reliability of x and y). If these differences were controlled statistically, the variation in values would shrink and converge toward an estimate of the true validity of x. If that true r is significant (practically and statistically), one can indeed generalize validity of x across situations. Validity thus is not viewed as situation specific.

Indeed, the results in the exhibit reveal that the average (weighted by sample size) uncorrected validity is $\bar{r}_{xy} = .30$, and the average (weighted by sample size) validity corrected for unreliability in the predictor and criterion is $\bar{r}_{xy} = .36$. In this example, fully two-thirds (66.62%) of the variance in the correlations was due to study artifacts (differences in sample size and differences in reliability of the predictor or the criterion). Put another way, the variability in the correlations is lower once they are corrected for artifacts, and the validities do generalize.

An enormous amount of evidence supporting the validity generalization premise has accumulated. Some experts argue that validity generalization reduces or even eliminates the need for an organization to conduct its own validation study. If validity generalization shows that a selection measure has a statistically significant and practically meaningful correlation with job performance, the reasoning goes, why go to the considerable time and expense to reinvent the wheel (to conduct a validation study when evidence clearly supports use of the measure in the first place)? There are two caveats to keep in mind in accepting this logic. First, organizations or specific jobs (for which the selection measure in question is intended) can sometimes be unusual. To the extent that the organization or job was not reflected in the validity generalization effort, the results may be inapplicable to the specific organization or job. Second, validity generalization efforts, while undoubtedly offering more evidence than a single study, are not perfect. For example, validity generalization results can be susceptible to "publication bias," where test vendors may report only statistically significant correlations. Although procedures exist for correcting this bias, they assume evidence and expertise usually not readily available to an organization.[18] Thus, as promising as validity generalization is, we think organizations, especially if they think the job in question differs from comparable organizations, may still wish to conduct validation studies of their own.

A particular form of validity generalization that has proved useful is meta-analysis. Returning to Exhibit 7.18, meta-analysis reveals that the average correlation between x and y (i.e., $\bar{r}_{xy}$) is $\bar{r}_{xy} = .36$, that most of the variability in the correlations is due to statistical artifacts (and not due to true substantive differences in validity across studies), and that the validity appears to generalize. Meta-analysis is very useful in comparing the relative validity of selection measures, which is precisely what we do in Chapters 8 and 9.

Staffing Metrics and Benchmarks

For some time now, HR as a business area has sought to prove its value through the use of metrics, or quantifiable measures that demonstrate the effectiveness (or