



HOWARD WAINER

with

Neil J. Dorans

Daniel Eignor

Ronald Flaugher

Bert F. Green

Robert J. Mislevy

Lynne Steinberg

David Thissen

**Computerized
Adaptive
Testing: A PRIMER**
Second Edition

Contents

Foreword to the First Edition	ix
<i>C. Victor Bunderson</i>	
Foreword to the Second Edition	xiii
<i>Drew H. Gitomer</i>	
Preface to the First Edition	xvii
<i>Howard Wainer</i>	
Preface to the Second Edition	xxi
<i>Howard Wainer</i>	
1 Introduction and History	1
<i>Howard Wainer</i>	
Prologue	1
The First Four Millennia of Mental Testing	2
The Origins of Mental Testing in the U.S. Military	4
The Origins of Admissions Testing for American Universities	6
Computerized Adaptive Testing	9
Important Issues in CAT	11
Challenges for the Future	16
The Structure and Use of a Gedanken Computerized Adaptive Test (The GCAT)	17

vi CONTENTS

Acknowledgments	19
Annotated References	19

2 System Design and Operation 23
Bert F. Green

The Test Scenario	23
System Issues	26
References	33
Exercises/Study Questions	35

3 Item Pools 37
Ronald Flaughter

Introduction	37
Steps in the Development of an Item Pool	39
Illustration of GCAT Item Pool Construction	45
On the Topic of Dimensionality	56
Acknowledgments	57
References	57
Exercises/Study Questions	59

4 Item Response Theory, Item Calibration,
and Proficiency Estimation 61
Howard Wainer and Robert J. Mislevy

Introduction	61
Item Response Theory	63
Estimating Proficiency	68
Estimating Item Parameters	75
Other Topics	88
Technical Appendix: Linking Procedures	94
References	97
Exercises/Study Questions	100

5 Testing Algorithms 101
David Thissen and Robert J. Mislevy

Introduction and Background	101
Item Selection CATs	105

	An Illustration from GCAT	115	
	Topics Requiring Special Consideration	119	
	Two-Stage Testing and Testlets	123	
	Technical Appendix: Statistical Inference in CAT	128	
	References	130	
	Exercises/Study Questions	133	
6	Scaling and Equating		135
	<i>Neil J. Dorans</i>		
	Scores and Scales	136	
	Scaling and CAT: A Perspective	140	
	The GCAT and GP&P Scales	141	
	Equating	143	
	Equating and CAT: A Perspective	152	
	GCAT to GP&P Equating: An Illustration	153	
	References	156	
	Exercises/Study Questions	158	
7	Reliability and Measurement Precision		159
	<i>David Thissen</i>		
	Introduction and Background	160	
	Measurement Error	165	
	Other Sources of Error	171	
	Composite Scores	174	
	Comparison with Paper-and-Pencil Batteries	175	
	Illustrations of GCAT Measurement Precision	177	
	References	183	
	Exercises/Study Questions	184	
8	Validity		185
	<i>Lynne Steinberg, David Thissen, and Howard Wainer</i>		
	Construct-Related Validity	188	
	Criterion-Related Validity	194	
	Threats to Validity	209	
	References	225	
	Exercises/Study Questions	229	

9	Future Challenges	231
	<i>Howard Wainer, Neil J. Dorans, Bert F. Green, Robert J. Mislevy, Lynne Steinberg, and David Thissen</i>	
	Introduction	231
	Practical Issues	232
	What if the Model is Wrong?	235
	Model Elaboration	238
	Testlets	245
	Policy Issues	254
	New Possibilities	259
	References	264
	Exercises/Study Questions	270
10	Caveats, Pitfalls, and Unexpected Consequences of Implementing Large-Scale Computerized Testing	271
	<i>Howard Wainer and Daniel Eignor</i>	
	Introduction	271
	Examinee Access	272
	Item-Pool Usage and Security	274
	Economic Realities of CAT	283
	Operational Attempts at and Suggestions for Enhancing Security	286
	What Have We Learned That Can Be Used in the Future?	294
	Conclusion	296
	References	297
	Exercises/Study Questions	299
	References	301
	Abbreviations and Acronyms Used	317
	Author Index	319
	Subject Index	323