

# Fundamental Considerations in Language Testing

Lyle F. Bachman

دکتر لیل ف. باخمان

Oxford University Press

# Acknowledgments

The author and publishers would like to thank the following for permission to reproduce the material below that falls within their copyright:

The American Council on Education for the extracts from R. L. Thorndike (ed.): *Educational Measurement* (Second edition) and R. L. Linn (ed.): *Educational Measurement* (Third edition)

The American Psychological Association for the extract from the paper by S. A. Messick in *American Psychologist* 30 .

Brooks/Cole Publishing Company for the table on page 30, adapted from M. J. Allen and W. M. Yen: *Introduction to Measurement Theory*

The Center for Applied Linguistics for the extract from the paper by J. O. Carroll in *Testing the English Proficiency of Foreign Students*

*Educational and Psychological Measurement* for the extracts from the paper by C. I. Mosier in Volume 7

Jay Haley for the extract from *Strategies of Psychotherapy*

The John Hopkins University Press for the figure on page 341, adapted from A. J. Nitko, 'Defining the criterion-referenced test' in R. A. Berk (ed.): *A Guide to Criterion-Referenced Test Construction*, and for the extract from the same paper

*Language Learning*, and the authors, for the table on page 196, from the paper by J. van Weeren and T. J. J. Theunissen in Volume 37

Newbury House Publishers for three figures (from J. A. Upshur, 'Context for language testing') in J. W. Oller and J. C. Richards: *Focus on the Learner*

Pergamon Press PLC for the extract from K. Johnson (ed.): *Communicative Syllabus Design and Methodology*

*Psychometrika*, and the authors, for the extract from the paper by K. K. Tatsuoka and M. M. Tatsuoka in Volume 52

# Contents

Preface	viii
<b>1</b> Introduction	
The aims of the book	1
The climate for language testing	2
Research and development: needs and problems	8
Research and development: an agenda	12
Overview of the book	13
Notes	15
<b>2</b> Measurement	
inProduction	18
Definition of terms: measurement, test, evaluation	18
Essential measurement qualities	24
Properties of measurement scales	26
Characteristics that limit measurement	30
Steps in measurement	40
Summary	49
Notes	50
Further reading	52
Discussion questions	52
<b>3</b> <b>Uses</b> of Language Tests	
Introduction	53
Uses of language tests in educational programs	53
Research uses of language tests	67
Features for classifying different types of language test	70
Summary	78
Further reading	79
Discussion questions	79
<b>4</b> Communicative Language <b>Ability</b>	
Introduction	81

vi *Contents*

Language proficiency and communicative competence	<b>82</b>
A theoretical framework of communicative language ability	84
Summary	107
Notes	108
Further reading	109
Discussion questions	109
<b>5 Test Methods</b>	
Introduction	111
A framework of test method facets	<b>116</b>
Applications of this framework to language testing	<b>152</b>
Summary	156
Notes	<b>157</b>
Further reading	158
Discussion questions	<b>159</b>
<b>6 Reliability</b>	
Introduction	<b>160</b>
Factors that affect language test scores	<b>163</b>
Classical true score measurement theory	<b>166</b>
Generalizability theory	<b>187</b>
Standard error of measurement: interpreting individual test scores within classical true score and generalizability theory	<b>197</b>
Item response theory	<b>202</b>
Reliability of criterion-referenced test scores	<b>209</b>
Factors that affect reliability estimates	<b>220</b>
Systematic measurement error	<b>222</b>
Summary	<b>226</b>
Notes	<b>227</b>
Further reading	<b>232</b>
Discussion questions	<b>233</b>
<b>7 Validation</b>	
Introduction	<b>236</b>
Reliability and validity revisited	<b>238</b>
Validity as a unitary concept	<b>241</b>
The evidential basis of validity	<b>243</b>
Test bias	271
The consequential or ethical basis of validity	<b>279</b>
Postmortem: face validity	<b>285</b>

Summary	289
Notes	291
Further reading	294
Discussion questions	294
<b>8 Some Persistent Problems and Future Directions</b>	
<b>Introduction</b>	296
Authentic language tests	300
Some future directions	333
A general model for explaining performance on language tests	348
<i>Apologia et prolegomenon</i>	351
Summary	356
Notes	358
Further reading	358
Discussion questions	358
<b>Bibliography</b>	361
<b>Author index</b>	395
<b>Subject index</b>	397