

- **Part I**
- **Introduction**
- **1. What Are we Testing and Why?**
- Introduction
- Tests as Tools
- Test Purposes and Types
 - Curriculum-Related Tests
 - Other Types of Tests
- Ways of Looking at Tests
 - Norm-Referenced and Criterion-Referenced Testing
 - Summative vs. Formative Assessment
 - Objective vs. Subjective Testing
 - Direct vs. Indirect Testing
 - Discrete-Point vs. Integrated Tests
 - Performance Assessments: Focus on Task Completion vs. Focus on Language Use
- What we Want Tests to Tell us
- How Good Is Good Enough?
- Why Graphs and Descriptive Statistics Matter
- Summary
- Further Reading
- Discussion Questions
- **2. Tasks - The Building Blocks of Tests**
- Introduction
- Why Focus on Task Types at This Point?
- Task Format
- Where Test Tasks Come From
- Commonly Used Task Formats
 - Selected Response Tasks
 - Limited Production Tasks
 - Deletion-Based Tasks
- Cloze Tests
- C-Tests
 - Extended Production Tasks
- Contents note continued: Extended Production Tasks With Written Responses
- Extended Production Tasks With Spoken Responses
- Portfolios
 - Translation as a Task Format
- Summary
- Further Reading
- Discussion Questions
- **3. Planning and Designing the Test**
- Why Do we Need to Plan Tests?
- Writing Specifications
- Test Context and Purpose Specifications
 - Why Context and Purpose Specifications Matter
 - What to Include in Context and Purpose Specifications
- Evaluating Existing Tests
- Specifications of the Overall Test Structure

- Summary
- Further Reading
- Discussion Questions

- **4. Writing the Specifications For Individual Test Tasks**

- Introduction
- Specifications For Individual Tasks: General Considerations
- Specifications For Item-Based Task Formats
- Specifications For Reading and Listening Comprehension Questions
- Specifications For Reading and Listening Comprehension Passages
 - General Considerations
 - Reading Comprehension Passages
 - Listening Comprehension Passages
- Specifications For Vocabulary and Grammar Tests
- Specifications For Speaking and Writing Prompts
- More on Specifications For Interview-Based Speaking Tests
- Specifications Development as an Iterative Process
- Reverse Engineering Specifications From an Existing Test
- Summary
- Further Reading
- Discussion Questions

- **5. Writing the Test**

- Introduction
- Following the Plan
- Writing Passage-Based Test Tasks
 - General Considerations For Reading and Listening Passages
 - Creating or Selecting Listening Passages
- General Suggestions on Item Writing
 - Keeping Questions Clear and Construct-Relevant
 - Not Giving Away the Answers
 - Maintaining Item Independence
- Item Writing For Selected Response Task Formats
 - Multiple-Choice Questions
 - True-False Questions
 - Matching Items
- Item Writing For Limited Production Tasks
 - Short Answer Questions
 - Contents note continued: Deletion-Based Tasks
- Writing Prompts For Extended Production Tasks
 - General Considerations
 - Where to Get Ideas For Prompts
- Reviewing Passages, Items, Keys, and Prompts Before Pilot Testing
- Compiling the Test
- Item Analysis: A Conceptual Overview
- Summary
- Further Reading
- Discussion Questions

- **6. Consistency of Measurement**

- Introduction
- Understanding Reliability

- Approaches to Estimating Reliability
 - Parallel Forms and Test-Retest Reliability
 - Internal Consistency Reliability
- Estimating the Accuracy of Individual Scores
- Improving the Reliability of a Norm-Referenced Test
- Score Dependability
- Classification Dependability
- Estimating the Reliability and Dependability of Performance Tests
- Interpreting Estimates of Reliability and Dependability
- Summary
- Further Reading
- Discussion Questions

- **7. Rating Scales For Extended Production Tasks**
- Rating Scales and Scoring Rubrics
 - Analytic vs. Holistic Rating Scales
 - "Trait" Scales
 - "Objective" Scoring Approaches For Speaking and Writing
- Developing or Adapting Rating Scales
 - Developing Rating Scales: First Steps
 - Writing the Descriptors
- Rater Training and Norming
- Rating Procedures
- Summary
- Further Reading
- Discussion Questions

- **8. Validation**
- Introduction
- Previous Conceptions of Validity
- More Recent Views of Validity
 - Construct Validity and Validity Arguments
 - Consequences of Test Use, Critical Approaches, and Test Fairness
- An Expansion of Validity Arguments
- Setting Realistic Expectations
- Summary
- Further Reading
- Discussion Questions

- **9. Test Administration**
- Introduction
- Purpose and Goals of This Chapter
- Pilot Testing
- Preparing For a "Live" Administration
 - Planning Ahead
 - Assembling and Copying the Test
 - Scheduling Facilities and Equipment
 - Planning Tests With a Speaking Component
- Registration
- Admitting and Seating the Students
- Proctoring the Test
- Scoring the Tests and Entering the Data

- Reporting the Results
- Archiving and Destroying Tests and Data
- Test Accommodations
- Summary
- Further Reading
- Discussion Questions

- **10. Other Important Topics in Testing**

- Introduction
- Computer-Based Testing
- Item Response Theory
- Generalizability Theory
- Standard Setting
- Assessing Young Learners
- Assessing Language For Specific Purposes
- Verbal Protocol Analysis
- Discourse Analysis
- Corpus-Based Research Approaches
- Summary
- Further Reading
- Discussion Questions

- **PART II**

- **Introduction**

- **11. Organizing Data in Microsoft Excel**

- Introduction
- Objectives
- Getting to Know the Layout of Excel
- Moving Around and Selecting Cells
- Inserting, Deleting, and Moving
- Getting Data Into Excel
- Some Tips For Formatting
- Basic Mathematical Operations and Functions
- Relative and Absolute Cell References
- Contents note continued: Worksheet 11.1 Creating a Class Gradebook in Excel
 - Introduction
 - Procedures
 - Discussion
- Summary

- **12. Descriptive Statistics and Standard Scores**

- Introduction
- Objectives
- Overview of Descriptive Statistics
 - The Normal Distribution, and Describing the Shape of a Distribution
 - Measures of Central Tendency
 - Measures of Dispersion
 - What Are Good Values For Descriptive Statistics?
- Worksheet 12.1 Calculating Descriptive Statistics in Excel
 - Introduction

- Procedures
 - Discussion
- Overview of Standard Scores
- Worksheet 12.2 Calculating Standard Scores in Excel
 - Introduction
 - Procedures
 - Discussion
- Summary
- **13. Creating Graphs or Charts in Excel**
- Introduction
- Objectives
- Overview of Visual Representations of Data
- Preparing Excel For Histogram Analyses
 - Installing the Analysis ToolPak in Excel 2007
 - Installing the Analysis ToolPak in Excel 2003
- Worksheet 13.1 Creating a Histogram and a Frequency Polygon in Excel
 - Introduction
 - Procedures
- Setting up the Bins and Creating the Histogram
- Reformatting the Histogram
- Copying a Histogram and Converting the Copy to a Frequency Polygon
 - Discussion
- Summary
- **14. Correlation**
- Introduction
- Objectives
- Overview of Correlation
 - Explaining Correlation Levels of Data
 - Pearsonian Correlation Coefficients
- Worksheet 14.1 Calculating Pearson r and Spearman p in Excel
 - Introduction
 - Procedures: Calculating Pearson r
 - Procedures: Calculating Spearman p
- Discussion
- Summary
- **15. Item Analysis For Norm-Referenced Testing**
- Introduction
- Objectives
- Overview of Item Analysis
 - Item Difficulty
 - Discrimination
 - Interpreting Item Analysis Results
- Worksheet 15.1 Performing Item Analysis For an NRT
 - Introduction
 - Procedures
 - Discussion
- Worksheet 15.2 Performing Item Analysis For an NRT With Polytomous Data
 - Introduction
 - Procedures

- Discussion

Summary

- **16.Item Analysis For Criterion-Referenced Testing**
- Introduction
- Objectives
- Overview of CRT Item Analysis
 - Item Difficulty in CRT
 - Item Discrimination in CRT
 - Interpreting Item Analysis Results in CRT
- Worksheet 16.1 Performing Item Analysis For a CRT
 - Introduction
 - Procedures
 - Discussion
- Summary
- **17.Distractor Analysis and Scoring Multiple-Choice Items in Excel**
- Introduction
- Overview of Scoring and Distractor Analysis
- Example of How to Apply Distractor Analysis to an Item
- Worksheet 17.1 Distractor Analysis and Multiple-Choice Scoring in Excel
 - Introduction
 - Procedures: Computing Response Frequencies
 - Procedures: Scoring Multiple-Choice Responses
 - Procedures: Calculating the Point-Biserial For Each Option
 - Discussion
- Summary
- **18.Reliability: Scoring Consistency For Norm-Referenced Tests**
- Introduction
- Objectives
- Overview of Reliability
- Worksheet 18.1 Calculating Cronbach's Alpha and the SEM in Excel
 - Introduction
 - Procedures
 - Discussion
- Worksheet 18.2 Applying the Spearman-Brown prophecy Formula
 - Introduction
 - Procedures
 - Discussion
- Summary
- **19.Dependability: Scoring or Classification Consistency For Criterion-Referenced Tests**
- Introduction
- Objectives
- Overview of Dependability
- Worksheet 19.1 Estimating ϕ , the CI_{crt} , and $\phi(\lambda)$ From NRT Reliability Statistics
 - Introduction
 - Discussion
- Summary

Glossary

Bibliography

Index

CD-ROM List of Contents