

CONTENTS

PART 1
The Foundations of Mental Measurement Theory

Chapter 1
Measurement in Psychology and Education

1.1	The need for a theory of mental testing	13
1.2	Psychological theory and its function	15
1.3	Measurement as a basis of model construction	16
1.4	The place of measurement in psychology	19
1.5	Levels of measurement	20
1.6	The specification of interval scales	21
1.7	Deterministic and probabilistic models	23
1.8	The assumptions underlying test theory models	24

Chapter 2
The Construction of True and Error Scores

2.1	Introduction	27
2.2	The distribution of measurements on a fixed person	29
2.3	True score as an expectation	30
2.4	The construction of the error random variable for a fixed person	31
2.5	The random selection of persons	32
2.6	Construction of the linear model	34
2.7	Derivation of the usual assumptions of the classical model	36
2.8	What is error?	38
2.9	The many concepts of true score	39
2.10	Experimental independence	44
2.11	Linear experimental independence	45
2.12	Replicate measurements	46
2.13	Parallel measurements and parallel forms	47

PART 2**The Classical Test Theory Model****Chapter 3****Basic Equations of the Classical Model for Tests of Fixed Length**

3.1	The classical linear model: restatement of assumptions	55
3.2	Expectations, variances, and correlations	56
3.3	Relationships based on parallel measurements	58
3.4	Definitions, interpretations, and applications	60
3.5	The validities of a test	61
3.6	An alternative statement of the classical model	63
3.7	Regression theory for the classical model	64
3.8	Errors of measurement, estimation, and prediction	66
3.9	Attenuation formulas	69
3.10	Elementary models for inferring true change	74

Chapter 4**Composite Tests**

4.1	Introduction	82
4.2	Composite measurements with two components	83
4.3	Composite measurements with n components	85
4.4	Coefficient α and the reliability of composite measurements	87
4.5	The internal structure of tests	95
4.6	Expectations, variances, and covariances of weighted composites	96
4.7	The correlation between two composite measurements	97

Chapter 5**Basic Equations of the Classical Model
for Homogeneous Tests of Variable Length**

5.1	Test length as a test parameter	103
5.2	The classical model with a continuous test length parameter	104
5.3	Statement of the assumptions	105
5.4	The true score as the observed score of a person on a test of infinite length	108
5.5	The fundamental theorem	108
5.6	Expectations and variances	109
5.7	Covariances	110
5.8	Correlations among observed, true, and error scores	111
5.9	Expectations, variances, and correlations of lengthened tests	111
5.10	The Spearman-Brown formula	112
5.11	The effect of test length on validity	114
5.12	Comparing reliabilities and validities of tests of differing lengths	118
5.13	The most reliable composite with a specified true score	119

5.14	Maximizing the reliability of the composite when component lengths are fixed	123
5.15	Maximizing the validity of a test battery as a function of relative test lengths for a fixed total testing time	124

Chapter 6

Factors Affecting Measurement Precision, Estimation, and Prediction

6.1	Introduction	129
6.2	Effect of group heterogeneity on test reliability	129
6.3	Speed and power tests	131
6.4	Conditions of measurement affecting reliability	133
6.5	Experimental problems in correcting for attenuation	137
6.6	Accuracy of the Spearman-Brown prophecy formulas	139
6.7	Reliability as a generic concept	139
6.8	Effect of explicit and incidental selection on test validity: the two-variable case	140
6.9	The effect of selection on test validity: the three-variable case	144
6.10	The effect of selection on test validity: the general case	146
6.11	Accuracy of the selection formulas	147

Chapter 7

Some Estimates of Parameters of the Classical Model

7.1	Introduction	151
7.2	Estimating true score	152
7.3	An unbiased estimate of the specific error variance from parallel measurements	153
7.4	The use of estimated error variances	159
7.5	Specific true-score variance estimated from an analysis of variance components	160
7.6	A general formulation of the estimation problem as an analysis of variance components	162
7.7	An estimate of an upper bound on the specific error variance from measurements that are not strictly parallel	166

PART 3

Other Weak True-Score Models

Chapter 8

Some Test Theory for Imperfectly Parallel Measurements

8.1	Defining true score	173
8.2	The generic error of measurement	176
8.3	The generic error variance	177

8.4	Basic properties of generic errors of measurement	180
8.5	Generic true-score variance	184
8.6	The relation between generic and specific true-score variances	185
8.7	Estimating generic parameters describing a single test form	187
8.8	Comparisons of estimates of error variance	191
8.9	Substantive considerations regarding choice among estimates	194

Chapter 9

Types of Reliability Coefficients and Their Estimation

9.1	Introduction	198
9.2	Estimating the specific reliability coefficient	200
9.3	Statistical properties of an estimated variance ratio	201
9.4	Specific reliability theory for composite tests	203
9.5	Maximum likelihood estimation of reliability for normally distributed scores	204
9.6	The frequency distribution of the estimated reliability	206
9.7	The generic reliability coefficient	208
9.8	Generic reliability for a single test	209
9.9	Use and interpretation of reliability coefficients	211
9.10	The reliability of ordinal measurements	214
9.11	Use of factor loadings as reliability coefficients	216
9.12	Estimating reliability without using parallel forms	216

Chapter 10

**Some Test Theory for τ -Equivalent Measurements,
Including Estimation of Higher-Order Moments**

10.1	Introduction and definitions	224
10.2	An assumption of linear experimental independence	225
10.3	Immediate implications	226
10.4	Basic theorem for τ -equivalent measurements	227
10.5	Third-order moments	228
10.6	Higher-order moments and cumulants	230
10.7	Regression of true score on observed score	230
10.8	Implications, applications, and limitations	232

Chapter 11

Item Sampling in Test Theory and in Research Design

11.1	Introduction	234
11.2	Matrix sampling	236
11.3	Generalized symmetric means	238
11.4	First- and second-degree gsm's	241
11.5	Estimating true-score moments	245
11.6	Estimating the relation of observed score to true score	248
11.7	Estimating the relation between scores on parallel test forms	248

11.8	Estimating the observed-score statistics for lengthened tests	249
11.9	Frequency distribution of errors of measurement for binary items	250
11.10	Item sampling as a technique in research design	252
11.11	Estimating a mean from a single item sample	253
11.12	Estimating a mean by multiple matrix sampling	255
11.13	Estimating group mean differences	257
11.14	Estimating observed-score variances by item sampling	259

PART 4

Validity and Test Construction Theory

Chapter 12

Validity

12.1	Introduction	261
12.2	Regression and prediction	262
12.3	Linear regression functions	264
12.4	Multiple and partial correlation	265
12.5	Partial and multiple correlation and regression in n variables	267
12.6	The screening of predictor variables	269
12.7	Suppressor variables, moderator variables, and differential predictability	271
12.8	Incremental validity	273
12.9	Validity and the selection ratio	275
12.10	Some remarks on the explication of the concept of validity as a correlation coefficient	277
12.11	Construct validity	278

Chapter 13

The Selection of Predictor Variables

13.1	Introduction	284
13.2	Some sampling problems	284
13.3	Formal procedures for selecting predictor variables	288
13.4	Prediction in future samples	289
13.5	The effect of relative test lengths on reliability and validity: the multiple predictor case	293
13.6	The determination of relative test lengths to maximize the multiple correlation	295

Chapter 14

Measurement Procedures and Item-Scoring Formulas

14.1	Introduction	302
14.2	Guessing and omitting	303

14.3	A simple formula score	305
14.4	Properties of the simple formula score	307
14.5	A simple regression model for scoring items	310
14.6	The regression method with a simple model that assumes partial knowledge	312
14.7	Other item-scoring formulas	313
14.8	An evaluation of partial knowledge	314
14.9	Methods for discriminating levels of partial knowledge concerning a test item	315
14.10	Assumptions underlying the personal probability approach to item scoring	319
14.11	Reproducing scoring systems	321

Chapter 15

Item Parameters and Test Construction

15.1	Introduction	327
15.2	Item difficulty	328
15.3	Item discriminating power	329
15.4	Item validity	332
15.5	Product moment correlations for dichotomous items	335
15.6	Biserial correlation	337
15.7	Comparison of biserial and point biserial coefficients	340
15.8	Tetrachoric correlation	345
15.9	A comparison of tetrachoric and phi coefficients	346
15.10	Considerations in the choice of test construction techniques	350
15.11	Formula scoring and corrections for chance success	352
15.12	Invariant item parameters	353

Chapter 16

Latent Traits and Item Characteristic Functions

16.1	Introduction	358
16.2	Latent variables	359
16.3	Local independence	360
16.4	Item-test regression	363
16.5	The normal ogive model	365
16.6	Conditions leading to the normal ogive model	370
16.7	Correlation matrix with one common factor	371
16.8	A sufficient condition for normal ogive item characteristic curves	374
16.9	Normal ogive parameters: item difficulty	376
16.10	Normal ogive parameters: item discriminating power	377
16.11	Practical use of normal ogive item parameters	379
16.12	Conditional distribution of test scores	384
16.13	A relation of latent trait to true score	386
16.14	Typical distortions in mental measurement	387

PART 5**Some Latent Trait Models
and Their Use in Inferring an Examinee's Ability***(Contributed by Allan Birnbaum)***Chapter 17****Some Latent Trait Models**

17.1	Introduction	397
17.2	The logistic test model	399
17.3	Other models	402
17.4	The test as a measuring instrument: examples of classification and estimation of ability levels by use of test scores	405
17.5	Information structure of a test and transformations of scale of scores	410
17.6	Transformations of scales of ability	411
17.7	Calculations of distributions of test scores	414
17.8	Quantal response models in general	420
17.9	Estimation of item parameters	420
17.10	Validity of test models	422

Chapter 18**Test Scores, Sufficient Statistics,
and the Information Structures of Tests**

18.1	Sufficient statistics: definition and interpretation	425
18.2	Conditions for sufficiency of a statistic	428
18.3	Test scores and sufficient statistics	429
18.4	Sufficiency and the logistic test model	431
18.5	Sufficiency and the information structures of tests	434

Chapter 19**Classification by Ability Levels**

19.1	Classification rules for distinguishing two levels of ability	436
19.2	Two-point classification problems	437
19.3	Locally best weights and classification rules	442
19.4	More general classification rules, composite scores, and statistical efficiency in general	444
19.5	Quantitative appraisal and efficient design of classification rules	446

Chapter 20**Estimation of an Ability**

20.1	Introduction	453
20.2	Some algebra of information functions	453
20.3	More general methods of estimation: maximum likelihood	455
20.4	The information functions of various test items	460