

EXHIBIT 7.7 Sources of Contamination Error and Suggestions for Control

Source of Contamination	Example	Suggestion for Control
Content domain	Irrelevant material on test	Define domain of test material to be covered
Standardization	Different time limits for same test	Have same time limits for everyone
Chance response tendencies	Guessing by test taker	Impossible to control in advance
Rater	Rater gives inflated ratings to people	Train rater in rating accuracy
Rating situation	Interviewees are asked different questions	Ask all interviewees the same questions

dures (test-retest and intrarater agreement) assess reliability between time periods. Reliability scores range from 0 (indicating that there is nothing but error in the actual scores) to 1.0 (indicating that the actual scores are the “true” scores). Like the squared correlation (r^2), you can think of reliability estimates as the percentage of variation in a given measure that is “true” as opposed to “error.”

Coefficient Alpha. Coefficient alpha may be calculated in instances in which there are two or more items (or raters) for a particular attribute. Its formula is

$$\alpha = \frac{n(\bar{r})}{1 + \bar{r}(n - 1)}$$

where $\bar{r}$ is the average intercorrelation among the items (raters) and n is the number of items (raters). For example, if there are five items ($n = 5$) and the average correlation among those five items is $\bar{r} = .80$, coefficient alpha is .95.

It can be seen from the formula and the example that coefficient alpha depends on just two things—the number of items and the amount of correlation between them. This suggests two basic strategies for increasing the internal consistency reliability of a measure—increase the number of items and increase the amount of agreement between the items (raters). It is generally recommended that coefficient alpha be at least .80 for a measure to have an acceptable degree of reliability. That means that 80% of the variation in scores is “true” variation and 20% is “error.”

Interrater Agreement. When raters serve as the measure, it is often convenient to talk about interrater agreement, or the amount of agreement among them. For