

Michael J. Kolen
Robert L. Brennan

Test Equating

Methods and Practices

With 36 Illustrations



Springer

Contents

Preface	vii
Notation	xv
CHAPTER 1	
Introduction and Concepts	1
Equating and Related Concepts	1
Equating in Practice—A Brief Overview of This Book	7
Properties of Equating	9
Equating Designs	13
Error in Estimating Equating Relationships	22
Evaluating the Results of Equating	23
Testing Situations Considered	24
Preview	25
Exercises	26
CHAPTER 2	
Observed Score Equating Using the Random Groups Design	28
Mean Equating	29
Linear Equating	30
Properties of Mean and Linear Equating	31
Comparison of Mean and Linear Equating	33
Equipercentile Equating	35
Estimating Observed Score Equating Relationships	47
Scale Scores	54
Equating Using Single Group Designs	62
Equating Using Alternate Scoring Schemes	63
Preview of What Follows	64
Exercises	64

CHAPTER 3	
Random Groups—Smoothing in Equipercentile Equating	66
A Conceptual Statistical Framework for Smoothing	67
Properties of Smoothing Methods	70
Presmoothing Methods	71
Postsmoothing Methods	83
Practical Issues in Equipercentile Equating	94
Exercises	103
CHAPTER 4	
Nonequivalent Groups—Linear Methods	105
Tucker Method	107
Levine Observed Score Method	111
Levine True Score Method	117
Illustrative Example and Other Topics	123
Appendix	134
Exercises	135
CHAPTER 5	
Nonequivalent Groups—Equipercentile Methods	137
Frequency Estimation Equipercentile Equating	137
Braun-Holland Linear Method	146
Chained Equipercentile Equating	147
Illustrative Example	149
Practical Issues in Equipercentile Equating with Common Items	154
Exercises	155
CHAPTER 6	
Item Response Theory Methods	156
Some Necessary IRT Concepts	157
Transformations of IRT Scales	162
Transforming IRT Scales When Parameters Are Estimated	167
Equating and Scaling	174
Equating True Scores	175
Equating Observed Scores	181
IRT True Score Versus IRT Observed Score Equating	184
Illustrative Example	185
Using IRT Calibrated Item Pools	200
Practical Issues and Caveat	207
Exercises	208
CHAPTER 7	
Standard Errors of Equating	210
Definition of Standard Error of Equating	211
The Bootstrap	213
The Delta Method	224

Using Standard Errors in Practice	240
Exercises	241
CHAPTER 8	
Practical Issues in Equating and Scaling to Achieve Comparability	244
Equating and the Test Development Process	246
Data Collection: Design and Implementation	250
Choosing From Among the Statistical Procedures	266
Choosing From Among Equating Results	272
Importance of Standardization Conditions and Quality Control	278
Conditions Conducive to Satisfactory Equating	283
Comparability Issues in Special Circumstances	283
Conclusion	292
Exercises	293
References	296
Appendix A: Answers to Exercises	307
Appendix B: Computer Programs and Data Sets	322
Index	324