With the SEM known, the range within which any individual's true score is likely to fall can be estimated. This range is known as a confidence interval or limit. There is a 95% chance that a person's true score lies within ±2 SEM of his or her actual score. Thus, if an applicant received a score of 22 on the test of knowledge of programming languages, the applicant's true score is most likely to be within the range of 22 ± 2(5), or 12–32.

Recognition and use of the SEM allow for care in interpreting people's scores, as well as differences between individuals in terms of their scores. For example, using the preceding data, if the test score for applicant 1 is 22 and the score for applicant 2 is 19, what should be made of the difference between the two applicants? Is applicant 1 truly more knowledgeable of programming languages than applicant 2? The answer is probably not. This is because of the SEM and the large amount of overlap between the two applicants' intervals (12–32 for applicant 1, and 9–29 for applicant 2).

In short, there is not a one-to-one correspondence between actual scores and true scores. Most measures used in staffing are unreliable to some degree, meaning that small differences in scores are probably due to error of measurement and should be ignored.

Relationship to Validity. The validity of a measure is defined as the degree to which it measures the attribute it is supposed to be measuring. For example, the validity of the test of knowledge of programming languages is the degree to which it measures that knowledge. There are specific ways to investigate validity, and these are discussed in the next section. Here, it simply needs to be recognized that the reliability with which an attribute is measured has direct implications for the validity of the measure.

The relationship between the reliability and the validity of a measure is

$$r_{xy} \leq \sqrt{r_{xx}}$$

where r_{xy} is the validity of the measure and r_{xx} is the reliability of the measure. For example, it had been assumed previously that the reliability of the test of knowledge of programming languages was r = .75. The validity of that test thus cannot exceed $\sqrt{.75} = 86$.

Thus, the reliability of a measure places an upper limit on the possible validity of a measure. It should be emphasized that this is only an upper limit. A highly reliable measure is not necessarily valid. Reliability does not guarantee validity; it only makes validity possible.

alidity of Measures

The validity of a measure is defined as the degree to which it measures the attribute it is intended to measure.[12] Refer back to Exhibit 7.1, which involved the