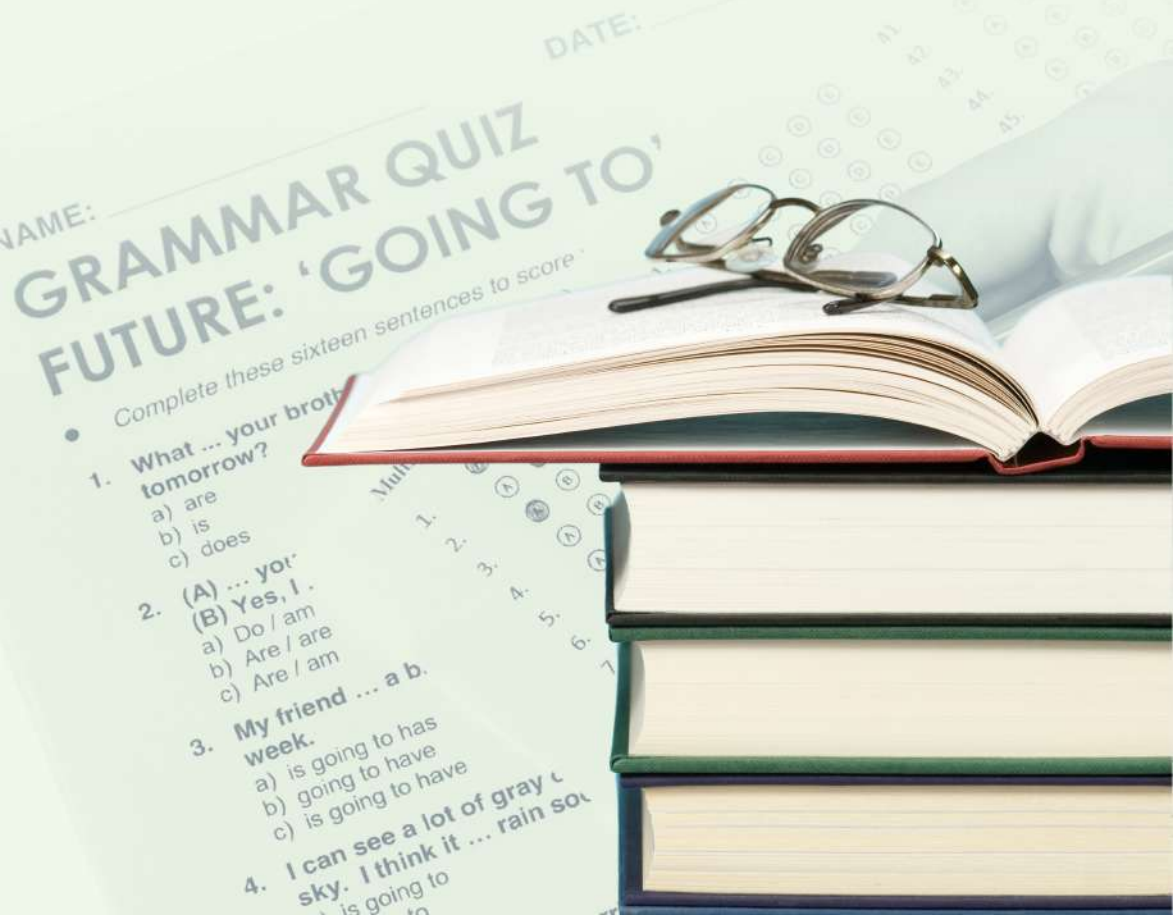


AN INTRODUCTION OF ENGLISH LANGUAGE TESTING

Editor : Dr. Bertaria Sohnata Hutauruk, S.Pd., M.Hum.

Penulis :

Sondang Manik | Ira Mutiaraningrum | Bahar Setiawan
Binti Su'aidah Hanur | Loso Judijanto | Lu'luil Maknun
Hera Hartati | Akhmad Hairul Umam | Muhammad Yazidus Syukri
Erika Sinambela | Weka Kusumastiti | Restu Arini | Amriani Hambali



Book Chapter

**An Introduction of
English Language Testing**

UU No 28 tahun 2014 tentang Hak Cipta

Fungsi dan sifat hak cipta Pasal 4

Hak Cipta sebagaimana dimaksud dalam Pasal 3 huruf a merupakan hak eksklusif yang terdiri atas hak moral dan hak ekonomi.

Pembatasan Pelindungan Pasal 26

Ketentuan sebagaimana dimaksud dalam Pasal 23, Pasal 24, dan Pasal 25 tidak berlaku terhadap:

- i. Penggunaan kutipan singkat Ciptaan dan/atau produk Hak Terkait untuk pelaporan peristiwa aktual yang ditujukan hanya untuk keperluan penyediaan informasi aktual;
- ii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk kepentingan penelitian ilmu pengetahuan;
- iii. Penggandaan Ciptaan dan/atau produk Hak Terkait hanya untuk keperluan pengajaran, kecuali pertunjukan dan Fonogram yang telah dilakukan Pengumuman sebagai bahan ajar; dan
- iv. Penggunaan untuk kepentingan pendidikan dan pengembangan ilmu pengetahuan yang memungkinkan suatu Ciptaan dan/atau produk Hak Terkait dapat digunakan tanpa izin Pelaku Pertunjukan, Produser Fonogram, atau Lembaga Penyiaran.

Sanksi Pelanggaran Pasal 113

1. Setiap Orang yang dengan tanpa hak melakukan pelanggaran hak ekonomi sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf i untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 1 (satu) tahun dan/atau pidana denda paling banyak Rp100.000.000 (seratus juta rupiah).
2. Setiap Orang yang dengan tanpa hak dan/atau tanpa izin Pencipta atau pemegang Hak Cipta melakukan pelanggaran hak ekonomi Pencipta sebagaimana dimaksud dalam Pasal 9 ayat (1) huruf c, huruf d, huruf f, dan/atau huruf h untuk Penggunaan Secara Komersial dipidana dengan pidana penjara paling lama 3 (tiga) tahun dan/atau pidana denda paling banyak Rp500.000.000,00 (lima ratus juta rupiah).

An Introduction of English Language Testing

Authors:

Sondang Manik
Ira Mutiaraningrum
Bahar Setiawan
Binti Su'aidah Hanur
Loso Judijanto
Lu'luil Maknun
Hera Hartati
Akhmad Hairul Umam
Muhammad Yazidus Syukri
Erika Sinambela
Weka Kusumastiti
Restu Arini
Amriani Hambali

Editor: Dr. Bertaria Sohnata Hutauruk, S.Pd., M.Hum.



AN INTRODUCTION OF ENGLISH LANGUAGE TESTING

Authors:

Sondang Manik
Ira Mutiaraningrum
Bahar Setiawan
Binti Su'aidah Hanur
Loso Judijanto
Lu'luil Maknun
Hera Hartati
Akhmad Hairul Umam
Muhammad Yazidus Syukri
Erika Sinambela
Weka Kusumastiti
Restu Arini
Amriani Hambali

Editor: **Dr. Bertaria Sohnata Hutauruk, S.Pd., M.Hum.**

Cover Design: **Nada Kurnia, S.I.Kom.**

Layouter: **Samuel, S.Kom.**

Size: **A5 Unesco (15,5 x 23 cm)**

Page: **xii, 247**

e-ISBN: **978-634-7037-56-5**

p-ISBN: **978-634-7037-57-2**

Terbit Pada: **December 2024**

Hak Cipta 2024, Pada Penulis

Isi diluar tanggung jawab percetakan

Copyright © 2024 by Future Science Publisher

All Right Reserved

Hak cipta dilindungi undang-undang
Dilarang keras menerjemahkan, memfotokopi, atau
memperbanyak sebagian atau seluruh isi buku ini
tanpa izin tertulis dari Penerbit.

PENERBIT FUTURE SCIENCE

(CV. FUTURE SCIENCE)

Anggota IKAPI (348/JTI/2022)

Jl. Terusan Surabaya Gang 1 A No. 71 RT 002 RW 005, Kel. Sumbersari, Kec. Lowokwaru, Kota
Malang, Provinsi Jawa Timur.
www.futuresciencepress.com

PREFACE

Above all, Thanks to Allah, the Almighty for the blessing and guidance to finish this book. I would like to thank to the authors for the idea and valuable knowledge to complete this book that entitled **An Introduction of English Language Testing**. Language testing is a broad category of testing that assesses aspects of a person's ability to understand or communicate in a particular language. Language testing is used for a variety of purposes. It is crucial to consider five primary principles of language assessment, such as practicality, reliability, validity, authenticity, and washback.

This book has been written by 13 authors with variety topics: (1) introduction, (2) testing, assessment and evaluation (3) pedagogy and classroom assessment, (4) test specification, (5) types of test, (6) criteria of a good test, (7) Integrated Language Test, (8) testing speaking (oral ability), (9) testing listening comprehension, (10) testing reading comprehension, (11) testing writing, (12) testing grammar and vocabulary, (13) scoring language test and assessment.

I also thank to a deepest appreciation to the publisher for giving support and encouragement in layouting, designing, editing and publishing the book. Again, I really appreciate for all the authors to share, write, and give the idea for this book. Thank you all for your valuable contribution to this book.

Pematang Siantar, November 2024

Editor,

Bertaria Sohnata Hutauruk

TABLE OF CONTENTS

PREFACE	v
TABLE OF CONTENTS	vi
CHAPTER 1 AN INTRODUCTION OF ENGLISH LANGUAGE TESTING.....	1
Sondang Manik	1
1.1. INTRODUCTION	1
1.2. CONNECTIONS TO OTHER DISCIPLINES.....	2
1.3. LANGUAGE TESTING.....	4
1.4. THE FIVE C'S	9
1.5. WHAT ARE THE APPROACHES TO LANGUAGE TESTING?.....	13
CHAPTER 2 TESTING, ASSESSMENT, AND EVALUATION. 19	
Ira Mutiaraningrum.....	19
2.1. INTRODUCTION	19
2.2. TESTING.....	21
2.2.1. TYPES OF TESTS	22
2.3. ASSESSMENT	25
2.3.1. TYPES OF ASSESSMENT.....	27
2.4. EVALUATION	32
2.4.1. TYPES OF EVALUATION	33
2.5. IMPLICATION	34
CHAPTER 3 PEDAGOGY AND CLASSROOM ASSESSMENT43	
Bahar Setiawan	43
3.1. INTRODUCTION	43

3.2.	CONCEPT OF PEDAGOGY	43
3.3.	CONCEPT OF CLASSROOM ASSESSMENT	45
3.4.	THE RELATIONSHIP BETWEEN PEDAGOGY AND CLASSROOM ASSESSMENT	48
3.4.1.	THE PEDAGOGICAL APPROACH UNDERPINNING CLASS ASSESSMENT	48
3.4.2.	ASSESSMENT AS A LEARNING INSTRUMENT	49
3.4.3.	AUTHENTIC ASSESSMENT IN THE CONTEXT OF MODERN PEDAGOGY	49
3.4.4.	FORMATIVE AND SUMMATIVE ASSESSMENT IN RELATIONSHIP WITH PEDAGOGY	50
3.4.5.	INTEGRATION OF COMPETENCY-BASED ASSESSMENT	51
3.4.6.	CHALLENGES IN HARMONIZING PEDAGOGY AND ASSESSMENT	51
3.4.7.	THE IMPORTANCE OF FEEDBACK IN EFFECTIVE JUDGMENT.....	52
3.5.	CONSTRUCTIVE PEDAGOGICAL APPROACH	52
3.5.1.	FUNDAMENTALS OF A CONSTRUCTIVE APPROACH.....	53
3.5.2.	BLOOM'S CONSTRUCTIVE APPROACH AND TAXONOMY	53
3.5.3.	IMPLEMENTATION OF CONSTRUCTIVE PEDAGOGY IN THE CLASSROOM	55
3.5.4.	CHALLENGES AND ADVANTAGES OF A CONSTRUCTIVE APPROACH.....	56
3.6.	CLASS ASSESSMENT RUBRIC	56
CHAPTER 4	TEST SPECIFICATION	61

	Binti Su'aidah Hanur	61
4.1.	INTRODUCTION	61
4.2.	TEST SPECIFICATION	61
4.3.	EXERCISES	72
4.4.	CONCLUSION.....	74
CHAPTER 5	TYPES OF TEST.....	77
	Loso Judijanto.....	77
5.1.	INTRODUCTION	77
5.2.	DEFINITION.....	79
5.3.	TYPE OF TEST.....	80
5.4.	THE IMPORTANCE OF ENGLISH LANGUAGE TESTING.....	82
5.5.	ENGLISH LANGUAGE WRITING SKILL TEST .	84
5.6.	ENGLISH LANGUAGE SPEAKING SKILL TEST.....	86
5.7.	ENGLISH LANGUAGE LISTENING COMPREHENSION SKILL TEST.....	89
5.8.	ENGLISH LANGUAGE READING COMPREHENSION SKILL TEST.....	90
5.9.	THE CAVEATS OF ENGLISH LANGUAGE TESTING.....	92
5.10.	ETHICAL ASPECTS OF ENGLISH LANGUAGE TESTING.....	95
5.11.	CONCLUSION.....	97
CHAPTER 6	CRITERIA OF A GOOD TEST	105
	Lu'luil Maknun.....	105
6.1.	INTRODUCTION	105

6.2.	VALIDITY	106
6.3.	RELIABILITY	111
6.4.	AUTHENTICITY	113
6.5.	WASHBACK	115
6.6.	FAIRNESS	118
6.7.	SCORABILITY	119
6.8.	CONCLUSION	121
CHAPTER 7	INTEGRATED LANGUAGE TEST	125
	Hera Hartati	125
7.1.	INTRODUCTION	125
7.2.	INTEGRATED LANGUAGE TEST	128
7.3.	INTEGRATED LANGUAGE TESTING: READING AND WRITING	130
7.4.	INTEGRATED LANGUAGE TESTING: LISTENING AND SPEAKING	133
7.5.	INTEGRATED LT READING, LISTENING AND WRITING	137
7.6.	INTEGRATED LANGUAGE TEST: LISTENING, READING AND SPEAKING	139
7.7.	CONCLUSION	141
CHAPTER 8	TESTING SPEAKING	145
	Akhmad Hairul Umam	145
8.1.	INTRODUCTION	145
8.2.	OVERVIEW OF LANGUAGE PROFICIENCY AND COMMUNICATIVE COMPETENCE.....	146
8.3.	IMPORTANCE OF AUTHENTICITY AND INTERACTION IN SPEAKING TESTS	146

8.4.	TYPES OF SPEAKING TESTS.....	147
8.5.	DESIGNING SPEAKING TESTS	148
8.6.	SCORING AND RATING SPEAKING TESTS....	150
8.7.	CHALLENGES IN SPEAKING ASSESSMENT ..	152
8.8.	BENEFITS AND LIMITATIONS OF TECHNOLOGY IN SPEAKING ASSESSMENT .	156
8.9.	CONCLUSION.....	161
CHAPTER 9	TESTING LISTENING COMPREHENSION	167
	Muhammad Yazidus Syukri	167
9.1.	DEFINITION OF LISTENING COMPREHENSION	167
9.2.	UNDERSTANDING THE PROCESS OF LISTENING COMPREHENSION.....	169
9.3.	ASSESSMENT OF LISTENING COMPREHENSION	170
9.4.	STRATEGIES FOR ENHANCING LISTENING COMPREHENSION	172
9.5.	SAMPLE ASSESSMENT MATERIALS	173
9.6.	STRATEGIES FOR IMPROVEMENT	175
9.7.	MODELS OF LISTENING COMPREHENSION .	175
9.8.	CONCLUSION.....	176
CHAPTER 10	TESTING READING COMPREHENSION	179
	Erika Sinambela.....	179
10.1.	INTRODUCTION	179
10.2.	STRATEGY IN TESTING READING COMPREHENSION	182
10.3.	TESTING READING COMPREHENSION	183

10.3.1. WORD AND SENTENCE MATCHING	183
10.3.2. TRUE/FALSE ITEMS	185
10.3.3. MULTIPLE CHOICE ITEM.....	186
10.3.4. GAP FILLING.....	188
10.3.5. COMPLETION ITEMS	189
10.3.6. REARRANGEMENT ITEMS.....	189
10.3.7. CLOZE PROCEDURE.....	191
10.3.8. OPEN-ENDED AND MISCELLANEOUS ITEMS	192
10.3.9. CURSORY READING	194
10.4. CONCLUSION	195
CHAPTER 11 TESTING WRITING	199
Weka Kusumastiti.....	199
11.1. INTRODUCTION	199
11.2. UNDERSTANDING WRITING	200
11.3. TEACHING WRITING	201
11.3.1. FOSTERING CREATIVITY AND PERSONAL EXPRESSION	202
11.3.2. TEACHING CRITICAL THINKING THROUGH WRITING.....	203
11.3.3. THE ROLE OF REVISION AND FEEDBACK....	204
11.4. TESTING WRITING	205
11.5. CONCLUSION	211
CHAPTER 12 TESTING GRAMMAR AND VOCABULARY ..	215
Restu Arini.....	215
12.1. INTRODUCTION TESTING GRAMMAR AND VOCABULARY	215

12.2.	TESTING GRAMMAR.....	216
12.3.	TESTING VOCABULARY	219
12.4.	CONCLUSION.....	225
CHAPTER 13	SCORING LANGUAGE TEST AND ASSESSMENT	229
	Amriani Hambali	229
13.1.	INTRODUCTION	229
13.2.	TYPES OF SCORING METHOD	232
13.3.	TYPES OF SCORING EVERY LANGUAGE SKILLS	235
13.4.	CONCLUSION.....	244

CHAPTER 6

CRITERIA OF A GOOD TEST

Lu'luil Maknun
Universitas Mercu Buana Yogyakarta, Yogyakarta
E-mail: lulu@mercubuana-yogya.ac.id

6.1. INTRODUCTION

Creating a good test in English language testing is crucial for accurately assessing a learner's proficiency and guiding their educational journey. A well-designed test not only measures the learner's current abilities but also identifies areas for improvement, providing valuable feedback for both students and educators. This process ensures that the test results reflect a true and fair evaluation of a learner's skills, rather than being influenced by poorly constructed questions or biases.

An effective English language test encompasses a variety of language skills, including reading, writing, listening, and speaking. This comprehensive approach allows for a more holistic assessment of a learner's abilities, as it captures their proficiency across different aspects of the language. For instance, a reading comprehension section might test a student's ability to understand and interpret written text, while a writing section evaluates their capacity to express ideas clearly and coherently. By incorporating multiple skills, a good test provides a balanced measure of overall language proficiency.

Additionally, well-crafted tests are essential for setting appropriate learning objectives and benchmarks. They help educators tailor their instruction to meet the specific needs of their students. For example, if a test reveals that a student struggles with grammar but excels in reading comprehension, the teacher can adjust their lessons to focus more on grammar

exercises. This targeted approach enhances the effectiveness of language instruction and supports students in achieving their learning goals more efficiently.

Moreover, creating a good test involves ensuring that it is fair and unbiased. This means avoiding questions that may favor certain cultural or linguistic backgrounds over others. A fair test gives all students an equal opportunity to demonstrate their abilities, regardless of their personal experiences or backgrounds. By adhering to principles of fairness and inclusivity, test designers contribute to a more equitable educational environment. In summary, the importance of creating a good test in English language testing lies in its ability to provide an accurate assessment of a learner's skills, guide targeted instruction, and ensure fairness. A well-designed test not only reflects a learner's proficiency but also supports their continued development and success in mastering the English language. Therefore, creating a good test in English Language Testing involves ensuring that the test is valid, reliable, practical, authentic, has positive washback, is fair, and is easy to score. By adhering to these criteria, educators and test designers can create assessments that are not only effective and reliable but also fair and meaningful. This book aims to provide a detailed exploration of each criterion, offering practical guidance and insights for creating high-quality English language tests, and delves into the essential criteria that ensure tests serve their intended purpose and positively impact language learning and teaching.

6.2. VALIDITY

Validity refers to the extent to which a test measures what it claims to measure. It is a crucial concept in psychological and educational testing because it indicates the accuracy and appropriateness of the inferences made based on the test scores.

Validity is a judgment that is made after considering evidence from all relevant areas (Nitko & Brookhart, 2011), other scholars said that validity is one of the core elements of measurement theory, and the focus of much of the practical work associated with psychometrics involves developing and reporting evidence for validity (Finch & French, 2019).

Types of Validity

There are several types of validity, each addressing different aspects of the test's effectiveness:

1. Content Validity

This type of validity assesses whether the test content covers the entire range of the concept being measured. For example, an English test with content validity would include a representative sample of all the topics in the curriculum.

2. Construct Validity

Construct validity evaluates whether the test truly measures the theoretical construct it intends to measure. This involves ensuring that the test is appropriately related to other measures and theories related to the construct. For instance, a test of intelligence should correlate with other established intelligence tests and theoretical understandings of intelligence.

3. Criterion-related Validity

This type examines how well one measure predicts an outcome based on another measure. It includes two subtypes:

- a. **Predictive Validity:** The test predicts future performance. For example, the SAT should predict college performance.
- b. **Concurrent Validity:** The test correlates well with a measure that has been previously validated and is taken

at the same time. For instance, a new depression scale might be validated against an established depression scale.

4. Face Validity

While not a scientific measure of validity, face validity refers to whether the test appears to measure what it is supposed to measure, based on a superficial examination. It is important for test takers' acceptance and trust in the test.

5. External Validity

This type assesses the extent to which the test results can be generalized to other contexts, groups, or settings. High external validity means the findings are applicable in various situations beyond the test environment.

6. Internal Validity

Internal validity refers to the degree to which the test results are attributable to the variables that the test aims to measure, rather than other confounding factors. Ensuring internal validity involves controlling for potential biases and confounding variables.

Each type of validity plays a critical role in establishing the overall validity of a test, ensuring that the test results are meaningful, accurate, and applicable to real-world scenarios.

Ensuring Validity in Test

Ensuring validity in test design is a multifaceted process that involves careful planning and attention to detail at every stage. Initially, the test's purpose must be clearly defined. Understanding what the test aims to measure guides the development of its content, ensuring that the test aligns with its intended objectives. This clarity is essential for achieving content validity, where the test must encompass all relevant aspects of the subject matter. For instance, if the test is designed

to assess English writing proficiency, it should cover all key areas such as grammar, punctuation, coherence, cohesive, reflecting the breadth of the curriculum. Ensuring validity in test design involves a systematic process to ensure that the test accurately measures what it is intended to measure. Here are steps to ensure various types of validity in test design:

1. Defining the Construct Clearly

By clarifying objectives, it clearly defines what the test is supposed to measure. This involves understanding the construct or domain comprehensively and reviewing literature, consult existing literature and theories related to the construct to ensure a thorough understanding.

2. Ensuring Content Validity

Involving subject matter experts in the test development process to ensure that all relevant content areas are covered, creating a test blueprint that outlines the content areas and the proportion of items dedicated to each area, and conducting pilot tests and gather feedback to refine the test items.

3. Enhancing Construct Validity

By developing clear operational definitions for each construct being measured, using statistical methods (e.g., factor analysis) to analyze item performance and ensure they align with the construct, and ensuring the test correlates with other measures of the same construct (convergent validity) and does not correlate with measures of different constructs (discriminant validity).

4. Establishing Criterion-related Validity

Incorporating criterion-related validity into test design involves examining how well the test predictions align with actual outcomes. For predictive validity, this means designing the test so that it accurately forecasts future performance or behaviors, such as using aptitude tests to

predict academic success. Concurrent validity is addressed by ensuring that the test scores correlate well with other established measures taken at the same time. This helps confirm that the test provides reliable and comparable results.

5. Improving Face Validity

By ensuring that the test items appear to measure what they are supposed to measure to the test takers and other stakeholders, and gathering feedback from potential test takers and stakeholders about the perceived relevance and clarity of the test items.

6. Maintaining External Validity

Using a sample that represents the population for which the test is intended and conducting Cross-validation means by testing the assessment in different settings, populations, and contexts to ensure the results are generalizable.

7. Ensuring Internal Validity

Implement procedures to minimize bias, such as random assignment and blinding; conduct pilot studies to identify and control potential confounding variables; and use statistical methods to control for confounding variables that might affect the test results.

8. Ongoing Evaluation and Revision**

Periodically review and update the test to ensure it remains relevant and accurate, continuously collect data on test performance and use it to make informed adjustments, and engage with stakeholders, including educators, psychologists, and test takers, to gather feedback and make necessary improvements.

In summary, ensuring validity in test design involves a comprehensive approach that integrates careful planning, rigorous testing, and ongoing evaluation. By addressing content,

construct, criterion-related, face, external, and internal validity, test developers can create assessments that are accurate, reliable, and meaningful.

6.3. RELIABILITY

Reliability refers to the consistency and stability of a measurement instrument or process over time. In other words, it's the extent to which an assessment tool produces stable and consistent results when the same thing is measured under similar conditions. Reliability can be defined as the proportion of variance in an observed test score that is accounted for by the construct being measured by the instrument (Finch & French, 2019). High reliability means that the tool will yield similar results under consistent conditions. Reliability in the context of English Language Testing refers to the consistency and stability of test scores across different administrations or conditions. In other words, a reliable test produces similar results under consistent conditions and with the same group of test-takers.

Types of Reliability

1. Test-Retest Reliability

This type assesses the consistency of test scores when the same test is administered to the same group of people at different points in time.

How It Works: After a period of time, the same test is given to the same group. High correlation between the two sets of scores indicates good test-retest reliability.

Example: Administering a vocabulary test to a group of students and then re-administering it a few weeks later to see if the results are similar.

2. Inter-rater Reliability

This type measures the degree to which different raters or scorers give consistent scores to the same test responses.

How It Works: Multiple raters score the same test responses independently. High agreement among the raters indicates high inter-rater reliability.

Example: In writing assessments, if different teachers score the same essay, their scores should be consistent for the test to be reliable.

3. Parallel-forms Reliability

This type examines the consistency of scores on different but equivalent forms of a test that measure the same construct.

How It Works: Two equivalent versions of a test are administered to the same group of people. High correlation between the scores on both forms indicates good parallel-forms reliability.

Example: Two versions of a grammar test with different questions but measuring the same grammar rules should yield similar scores.

4. Internal Consistency Reliability

This type assesses how consistently different items on the same test measure the same construct.

How It Works: It involves evaluating the correlation between individual items on the test. High internal consistency means that items on the test are measuring the same underlying construct.

Example: A reading comprehension test where all questions are related to the same passage should show high internal consistency if the questions are well-aligned with the passage content.

Factors Affecting Reliability in English Language Testing

1. **Test Length:** Longer tests generally provide more reliable estimates of a test-taker's ability because they cover a broader range of content.

2. **Test Administration Conditions:** Variability in testing conditions, such as noise or time of day, can affect reliability.
3. **Test-taker Factors:** Variability in test-takers' mood, health, or test-taking strategies can impact their performance and thus the reliability of the test scores.
4. **Item Quality:** Poorly written or ambiguous items can affect the consistency of test scores.

Methods to Improve Reliability in English Language Testing

1. **Standardization:** Ensure that test administration, scoring, and interpretation procedures are consistent across different test-takers and test administrations (Hambleton, R. K., & Jones, R. W., 1993).
2. **Clear and Precise Items:** Develop test items that are unambiguous and clearly aligned with the test objectives.
3. **Training for Raters:** Provide thorough training for raters to ensure consistent scoring and reduce variability in inter-rater reliability (McNamara, T., 1996).
4. **Pilot Testing:** Conduct pilot tests to identify and address potential issues with test items or administration procedures before the actual test administration (Bachman, 1990)
5. **Using Multiple Forms:** Create multiple equivalent forms of the test and use them in a balanced manner to minimize the effects of any one form.

Improving and ensuring reliability is crucial for creating fair and valid assessments in English language testing.

6.4. AUTHENTICITY

Authenticity in English language testing refers to the degree to which a test reflects real-world language use. An authentic test is designed to measure language skills in contexts and

situations that are representative of those the test-taker would encounter outside the test environment. This involves ensuring that the language tasks and materials used in the test are similar to those the learner would face in everyday life or in specific contexts relevant to their needs (Hasrol et al., 2022 & Widdowson, 1978)

Real-World Relevance in English Language Testing

Real-world relevance is crucial because it enhances the validity and practicality of the test. When tests reflect real-life situations, they are more likely to:

1. **Motivate Learners:** Learners are more engaged when they see the direct application of their skills.
2. **Provide Accurate Assessments:** Tests that mimic real-world tasks are better indicators of how well learners can use English in practical situations.
3. **Prepare Learners for Real Situations:** Authentic testing prepares learners for actual communication scenarios they might encounter, such as writing emails, giving presentations, or participating in conversations.

Designing Authentic Test Tasks in English Language Testing

Designing authentic test tasks involves creating test components that closely resemble real-world language use. Here's how to approach it:

1. **Identify Real-World Contexts:** Determine the contexts in which learners will use English (e.g., academic settings, business environments, social interactions).
2. **Incorporate Genuine Materials:** Use texts, dialogues, and scenarios from real-life sources such as newspapers, job interviews, or social media.

3. **Simulate Real Tasks:** Design tasks that mimic actual language use, such as writing a cover letter, giving a presentation, or participating in a group discussion.
4. **Ensure Cultural Relevance:** Make sure tasks are culturally appropriate and relevant to the learners' backgrounds and goals.

Examples of Authentic Test Tasks in English Language Testing

Examples of authentic test tasks might include:

1. **Email Writing:** Task learners with writing an email to a colleague requesting information or providing an update, reflecting a common workplace scenario.
2. **Role-plays:** Have learners engage in role-play exercises where they negotiate a business deal or resolve a customer service issue.
3. **Reading Comprehension:** Use real-world documents such as news articles or company reports and ask learners to summarize or respond to questions about the content.
4. **Presentations:** Ask learners to prepare and deliver a short presentation on a topic of their choice, mimicking a real-life situation like a conference or meeting.
5. **Interviews:** Simulate a job interview where learners must answer questions about their skills and experiences, reflecting real-world job-seeking scenarios.

These approaches ensure that language testing is not just about assessing knowledge but about evaluating practical language use in contexts that matter to learners.

6.5. WASHBACK

Washback is the effect that tests have on teaching and learning processes. It encompasses both the influence of tests on

teaching methods and the impact on what students learn. Green (2020) said that washback is a term used in education to describe the influence, whether beneficial or damaging, of an assessment on the teaching and learning that precedes and prepares for that assessment. In essence, washback describes how tests can shape the curriculum, instructional strategies, and student learning experiences.

Positive vs. Negative Washback

No	Positive Washback	Negative Washback
1	Encourages Better Teaching Practices: Tests that are well-aligned with learning objectives can motivate teachers to adopt effective teaching strategies that enhance students' skills in relevant areas.	Narrowing of Curriculum: Tests that focus on specific skills or content may lead teachers to concentrate only on what will be tested, neglecting other important areas of language learning.
2	Focuses on Relevant Content: Tests that accurately measure the desired language skills (e.g., reading, writing, speaking) can lead to a curriculum that emphasizes these skills.	Teaching to the Test: Teachers might focus primarily on test preparation rather than a holistic understanding of the language, which can limit students' overall language proficiency.
3	Promotes Student Engagement: When tests are perceived as fair and	Increased Stress: High-stakes tests can create pressure and anxiety for

	<p>meaningful, students are more likely to be motivated and engaged in their learning.</p>	<p>students, which can hinder their learning experience and performance.</p>
--	--	--

Strategies to Promote Positive Washback in English Language Testing

1. **Align Tests with Learning Objectives**
 Ensure that tests assess the skills and knowledge that are genuinely valued in the language learning curriculum.
2. **Use a Variety of Assessment Types**
 Incorporate different types of assessments (e.g., formative, summative, performance-based) to provide a well-rounded evaluation of students' language abilities.
3. **Promote Authentic Assessment**
 Design tests that reflect real-life language use and communication tasks, rather than just academic exercises.
4. **Encourage Comprehensive Teaching**
 Design tests in a way that encourages teachers to cover a broad range of language skills and not just test-specific content.
5. **Provide Feedback and Support**
 Offer constructive feedback on tests and provide support for both students and teachers to address any gaps or weaknesses identified.
6. **Minimize High-Stakes Pressure**
 Where possible, reduce the emphasis on high-stakes testing and incorporate continuous assessment methods to reduce stress and promote a more balanced approach to learning.

By considering these factors, educators can design tests that not only accurately measure language proficiency but also contribute positively to teaching and learning processes.

6.6. FAIRNESS

Fairness in the context of English testing refers to the principle that all test-takers should have an equal opportunity to demonstrate their language proficiency without any undue advantage or disadvantage. Bazvand and Rasooli (2022) said that fairness is an underpinning quality of assessment. This means the test should be designed and administered in a way that is free from bias and ensures that the results accurately reflect the true abilities of all individuals, regardless of their backgrounds or characteristics.

Avoiding Bias in Testing

1. Cultural Bias

Cultural bias occurs when the content of a test favors individuals from certain cultural backgrounds over others. This can happen if test questions assume specific cultural knowledge, experiences, or perspectives that are not universally shared. For example, a reading comprehension passage about a culturally specific holiday or event may disadvantage test-takers who are not familiar with that culture.

Ways to Avoid Cultural Bias

- a. **Diverse Content Review:** Include a diverse group of educators and experts in the test development process to review test items for cultural bias.
- b. **Inclusive Content:** Use texts and materials that are culturally neutral or represent a wide range of cultures.
- c. **Field Testing:** Conduct field tests with diverse groups to identify potential biases and adjust the test accordingly.

2. Linguistic Bias

Linguistic bias occurs when the language used in the test is more accessible to some test-takers than others. This can

happen if the test includes idioms, slang, or complex syntax that may not be familiar to all learners, particularly non-native speakers or those from different dialectal backgrounds. Ways to Avoid Linguistic Bias:

- a. **Simplified Language:** Use clear, straightforward language and avoid idiomatic expressions or culturally specific references.
- b. **Variety in Language Use:** Ensure that test items reflect a range of language uses and contexts to be more inclusive.
- c. **Review by Linguists:** Have linguists review test items to ensure they are free from unnecessary complexity and bias.

Ensuring Equal Opportunities for All Test-Takers

1. **Accommodations:** Provide necessary accommodations for test-takers with disabilities, such as extended time, alternate formats (e.g., Braille, large print), or assistive technologies.
2. **Access to Preparation Resources:** Ensure that all test-takers have access to preparation materials and practice tests. This can help level the playing field by giving everyone a fair chance to prepare for the test.
3. **Clear Instructions:** Provide clear and unambiguous instructions for all sections of the test to ensure that all test-takers understand what is expected of them.
4. **Consistent Testing Conditions:** Maintain consistent testing conditions, such as quiet environments and proper lighting, to ensure that external factors do not unfairly affect performance.

6.7. SCORABILITY

Scorability refers to the ease and accuracy with which test responses can be evaluated and assigned a score. High

scorability ensures that different scorers can assign similar scores to the same responses, ensuring fairness and consistency in the assessment process.

Objective vs. Subjective Scoring

	Objective Scoring	Subjective Scoring
Definition	Objective scoring involves test items with clear, unambiguous answers. Examples include multiple-choice questions, true/false questions, and matching exercises (Nitko, A. J., & Brookhart, S. M., 2011).	Subjective scoring involves evaluating responses that require interpretation, such as essays, short answers, and oral presentations (Gipps, C. V., 1994).
Advantages	<p>Consistency: Objective scoring is highly consistent, as there is little room for scorer interpretation.</p> <p>Speed: Tests can be scored quickly, often using automated systems.</p> <p>Fairness: Minimizes scorer bias, providing a level playing field for all test-takers.</p>	<p>Depth: Allows assessment of complex skills and deeper understanding of language use.</p> <p>Flexibility: Can accommodate a wide range of responses and creative answers.</p>

	(Brown, G. T. L., & Harris, L. R., 2014).	
Limitations	May not fully capture complex language abilities, such as writing proficiency or speaking skills.	Variability: Higher potential for scorer inconsistency and bias. Time-Consuming: Requires more time to evaluate and score responses.

6.8. CONCLUSION

In English language testing, a good test must meet several key criteria to ensure its effectiveness, fairness, and reliability. These criteria include Validity, Reliability, Practicality, Authenticity, Fairness, Washback, Transparency, and Comprehensiveness that should cover a wide range of language skills and abilities, including listening, speaking, reading, and writing. It should also assess both the understanding and use of language in various contexts. By adhering to these criteria, English language tests can provide accurate, reliable, and fair assessments of language proficiency, which are crucial for making informed decisions in educational, professional, and personal contexts.

REFERENCES

- Bachman, L. F. (1990). *Fundamentals of Language Testing*. Oxford University Press.
- Bazvand, A. D., & Rasoolo, A. (2022). Students' experiences of fairness in summative assessment: A study in a higher education context. *Studies in Educational Evaluation*, 72. <https://doi.org/10.1016/j.stueduc.2021.101118>

- Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*, 3(1), 22-30. <https://doi.org/10.14786/flr.v2i1.24>
- Finch, W.H., & French, B. (2019). *Educational and Psychological Measurement*. Routledge, Taylor&Francis Group.
- Gipps, C. V., (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. Routledge. <https://doi.org/10.4324/9780203486009>
- Green, A. (2020). Washback in language assessment. *Wiley Online Library*. <https://doi.org/10.1002/9781405198431.wbeal1274.pub2>
- Hambleton, R. K., & Jones, R. W. (1993). *Comparison of Item Response Theory and Classical Test Theory in the Context of Test Development*. In R. W. Jones & R. K. Hambleton (Eds.), *Handbook of Test Development* (pp. 83-114). Lawrence Erlbaum Associates.
- Hasrol, S., Zakaria, A., Aryadoust, V. (2022). A systematic review of authenticity in second language assessment. *Research Method in Applied Linguistic, 1*. <https://doi.org/10.1016/j.rmal.2022.100023>.
- McNamara, T. (1996). *Measuring Second Language Performance*. Longman.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational Assessment of students*. (6th ed.). Pearson Education.
- Widdowson, H. G. (1978). *Teaching language as communication*. Oxford University Press.

AUTHOR' PROFILE



Lu'luil Maknun

The author was born in Brebes on January 2, 1987. She is a permanent lecturer in the English Education Study Program at the Faculty of Teacher Training and Education, Universitas Mercu Buana Yogyakarta. She earned her Bachelor's degree in English Education and her Master's degree in the English Department at Yogyakarta State University. The author is currently pursuing a doctoral degree in Language Education Science at Yogyakarta State University.

AN INTRODUCTION OF ENGLISH LANGUAGE TESTING

- BAB 1** : An Introduction of English Language Testing
Sondang Manik
- BAB 2** : Testing, Assessment, and Evaluation
Ira Mutiaraningrum
- BAB 3** : Pedagogy and Classroom Assessment
Bahar Setiawan
- BAB 4** : Test Specification
Binti Su'aidah Hanur
- BAB 5** : Types of Test
Loso Judijanto
- BAB 6** : Criteria of a Good Test
Lu'luil Maknun
- BAB 7** : Integrated Language Test
Hera Hartati
- BAB 8** : Testing Speaking
Akhmad Hairul Umam
- BAB 9** : Testing Listening Comprehension
Muhammad Yazidus Syukri
- BAB 10** : Testing Reading Comprehension
Erika Sinambela
- BAB 11** : Testing Writing
Weka Kusumastiti
- BAB 12** : Testing Grammar and Vocabulary
Restu Arini
- BAB 13** : Scoring Language Test and Assessment
Amriani Hambali



FUTURE SCIENCE

Jl. Terusan Surabaya, Gang 1 A No. 71 RT 002 RW 005,
Kel. Sumbarsari, Kec. Lowokwaru, Kota Malang,
Provinsi Jawa Timur.
Website : www.futuresciencepress.com



IKAPI
IKATAN PENGAJARIT INDONESIA

No. 348/JTI/2022

ISBN 978-634-7037-57-2
ISBN 978-634-7037-56-5 (PDF)



9 786347 037572