effect size that is comparable across studies. The proportion of variance is reported for each facet in a study. For example, the proportion of variance for persons is provided in Equation 8.14.

Equation 8.14 shows that the person effect accounts for approximately 32% of the variability in rating scores among persons. Next, in Equation 8.15 we calculate the proportion of variance for the rater effect.

We see from Equation 8.15 that the rater effect accounts for approximately 56% of the variability in memory score performance ratings. From this information we conclude

**Equation 8.12.** Generalizability coefficient for rating data with residual averaged over raters

$$\hat{\rho}^2_{\text{raters*}} = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \frac{\hat{\sigma}_e^2}{n_i^{'}}} = \frac{1.5}{1.5 + \frac{.53}{3}} = \frac{1.5}{1.5 + .18} = \frac{1.5}{1.5 + .18} = \frac{1.5}{1.68} = .89$$

*Note*. The asterisk (*) signifies that the G coefficient can be used for a D-study with persons crossed with raters (i.e., the measurement conditions). Notation is from Crocker and Algina (1986, p. 167).

**Equation 8.13.** Revised generalizability coefficient for rating data

$$\hat{\rho}^2_{\text{raters*}} = \frac{\hat{\sigma}^2_p}{\hat{\sigma}^2_p + \dfrac{\hat{\sigma}^2_e}{n_i^{'}}} = \frac{1.5}{1.5 + \dfrac{.53}{2}} = \frac{1.5}{1.5 + .27} = \frac{1.5}{1.5 + .27} = \frac{1.5}{1.77} = .85$$

*Note*. The asterisk (*) signifies that the G coefficient can be used for a D-study with persons crossed with the average number of raters (i.e., the measurement conditions).

**Equation 8.14.** Proportion of variance for persons

$$\frac{\hat{\sigma}^2_p}{\hat{\sigma}^2_p + \hat{\sigma}^2_r + \hat{\sigma}^2_{\text{residual}}} = \frac{1.5}{1.5 + 2.62 + .53} = \frac{1.5}{4.65} = .32$$

**Equation 8.15.** Proportion of variance for raters

$$\frac{\hat{\sigma}^2_r}{\hat{\sigma}^2_p + \hat{\sigma}^2_r + \hat{\sigma}^2_{\text{residual}}} = \frac{2.62}{1.5 + 2.62 + .53} = \frac{2.62}{4.65} = .56$$

that the rater effect is moderate (i.e., raters account for or capture a medium amount of variability among the raters). Another way of interpreting this finding is that the raters are moderately similar or consistent in their ratings.

## 8.12 DESIGN 3: TWO-FACET DESIGN WITH THE SAME RATERS ON MULTIPLE OCCASIONS

In Design 3, we cover a G-study where the ratings are averaged, a strategy used to reduce the error variance in the measurement condition. We can average over raters because the *same observers are conducting the ratings on each occasion* for persons (i.e., raters are not different for persons). Averaging over raters involves dividing the appropriate error component by the number of raters and occasions. For example, in Equation 8.16 the error

**Equation 8.16.** Generalizability coefficient for two-facet design—same 3 raters and 2 occasions

$$\hat{\rho}^2_{\text{RATERS}} = \frac{\hat{\sigma}^2_p}{\hat{\sigma}^2_p + \dfrac{\hat{\sigma}^2_{\text{raters}} + \hat{\sigma}^2_{\text{error}}}{n'_{\text{raters}}}} = \frac{1.5}{1.5 + \dfrac{.104 + .53}{3}} = \frac{1.5}{1.5 + .21} = \frac{1.5}{1.71} = .88$$

*Note*. The asterisk (*) signifies that the G coefficient can be used for a D-study with persons crossed with the average number of raters (i.e., the measurement conditions). Capital notation for RATERS signifies that the error variance is divided by 3, the number of raters in a D-study. The symbol $n'_{\text{raters}}$ signifies the number of ratings to form the average. Notation is from Crocker and Algina (1986, p. 167).

variance component is divided by 3 ([.104 + .53]/3). In our example data, the change realized in the G coefficient by averaging over raters is from .89 to .88 (Equation 8.16).

There is little decrease in the G coefficient (i.e., from .89 in Design 2 to .88 in Design 3), telling us that when it is reasonable to do so, averaging over raters is an acceptable strategy.

## 8.13 DESIGN 4: TWO-FACET NESTED DESIGN WITH MULTIPLE RATERS

In Design 3, we illustrated the situation in which each person is rated by the same raters on multiple occasions. In Design 4, each person has three ratings (on three occasions), *but each person is rated by a different rater*. For example, this may occur in the event that a large pool of raters is available for use in a G-study. In this scenario, raters are *nested within persons*. Symbolically, this nesting effect is expressed as $r : p$ or $r(p)$. In this design, differences among persons are influenced by (1) rater differences plus (2) universe score differences for persons and (3) error variance. To capture this variance, the observed score variance for this design is $\sigma^2_p + \sigma^2_{\text{raters}} + \sigma^2_e$, where the variance component symbols are the same as in Design 2. Using the same mean square information in Equations 8.9, 8.10, and 8.11, we find that the G coefficient for Design 4 is provided in Equation 8.17.

We see that there is substantial reduction in the G coefficient from .89 (Design 2) or .88 (Design 3) to .70 (Design 4). Knowing this information about the reduction of the G coefficient to an unacceptable level, we can plan accordingly by using Design 2 or 3 rather than Design 4.

**Equation 8.17.** Generalizability coefficient for Design 4

$$\hat{\rho}^2_{\text{RATERS}} = \frac{\sigma^2_p}{\hat{\sigma}^2_p + \hat{\sigma}^2_{\text{raters}} + \hat{\sigma}^2_{\text{residual}}} = \frac{1.5}{1.5 + .104 + .53} = \frac{1.5}{1.5 + .63} = \frac{1.5}{2.13} = .70$$

*Note*. No asterisk (*) is included in the equation after "raters," signifying that this is a D-study and the measurement condition of ratings is nested within persons.

## 8.14 DESIGN 5: TWO-FACET DESIGN WITH MULTIPLE RATERS RATING ON TWO OCCASIONS

In Design 4, the scenario was illustrated where different raters rate each person and each person is rated on three occasions. Our strategy in Design 5 with multiple raters and occasions of measurement is to average over ratings. The G coefficient for Design 5 is provided in Equation 8.18.

Table 8.9 summarizes the formulas for the four G coefficients based on the designs covered to this point (excluding Design 5, which is a modification of Design 4).

**Equation 8.18.** Generalizability coefficient for Design 5

$$\hat{\rho}^2_{\text{RATERS}} = \frac{\hat{\sigma}^2_p}{\hat{\sigma}^2_p + \dfrac{\hat{\sigma}^2_{\text{raters}} + \hat{\sigma}^2_{\text{residual}}}{n'_{\text{raters}}}} = \frac{1.5}{1.5 + \dfrac{.104 + .53}{3}}$$

$$= \frac{1.5}{1.5 + .21} = \frac{1.5}{1.71} = .88$$

*Note*. The word RATERS in capital letters signifies that the measurement condition, ratings, are averaged over raters. The symbol $n'_{\text{raters}}$ signifies the number of ratings to form the average. Notation is from Crocker and Algina (1986, p. 167).

**TABLE 9.1. Subtest Variables in the *GfGc* Dataset**

| | Name of subtest | Number of items | Scoring |
|---|---|---|---|
| Fluid intelligence (*Gf*) | | | |
| Quantitative reasoning—sequential | Fluid intelligence test 1 | 10 | 0/1/2 |
| Quantitative reasoning—abstract | Fluid intelligence test 2 | 20 | 0/1 |
| Quantitative reasoning—induction and deduction | Fluid intelligence test 3 | 20 | 0/1 |
| Crystallized intelligence (*Gc*) | | | |
| Language development | Crystallized intelligence test 1 | 25 | 0/1/2 |
| Lexical knowledge | Crystallized intelligence test 2 | 25 | 0/1 |
| Listening ability | Crystallized intelligence test 3 | 15 | 0/1/2 |
| Communication ability | Crystallized intelligence test 4 | 15 | 0/1/2 |
| Short-term memory (*Gsm*) | | | |
| Recall memory | Short-term memory test 1 | 20 | 0/1/2 |
| Auditory learning | Short-term memory test 2 | 10 | 0/1/2/3 |
| Arithmetic | Short-term memory test 3 | 15 | 0/1/2 |

deductive reasoning) does not correlate at even a moderate level with graphic orientation and graphic identification. Additionally, inspection of the unshaded cells in Table 9.2 reveals that the subtests in the theoretical clusters also correlate moderately (with the exception of subtest 10 on inductive and deductive reasoning) with subtests that are not part of their theoretical cluster.

## 9.4 ESTIMATING FACTORS AND FACTOR LOADINGS

At the heart of FA is the relationship between a correlation matrix and a set of **factor loadings**. The intercorrelations among the variables and the factors share an intimate relationship. Although factor(s) are *unobservable variables,* it is possible to calculate the correlation between factors and variables (e.g., subtests in our *GfGc* example). The correlation between factors and the *GfGc* subtests are called factor loadings. For example, consider questions 1–4 originally given in Section 9.1.

1. What role does the pattern of intercorrelations among the variables or subtests play in identifying the number of factors?
2. What are the general steps in conducting a factor-analytic study?
3. How are factors estimated?
4. How are factor loadings interpreted?

Through these questions, we seek to know (1) how the pattern of correlations among the variables inform what the factor loadings are, (2) how the loadings are estimated; and