Abstract
The studies of Reliability Generalization (RG) analyze estimates of the reliability of scores from a test provided by a set of studies. As their goals and designs are usually very varied, the sampling of individuals obeys very different schemas. Thus, the variances of the scores might be more heterogeneous than expected from random sampling. Two main problems associated with this potential source of heterogeneity should be taken into account. First, heterogeneity has been usually identified subjectively, not very rigorously. Second, once identified, it has not been taken into account in subsequent analyses. In previous papers, various ways to face both problems have been proposed. The procedures are summarized and applied to a set of 65 independent studies that report estimates of the internal consistency of the Beck Depression Inventory. The results show why any study of RG should take into account the heterogeneity of the variances. In addition to this, the only source that additionally accounts for significant variance in the coefficients is the version of the test employed: the second and third versions of the test involve significant increases in the internal consistency. The consequences of ignoring the heterogeneity of the variances are discussed.