Linear Dependence of Gain Scores on Their Components Imposes Constraints on Their Use and Interpretation: Comment on "Are Simple Gain Scores Obsolete?"

Lloyd G. Humphreys

University of Illinois, Urbana-Champaign

The properties of gain scores are linearly determined by the properties of their components. Thus, the reliability of a gain is uniquely determined by the reliabilities of the components, the correlation between them, and their standard deviations. Reliability is not inherently low, but the components of gains used in many investigations make low reliability likely. Correlations of the difference between two measures and a third variate are also determined uniquely by three correlations and two standard deviations. Raw score standard deviations frequently tell more about the measurement metric and how it is used than about the psychological processes underlying the measurements. Correlations involving gains/differences cannot be understood adequately unless the essential sample statistics of the components are known and reported. *Index terms: change scores, classical test theory, difference scores, gain scores, intraindividual differences, measurement of growth, reliability, test theory, validity.*

The defense of gain scores by Williams & Zimmerman (1996) contains many examples of the critical problem with gain scores: namely, linear determination of their properties by the properties of their components. This is demonstrated in the succession of graphs in which reliabilities and validities of gain scores are uniquely determined by the standard deviations and correlations of the components. However, Williams and Zimmerman did not discuss what linear determination implies for the planning and interpretation of research results. Many investigators will use the defense of gain scores as a green light to continue careless use and erroneous interpretations of gain scores.

Reliability

Gain scores are certainly not inherently unreliable. Increased reliability can occur if the pretest and posttest are negatively correlated. Nevertheless, gain scores not infrequently have reliabilities approaching 0 in psychological data. An increase is indeed a rare event. Investigators can readily estimate a reliability for a given gain score by plugging-in sample statistics of its components in a simple equation (see Williams & Zimmerman, 1996). The parallel forms reliability of the gain requires the parallel forms reliabilities of the components, the correlation between components, and the two standard deviations. In planning research, estimates can be substituted for the two reliabilities.

Reliabilities of differences must be sufficiently greater than 0 to meet the goals of the research. In research on individual differences, the expected maximum correlation of the difference between two components with a third variate is equal to the square root of the product of the reliabilities of the difference and the third variate. In experimental research in which a difference is the dependent measure, the combination

293

of low reliability and small effect size reduces power substantially (Humphreys & Drasgow, 1989).

Validity

The interpretation of the relations of differences with a third variate frequently neglects the linear determination of the properties of differences by the correlations and standard deviations of the components. This neglect can produce egregious errors of interpretation, overlooking that no new information is contained in a distribution of differences. The whole is always precisely equal to the sum of its parts.

A useful consequence of linear determination is that differences for each pair of scores of the components need not be computed; only the sample statistics of the components are needed. The equation for the correlation of the difference with a third variate provides a level of accuracy in each and every sample that is limited only by the number of decimal places retained.

The decision not to make the subtractions and to use an equation reinforces caution in interpretation and provides an opportunity for the investigator to select theoretically best weights for the correlations of the components with the third variate. There can be substantial gain by substituting a priori weights for the raw score standard deviations. When physical scales of measurement are used, arbitrariness of experimental arrangements can affect distributions of scores and thus the standard deviations. Many scales of measurement used by psychologists also have arbitrary zeros and units. Standard scores, as an alternative to obtained scores, can be investigated by substituting unities for the standard deviations of raw scores in the equation.

A focus on the properties of the components early in the analysis helps steer investigators to include in the research report full information about the components. Full information is required to understand relations with differences. If information about the parts is not published, inferences that investigators may not understand their own data are not unreasonable. Readers cannot understand the data without that knowledge. Editors should require component information in all research reports involving differences.

References

Humphreys, L. G., & Drasgow, F. (1989). Some comments on the relations between reliability and statistical power. *Applied Psychological Measurement, 13,* 419–425.

Williams, R. H., & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement, 20,* 59–69.

Author's Address

Send requests for reprints or further information to Lloyd G. Humphreys, Department of Psychology, University of Illinois at Urbana-Champaign, 603 E. Daniel Street, Champaign IL 61820, U.S.A. Email: lhumphre@s.psych.uiuc.edu.