

In staffing, for practical reasons, the scores of individuals are treated as if they were the attribute itself rather than merely indicators of the attribute. For example, scores on a mental ability test are interpreted as being synonymous with how intelligent individuals are. Alternatively, individuals' job performance ratings from their supervisors are viewed as indicators of their true performance.

Treated in this way, scores become a major input to decision making about individuals. For example, scores on the mental ability test are used and weighted heavily to decide which job applicants will receive a job offer. Alternatively, performance ratings may serve as a key factor in deciding which individuals will be eligible for an internal staffing move, such as a promotion. In these and numerous other ways, management uses these scores to guide the conduct of staffing activities in the organization.

The quality of the decisions made and the actions taken is unlikely to be any better than the quality of the measures on which they are based. Thus, a lot is at stake in the quality of the measures used in staffing. Such concerns are best viewed in terms of reliability and validity of measures.⁸

Reliability of Measures

Reliability of measurement refers to the consistency of measurement of an attribute.⁹ A measure is reliable to the extent that it provides a consistent set of scores to represent an attribute. Rarely is perfect reliability achieved, because of the occurrence of measurement error. Reliability is thus a matter of degree.

Reliability of measurement is of concern both within a single time period in which the attribute is being measured and between time periods. Moreover, reliability is of concern for both objective and subjective measures. These two concerns help create a general framework for better understanding reliability.

The key concepts for the framework are shown in Exhibit 7.5. In the exhibit, a single attribute, "A" (e.g., knowledge of programming languages), is being measured. Scores (ranging from 1 to 5) are available for 15 individuals. A is being measured in time period 1 (T_1) and time period 2 (T_2). In each time period, A may be measured objectively, with two test items, or subjectively, with two raters. The same two items or raters are used in each time period. (In reality, more than two items or raters would probably be used to measure A, but for simplicity only two are used here.) Each test item or rater in each time period is a submeasure of A. There are thus four submeasures of A—designated X_1 , X_2 , Y_1 , and Y_2 —and four sets of scores. In terms of reliability of measurement, the concern is with the consistency or similarity in the sets of scores. This requires various comparisons of the scores.

Comparisons Within T_1 or T_2

Consider the four sets of scores as coming from the objective measure, which used test items. Comparing sets of scores from these items in either T_1 or T_2 is called