

1 Overview

There are six parts to this volume. The first part, *Foundations*, sets the stage for the remainder of the volume, providing an historical perspective on score linking, definitions of types of score linkage, and background information on data collection designs, linking methods, and related assumptions. The remaining five parts deal with different types of linking scenario.

In the not too far past, linkings of any sort were treated as equatings. In this volume, we use equating to refer to the pinnacle of linking, the type of linkage that is sought, rarely achieved, and too often mistakenly presumed to have been attained. Equating is what large-scale testing programs engage in when they use large representative samples of examinees, sound data collection practices, and appropriate methods to link test editions built to the same set of specifications. Equating adjusts for differences in difficulty that occur with the use of different sets of similar test questions. Equating ensures that examinees are treated fairly. Part 2, *Equating*, focuses on linking scenarios in which the assumptions of equating are met.

A slightly different linking scenario arises when a testing program implements some form of change to their test, and wants to link scores across the old and new versions. The change might be in content, test administration conditions, or mode of administration. The change might be small or large. Testing programs that are in transition due to changes of any nature must face the question of whether scores from the previous version of the test can be viewed as interchangeable with scores from the new version of the test. Part 3, *Tests in Transitions*, discusses linking issues associated with this scenario.

Another linking scenario occurs when there is an interest in linking scores across related but distinct tests. Typically, the tests measure similar constructs, are administered to similar kinds of examinees, and are used for the same purpose, but differ in terms of specifications and perspective. Part 4, *Concordance*, deals with linking issues associated with this scenario.

An alternate linking scenario arises when there is an interest in making comparisons of performance across different levels of difficulty for a given construct. In the realm of K-12 testing, test scores are often compared across grades even though test content and test populations differ. Linkages of this sort must ensure that the comparisons are meaningful

despite the changes in content and examinees. Part 5, *Vertical Scaling*, discusses linking issues associated with this scenario.

The final linking scenario considered in this volume occurs when there is an interest in linking group-level scores to individual-level scores. For example, the accountability movement has triggered an interest in making meaningful quantitative comparisons across scores on the National Assessment of Educational Progress and assessments designed to measure whether individuals meet state standards. Part 6, *Linking Group Assessments to Individual Assessments*, looks at linking issues under this scenario.

In the final chapter of this volume, *Postscript*, we briefly review the transitions that occur from Parts 2 to 6. We take note of the diversity of perspective within parts. We discuss the descent of linking from the ideal state of equating to the realities faced by professionals who have to operate in arenas where the need for comparability is great and the capacity to achieve it is limited. We address the inevitable conflict that occurs when expectations exceed professional capabilities.