Postscript to "The Reliability of a Linear Composite of Nonequivalent Subtests"

William W. Rozeboom
University of Alberta

Some practicality clarifications for use of the composite-reliability formula developed in Rozeboom (1989) are described. A microcomputer program is announced to implement the calculations, and a prior source of related work is identified. *Index terms: composite reliability, item weighting, nonequivalent subtests, non-homogeneous item composites.*

Inquiries in response to Rozeboom's (1989) recent article on the reliability of composite tests whose makeup violates the classical equivalent-components premise of internal-consistency reliability estimates (Spearman-Brown, split-half, coefficient α, and others) show that there is indeed an applied need for the estimation formula developed there. However, further clarification and computational assistance appears to be needed. In particular, it is important to identify specific conditions under which Rozeboom's Equation 9 can and should be preferred in practice to more traditional reliability estimates.

Clarification 1

Rozeboom's (1989) Equation 9 is

$$r_{x^*} = 1 - \left[\sum_{i=1}^{m} w_i^2\sigma_i^2(1 - r_{x_i})\right] / \sigma_{x^*}^2 \quad , \tag{1}$$

where r_{x^*} is the (estimated) reliability of weighted composite x^*, σ_{x^*} is its standard deviation, and for each $i = 1, \ldots, m$, σ_i and r_{x_i} are, respectively, the standard deviation and (estimated) reliability of constituent item x_i.

Although Equation 1 [i.e., Equation 9 in Rozeboom (1989)] has been identified as the reliability of a linear composite

$$x^* = w_1x_1 + \ldots + w_mx_m \tag{2}$$

of nonequivalent subtests, there is no *requirement* on Equation 1 that its summed components w_ix_i be nonequivalent. Instead—unlike traditional formulas for the reliability of item sums—Equation 1 is not affected by how psychometrically equivalent (ideally, having true-score correlations approaching +1 and nearly identical true-score variances) the w_ix_i may be. Neither is there any requirement that the constituent variables $\{x_i\}$ in x^* all be psychometric subtests combining scores on a multiplicity of test items: They can be any mix of subtests, single items, and any other examinee data such as age, physical measurements, or school and employment records that may be relevant to the prediction target.

Moreover, the same composite test x^* can generally be parsed as a sum of weighted constituents

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 15, No. 1, March 1991, pp. XXX-XXX

0146-6216/91/010099-03$1.40

in many different ways, especially when some of its constituent x_i in one parsing are themselves composites of subconstituents. With one small caveat—namely, that different constituents should not have any subconstituents in common—Equation 1 applies equally well to all these parsings. Unless a test-retest or parallel-forms correlation is available for the entirety of x^*, this is the only plausible way to estimate the reliability of x^* if its weighted constituents have dubious homogeneity, particularly if some of its constituents that are positively correlated prior to weighting receive weights opposed in sign.

Clarification 2

However, the use of Equation 1 has an important prerequisite that strongly constrains its applicability in practice: Its estimate of x^* parsed as $w_1x_1 + , \ldots, + w_mx_m$ requires a reliability estimate for each constituent x_i. The weaker the confidence in those estimates, the less the yield of Equation 1 can be trusted for this parsing of x^*. Conversely, those parsings of x^* under which the estimates of the constituent reliabilities seem most plausible are clearly the ones to prefer. The primary alternatives for estimating the reliability of a prospective constituent x_i of x^* are as follows.

Test-Retest Correlation

Even when test-retest data on the entirety of x^* are unavailable, test-retest or parallel-forms correlations across different testing occasions (which, for simplicity, can be treated as the same) may be available for a constituent x_i. If contrivable, this should be the first choice; and all items in x^* for which simultaneous test-retest data are available should be combined into a single constituent of the preferred parsing of x^*.

Internal Consistency

When x_i is a weighted or unweighted sum of items, it is straightforward to compute an α estimate of the reliability of x_i by incorporating the items' weights, if any, into their scaling units. If there is reason to distrust the item homogeneity in x_i, a better solution is to use the composite's best split-half reliability in preference to α (see Rozeboom, 1966, p. 455ff on the superiority under item heterogeneity of well-chosen split-half reliabilities to more finely divided α coefficients.)

Educated Guessing

When there is no direct empirical measure of the reliability of x_i, rough estimates may be available through rational considerations. Some variables, such as examinees' vital statistics, are presumably measured with reliability approaching unity. In other cases, previous reliability data on measures similar in kind to x_i in examinee samples similar to the present target population may provide usable approximations. One virtue of Equation 1 is that if the reliability of any constituent x_i is uncertain but its upper and lower bounds can be plausibly estimated, Equation 1 can be used to map these into corresponding bounds on any composite reliability to which it contributes.

Program RELIAB

Although Equation 1 is quite simple algebraically, its numerical computation becomes increasingly tedious and error-prone as its constituent variables become more numerous than two or three, unless a computer is used. Microcomputer program RELIAB is now available that provides the reliability of a composite test, given the intercorrelation matrix of the constituent variables, their compositing weights, their standard deviations (needed if the weights are for raw scores), and their estimated reliability coefficients. This input is entered on the console keyboard one line at a time in response to screen prompts that provide essentially all the guidance a user needs. A copy of this program, its source and DOS-executable code, and a brief printable READ.ME documentation are available by sending $2 to the author at the address below.

References

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.) New York: McGraw-Hill.

Rozeboom, W. W. (1966). *Foundations of the theory of prediction.* Homewood IL: Dorsey Press.

Rozeboom, W. W. (1989). The reliability of a linear composite of nonequivalent subtests. *Applied Psychological Measurement, 13,* 277–283.

Acknowledgments

Nunnally (1978, pp. 246–254) established and discussed at some length a formula virtually identical to Equation 1. Information on additional precedents in the literature that merit recognition should be addressed to the author.

Author's Address

Send requests for reprints or further information to William W. Rozeboom, Department of Psychology, University of Alberta, Edmonton, Alberta, T6G 2E9 Canada.