VFAST Transactions on Education and Social Sciences
http://vfast.org/journals/index.php/VTESS@ 2022, ISSN(e): 2309-3951, ISSN(p): 2411-0221
Volume 10, Number 2, April-June, 2022 *pp:73-82*

Development and Analysis of Objective test Items of English at Intermediate Level (Part-I)

Amir Hafeez Malik[1], Rashda Perveen[2], Abdullah Shafiq[3], Abdul Rashid*[4]

[1]Assistant Professor of English, Emerson University, Multan, Pakistan
[2]visiting Lecturer in English, IBF, Bahauddin Zakariya University, Multan, Pakistan
[3]Lecturer in English, Institute of Central Asia, Multan, Pakistan
[4]Lecturer in English, Bahadur Sub Campus Layyah, Bahauddin Zakariya University, Multan, Pakistan
*Corresponding author email address: abdulrashid@bzu.edu.pk

ABSTRACT

The present study presents the analysis of current practices in development and analysis of objective test items for Intermediate Part-I level in Punjab, Pakistan. The study is preceded with the help of teachers teaching English to Intermediate Part-I class. A Test consisting of 60 items was produced for students. This Test was the basic tools for collections of data for the study. For each item of the Test, four options were given and students had to tick only one of them. The data collected through Test was then analyzed, tabulated, difficulty level and discrimination powers were calculated and each item was interpreted separately. The key point of the study is that a number of issues are being faced by both the teachers in class, as test designers, as examiners and also by the students as testees. Therefore, this study provides guides to develop and analyze object test items for Intermediate Part-I level. It is aimed to bring reliability and validity to development and analysis of objective test items for a better education system.

KEYWORDS

Testing & Evaluation, Test items, Objective, Practice

JOURNAL INFO

HISTORY: Received: March 14, 2022 Accepted: May 12, 2022 Published: May 19, 2022

INTRODUCTION

Language testing is the most significant component of academic process and system. The complete process of testing and its results is largely and highly dependent upon the test items designed for the purpose of evaluation. The more valid and reliable the test items are the more standardized results become. Imparting knowledge is one thing and measuring the same is another. The measurement and evaluation of the level of knowledge of learners requires a particular insight and different skills. In today's educational world, the scholars have accorded the status of an art to the capability of a teacher, which he possesses and exhibits in testing and ranking his students at various levels and situations. Any lad or deficiency is ill-affordable as test results these days have an undeniable significance in determining the course of further education. It has been observed and verified that object test items of Intermediate Part-I level in Punjab, Pakistan need to be paid due care as these items at present cannot be labeled valid, reliable and standardized test items. A number of related issues can be highlighted and handled with the help of studies and researched like the present study. In current academic world, we need expert evaluators and examiners, who specialize in designing, conducting and evaluating different tests and in different examinations. This has specially become true with an ever-increasing importance of objective tests in modern times. The prime point is to avoid any possible weakness in test items to give reliable results to the students and implement an authentic examination and evaluation system.

STATEMENT OF THE PROBLEM

All examining bodies (Boards and Universities) in Pakistan include objective type test items in all subject papers yet standard items are needed to evaluate the standards (difficulty level and discrimination power) of these objective type test items in the papers. This study has focused on development and analysis of objective test items of English at Intermediate Level (Part-I) with the view to provide solution to the problem that currently the objective test items are not standardized and need improvement.

SIGNIFICANCE OF THE STUDY

1. This research may provide an item-bank of standard test items for the teachers and the Paper-setters of Objective Tests at Intermediate (Part-I) Level.
2. It may provide effective guideline to the students for the preparation of their Objective Test Examinations.
3. This research work may guide the new researchers of objective test items.

4. The Educational Institutions may benefit from this research work to improve their standard of test designing and examination by having the standard test items.

REVIEW OF LITERATURE

Language testing can be studied, explored and researched under the umbrella of applied linguistics. The role and place of language testing is self-evident in teaching learning process. It has multiple importance and implementations. As a matter of fact language testing serves the purpose of making needed changes in teaching, placing students, diagnosing the teaching learning issues and offering appropriate solutions for improvement and achieving set targets.

Language testing has its history. It has been the subject of research since its introduction. Language testing has gone through different phases of development from past to present. The same will continue in future. Bachman (1999) presents a brief review of language testing by investigating language from 1977. Bachman (1982) stated that in the 1980's language testing further extended significantly under the influence of SLA research. SLA research prompted language testers to investigate a wide variety of factors such as aptitude, affect, field dependence, academic discipline and discourse domains on language test performance and the strategies involved in the process of test-taking itself. The work of applied linguists such as Widdowson (1983), Savignon (1983), Canale and Swain (1980) prove extremely significant and effective to language testing in the decade to come. The 1980's experienced a broadening of the issues and concerns of language testing into other areas of applied linguistics. Bachman (1999) maintains that the 1990's have seen expanded connections between language testing and mainstream educational measurement.

It can be said that effective language teaching learning is largely dependent on standardized language testing. It goes without saying that language testing has been a tradition in language teaching learning since ages. Its significance has multiplied with the passage of time. The major purpose of language testing is to determine the level of learning, granting of grades to the students and to distinguish the achievers as high, low, and no achievers. Language testing has multiple implementations for both the teachers and the learners. In teaching learning process, it is sharing of knowledge as well as assessment of knowledge. Language testing has a variety of language tests such as proficiency tests, achievement tests, aptitude tests, placement tests and diagnostic tests. Lewkowicz (2000) points out the concern that language learners exposed to test materials may not be representatives of target language.

Spolsky (1978) divides language testing into three periods such as: first is pre-scientific, second is the psychometric structuralists and the third is integrative sociolinguistic period. Heaton (1990) classifies the history of language testing into four major eras of: essay translation, the structuralist, the integrative and the communicative approach. Brown (2005) highlights five requirements of communicative tests. Oller (1979) states that cloze test or dictation can provide measures of overall language proficiency. Canale and Swain (1980) and Batchman (1990) opined that language testing should focus competence and performance. Brown (1996) used the term movements instead of period on the ground that somehow these overlap chronologically as well as co-exist. It is evident that the history of language testing has been the study and investigation of different trends or movements in the realm of language testing from the onset. Historically speaking, it can be said that language testing as a field of study has its past, present and will have its future in days to come.

Table1. Types of Test

1. Objective Test Types	
1. SUPPLY TYPE:	**2. SELECTION TYPE**
(i) Short Answer **(ii)**Completion	**(i)** True / False **Or** Alternate Response **(ii)** Matching **(iii)** Multiple Choice
2. Essay Test Types	
(i) Extended Response	**(ii)** Restricted Response

OBJECTIVE TESTS

Objective tests measure both your ability to remember facts and figures and your understanding of course materials. These tests are often designed to make you think independently, so don't count on recognizing the right answer. Instead, prepare yourself for high level critical reasoning and making fine discriminations to determine the best answer. The most common objective test questions are multiple-choice, true-false, and matching items. Doing well on these questions requires that you not only master the information but also interpret the test-maker's intentions. You know you have mastered the information if you can:

1. Recall specific terms, facts, names, and other key words; become proficient in the language of the course.
2. Distinguish the ways in which ideas, facts, theories, or other observations differ from each other and categorize ideas, facts, theories, or other observations according to the ways these are similar.
3. Answer the questions and solve the problems in the text and create your own questions or problems.

PREPARING FOR OBJECTIVE TESTS

1. Review notes and text(s) - list the major concepts that have been covered.
2. Highlight topics that were stressed. Note why they were stressed.
3. Think vocabulary. Every field of study has its own vocabulary, so identify words and terms used to represent specific concepts (i.e., the word "paradigm" in a social science course), and treat them as you would a foreign language. Make flash cards for frequent drills, and try to use these words whenever you work with course-related materials.
4. Compare and contrast. Sometimes objective questions can be used to test your ability to distinguish concepts, ideas, theories, events, facts from each other. Construct diagrams, charts, tables or lists to summarize relationships.
5. Recite for precision. Review your retention of the information by recalling it often. Use odd moments, in addition to 15-20 minute review sessions, to say or write out complete ideas and facts. It is very important to verbalize the recalled information completely and in a detailed manner so that you will have a precise idea of your mastery of the material.

MULTIPLE CHOICE QUESTIONS

1. Probably the most commonly used objective question, the multiple choice question, consists of two parts:
    1. The stem - the statement or question.
    2. The choices - also known as the distracters. There are usually 3 to 5 options to complete the stem statement or question.
2. Read the stem as if it were an independent, free-standing statement. Anticipate the phrase that would complete the thought expressed, and then compare each answer choice to your anticipated answer. It is important to read each choice, even if the first choice matches the answer you expected, because there may be a better answer listed.
3. Another evaluation technique is to read the stem together with each answer choice as if it were a true-false statement. If the answer make the statement as false one, cross it out. Check all the choices
4. The complete the stem as a true statement. Try to suspend judgment about the choices you think are true until you have read all the choices.
5. Beware of words like not, but, except. Mark these words because they specify the direction and limits of the answer.
6. Also watch out for words like always, never, and only. These must be interpreted as meaning all of the time, not just 99% of the time. These choices are frequently incorrect because there are few statements that have no exceptions (but there are a few).
7. If there are two or more options that could be the correct answer, compare them to each other to determine the differences between them, and then relate these differences with the stem to deduce which of the choices is the better one. (Hint: select the option that gives the most complete information.)
8. If there is an encompassing answer choice, for example "all of the above and you are able to determine that there are at least two correct choices, select the encompassing choice.
9. Use hints from questions you know to answer questions you do not.
10. If you do not find an answer, try to relate each answer to the stem to evaluate which one logically completes the thought.
11. Make educated guesses--eliminate options any way you can.

MULTIPLE CHOICE TESTS CHARACTERISTICS

1. It is a selection type test
2. It consists of a problem and list of suggested solutions
3. The problem is called the stem of the item
4. The list of suggested solutions is called alternatives or choices or options
5. The problem may be a direct question or incomplete statement
6. The alternatives may include words, numbers, symbols or phrases
7. The pupils are requested to select one correct or best alternative
8. The correct alternative is called answer
9. Remaining alternatives are called distracters as these distract the student from correct answer

ADVANTAGES

1. The Multiple Choice item is the most versatile type measuring simple learning outcome (information) to complex learning outcome (understanding, application)
2. Use of scoring machine or key makes the scoring easy
3. The degree of guessing is low (25%) as compared with true / false items (50%)

4. These items are less ambiguous as compared to true / false items

LIMITATIONS

1. It is difficult to construct MCQs. It needs expertise
2. More time is required to attempt MCQs in comparison to other types
3. Examinees who are familiar with the technique of MCQs, get high marks while having low capability

PRINCIPLES FOR CONSTRUCTING MCQS

1. The stem of the item should be meaningful and should present a definite problem
2. The stem should include much of the item and alternatives should be brief
3. An item should contain only one correct or clearly best answer
4. Alternatives should be homogeneous
5. Verbal associations between the stem and correct answer should be avoided
6. The correct answer should appear in random order
7. Use sparingly special alternatives such as “none of the above” or “all of the above”
8. All of the alternatives should be grammatically consistent with the stem of the item
9. Avoid the use of synonyms as options alternatives
10. Use guessing correction formula for scoring
11. Avoid a negatively stated item stem
12. The relative length of the alternatives should be approximately equal

ROLE OF VALIDITY & RELIABILITY IN OBJECTIVE TESTS

Tests and other evaluation instruments serve a variety of uses in the school. For example, tests of achievement might be used for selection, placement, diagnosis, or certification of mastery; aptitude tests might be used for predicting success in future learning activities or occupations; and appraisals of personal-social development might be used to understand better pupils’ learning problems or to evaluate the effects of a particular school program. Regardless of the type of instrument used or how the results are to be used, however, all of the measurements should possess certain characteristics. The most essential of these are validity, reliability, usability and objectivity.

Validity

Validity is the degree to which a test measures what it is supposed to measure.

Meter is used for the purpose of measuring length while scale is used for the purpose of measuring weight. If meter is used for measuring length, it will be valid because it is supposed to measured length and if it is used for measuring weight, it will not be valid.

Suppose a test is developed to measure the scientific concepts for a particular class, it will be valid for this specific purpose. It will not be valid for any other purpose. If someone tries to assess English comprehension from this test, then the test will not be valid.

Tests measuring the personality: In the same way are designed for a variety of purposes. A test designed for measuring achievement in the subject of computer science will not be valid for a test designed for measuring achievement in the subject of computer science for class 5 will not be valid for measuring achievement in the subject of computer science.

Types of Validity:

(i) Construct validity: is the degree to which a test measures an intended hypothetical construct.

(ii) Content validity is the degree to which a test measures an intended content area.”

RELIABILITY

Reliability is the degree to which a test consistently measures whatever it measures.

A reliable test gives the same scores when administered and re-administered while an unreliable test does not give the same scores. If an intelligence test was unreliable, then a student scoring and IQ of 120 today might scores an IQ of 140 tomorrow and a 95 day after tomorrow. If the test was reliable and if the student IQ was 110, then we would not expect fluctuation in score from testing to testing. A score of 105 would not be unusual but a score of 145 would be very unlikely. Reliability is expressed numerically as a coefficient. A high coefficient indicates high reliability and a low coefficient indicates low reliability. If a test were perfectly reliable, the coefficient would be 1.00.

A valid test is always reliable but a reliable test is not necessary valid.

TYPES of RELIABILITY

(i) Test Retest Reliability

Test retest reliability is the degree to which scores are consistent over time. Test retest reliability is established by determining the relationship between scores resulting from administering the same test, to the same group at different times.

Steps

- Administer the test

- After sometime (say two weeks) administer the same test to same group
- Correlate the two sets of scores
- High correlation coefficient indicates good test retest reliability

(ii) Equivalent-Forms Reliability:

Equivalent forms of a test are two tests that are identical in every way except for the actual items included.
The two tests are identical in term of

- Number of items
- Structure
- Difficulty level
- Directions for administering, scoring and interpretation
- Measuring variable

(iii) Split-Half Reliability:

Split-half reliability is determined by establishing the relationship between the scores on two equivalent halves of a test administered to a group at one time.

In this type, test items are divided into two equal halves. Suppose a test consists of 20 items, it can be divided into two equal halves as under:

1. Odd numbers: 1,3,5,7,9,11,13,15,17,19
2. Even numbers: 2,4,6,8,10,12,14,16,18,20

The scores on odd numbers are correlated with even numbers to compute the split-half reliability. The correction formula known as Spearman-Brown prophecy formula is used which is as under:

$$\textbf{r Total test}=\frac{2rSplit-half}{1+rSplit-half}$$

Item Difficulty

It deals with how difficult is the test item. It is indicated by the percentage of pupils who got the item right. It is recommended that test item would be neither too easy nor too difficult.

Steps:

1. Arrange the papers in order from the highest to the lowest scores (say 80 papers)
2. Select 25% papers with the highest scores (high achievers)-20 papers
3. Select 25% papers with the lowest scores (low achievers)-20 papers
4. 50% papers in the middle (40 papers) would not be taken in account
5. Calculate the correct responses of high achievers and low achievers on each test item
6. Apply the formula and calculate F (Facility index)

$$F=\frac{NR}{NT}\ \text{x } 100$$

NR= Number of students who got the item right
NT=Total number of students

7. F is acceptable when it rangers from 30% to 70%
8. Value more than 70% indicates that item is very easy
9. Value less than 30% indicates that item is very difficult

Example:

Total Paper = 80
High Achievers = 20
Low Achieves = 20

Table 2.Results of the High and Low achievers

	N	a	b	c	D	omit
25% High achievers	20	5	10	0	5	0

25% Low achievers	20	4	2	0	14	0
			12			

"b" is the correct answer

$$F=\frac{NR}{NT}\text{ x100}$$

$$=\frac{12}{40}\text{ x100}$$

$$=30\%$$

6. DIFFERENTIABILITY

It refers to the degree to which test item discriminates between pupils with high and low achievement. One purpose of testing is to discriminate between high and low achievers.

STEPS

1. Arrange the papers in order from the highest to the lowest scores (say 80 papers)
2. Select 25% papers with the highest scores (high achievers) 20 papers
3. Select 25% papers with the lowest scores (low achievers) 20 papers
4. 50% papers in the middle (40 papers) would not be taken in account
5. Calculate the correct responses of high achievers and low achievers on each test item
6. Apply the formula and calculate D

$$D=\frac{NH-NL}{n}$$

Where n=Number of high or low achievers
NH=Number of high achievers who got the item right
NL=Number of low achievers who got the item right

7. D is acceptable when value rangers form 1-0.30
8. Value of 1 indicates 100% discrimination
9. Value less than 0.30 indicates incapability of the item to discriminate

Example:

Total Paper=80
High Achievers=20
Low Achievers=20

Table 3. Results of students on the items of differentiability

	N	**a**	**b**	**C**	**d**	**omit**
25% High achievers	20	5	10	0	5	0
25% Low achievers	20	4	12	0	14	0

"b" is the correct answer

$$D=\frac{NH-NL}{n}\qquad=\frac{10-2}{20}$$

$$=\frac{8}{20}=0.40$$

RESEARCH METHODOLOGY

The study is descriptive in nature. The researcher used the following method to conduct the research:

Population

All the students of Intermediate Level (Part-I) of public and private Colleges of Multan city were included in the population of the study.

Sample of the Study

The researcher selected 100 students by convenient sampling method.

Tool of Research

A test of 60 MCQS test items was developed as a tool from Intermediate Part-I Punjab Text book Board prescribed textbooks I & III and the Past Question Papers of All Boards of the Punjab.

Table 4. List of the items used for data collection

Research Tool		
Sr.	**Category**	
1	Test Items	60
2	Punjab Textbook Board English Textbook-I	Intermediate Part-I
3	Punjab Textbook Board English Textbook-III	Intermediate Part-I
4	Past Papers	All Boards of Punjab

Data Collection

The data was collected from 100 male and female students of Intermediate Level (Part-I) of public and private Colleges of Multan city.

Administration of Test

The test was administered to 50 Male & 50 Female of Intermediate Part-I Class students from different public and private colleges of Multan city. They were given instructions and proper time to solve them. All the answers books were received personally by the researcher.

Data Analysis

The collected data was analyzed and tabulated in a master sheet and their D.L and D.P were calculated.

The sample table used for analysis is given below:

D.L=46% D.P=0.2

This is valid, effective and reliable item. It may be included after correction.

Table.5 Cumultaive Table For Findings

Sr.	D.L	D.P
1	46%	0.2
2	30%	0.2
3	66%	0.44
4	30%	0.2
5	62%	0.44
6	46%	0.2
7	48%	0.16
8	72%	0.56
9	48%	0.32
10	72%	0.56
11	74%	0.52
12	48%	0.32
13	72%	0.56
14	74%	0.52
15	72%	0.56
16	48%	0.32
17	48%	0.16
18	72%	0.56

19	74%	0.52
20	74%	0.44
21	46%	0.2
22	30%	0.2
23	60%	0.32
24	74%	0.44
25	48%	0.32
26	72%	0.56
27	48%	0.32
28	46%	0.2
29	62%	0.44
19	74%	0.52
20	74%	0.44
21	46%	0.2
22	30%	0.2
23	60%	0.32
24	74%	0.44
25	48%	0.32
26	72%	0.56
27	48%	0.32
28	46%	0.2
29	62%	0.44
30	30%	0.2
31	66%	0.44
32	46%	0.2
33	62%	0.44
34	48%	0.16
35	74%	0.44
36	68%	0.48
37	66%	0.68
38	66%	0.2
39	60%	0.64
40	76%	0.48
41	68%	0.48
42	54%	0.28
43	48%	0.32
44	48%	0.16
45	68%	0.24
46	46%	0.2
47	72%	0.56
48	48%	0.16
49	60%	0.4
50	74%	0.44
51	74%	0.52
52	74%	0.52
53	68%	0.48
54	66%	0.68
55	58%	0.2
56	54%	0.28
57	62%	0.2
58	74%	0.52
59	62%	0.2
60	68%	0.48

Table 6.Cumulative List of students

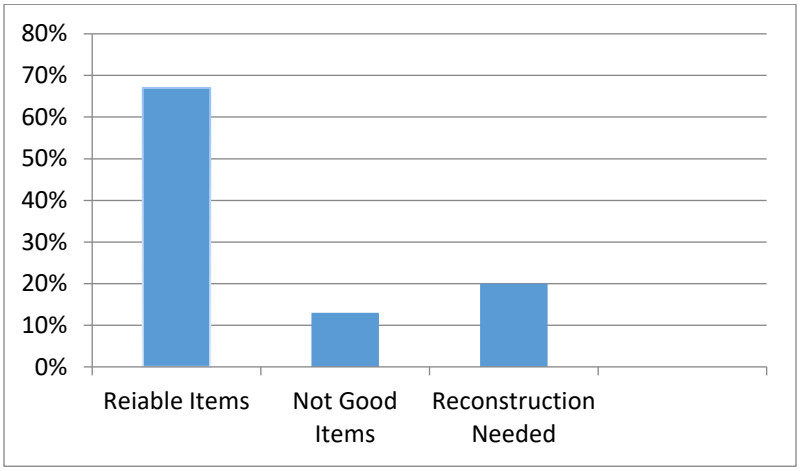
STUDENT DATA		
Sr.	**Category**	**Number**
1	High Achievers	**20**
2	Mid Achievers	**50**

3	Low Achievers	**30**
	Total	**100**

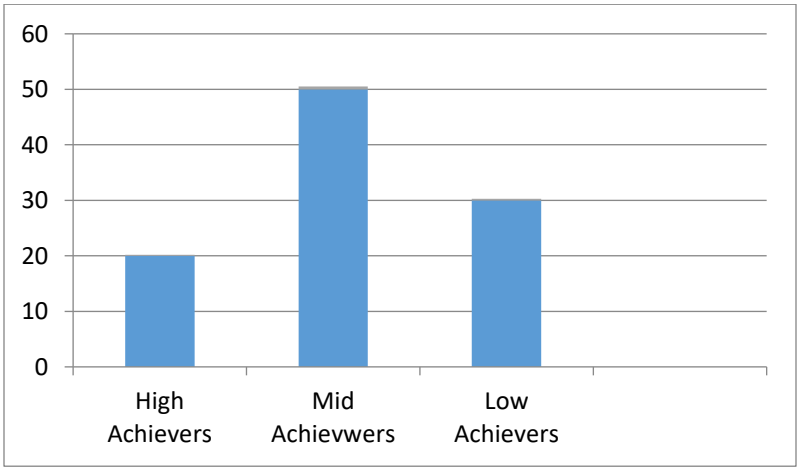
FINDINGS

The findings of the research are followings**:**

1. 67% test items are Reliable and Valid and have discrimination power to distinguish between high and low achievers.
2. 13% test items are Not Good. So not useable.
3. 20% test items Need Correction and Reconstruction. Can be used after needed necessary correction.



Test Items Result Data (Total 60 Test Items)
(67% Reliable 13% Not Good 20% Reconstruction Needed)
Figure 1.Cumulative Graphical Representation of Test Items



Students Result Data (Total 100 Students Tested)
(30 HA 50 MA 20 LA)
Figure 2. cumulative graphical representation students score achieved during Test

CONCLUSIONS

The following conclusions were acheived:

1. **40** test items of this test are effective and reliable and should be included in the test. These are: 3,5,8,9,10,11,12,13,14,15,16,18,19,20,23,24,25,26,27,29,31,33,34,35,36, 37,39,41,42,43,47,49,50,51,52,53,54,56,58,60.

2. **08** test items of this test are not good, valid and reliable and it should not be included in the test. 2,4,7,17,22,30,44,48
3. **12** test items of the test can be added to the test after correction. 1,6,21,28,32,38,40,45,46,.55,57,59.

1. Out of 100 students tested 20 were high achievers, 50 were mid achievers and 30 were low achievers.

RECOMMENDATIONS

The recommendations of this article are as follows:

4. It is recommended that objective test items in all boards papers must have discrimination power.
5. It is recommended that only Standard objective type questions must be included in BISE's papers.
6. It is recommended that Objective test items should be simple and clear. Their language must not be complicated and tricky.
7. Future researchers should work on objective type items of other subjects for all classes.
8. Future researchers should work on completion items of all subjects for all classes.

CREDIT AUTHOR STATEMENT

Amir Hafeez Malik: Conceptualization, Methodology, Software, **Rashda Perveen.** Writing- Original draft preparation. **Abdullah Shafiq**: Visualization, Investigation. **Abdul Rashid** Supervision, Validation of Results, Writing- Reviewing Editing and corresponding

COMPLIANCE WITH ETHICAL STANDARDS

It is declare that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

REFERENCES

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests* (Vol. 1). Oxford University Press.

Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language testing*, *17*(1), 1-42.

Bloor, M., Bloor, T., Forrest, R., & Relton, H. (1970). *Objective Tests in English as a Foreign Language: Pupils' Book: Prepositions, Particles and Mixed Tests*.

Cangelosi, J. S. (1990). *Designing tests for evaluating student achievement*. Longman Publishing Group.

Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics*, *20*, 147-161.

Davies, A. (1997). Demands of being professional in language testing. *Language testing*, *14*(3), 328-339.

Douglas, D. (1995). Developments in language testing. *Annual Review of Applied Linguistics*, *15*, 166-187.

Fulcher, G. (1999). Assessment in English for academic purposes: Putting content validity in its place. *Applied linguistics*, *20*(2), 221-236.

Groot, P. J. (1990). Language testing in research and education: The need for standards. *AILA Review*, *7*, 9-23.

Hughes, A. (2003). *Testing for language teachers*. Cambridge university press.

Lyman, H. B. (1998). *Test scores and what they mean*. Allyn & Bacon.

McNamara, T. F. (1996). *Measuring second language performance*. Longman Publishing Group.

Shepard, L. A. (1993). Chapter 9: Evaluating test validity. *Review of research in education*, *19*(1), 405-450.

Wesman, A. G. (1971). Writing the Test Item (w:) RL Thorndike (ed.) Educational Measurement.