ISSN : 2277 - 4262
Scientific Journal IF:5.744
INTERNATIONAL MULTIDISCIPLINARY E-JOURNAL
An International Peer Reviewed, Refereed Journal

The Development of Reading Comprehension Ability Test for Yemeni undergraduate EFL learners

Iftikhar Yusuf Al-Ariqi[1]
Ph.D Scholar
English Department, Kuvempu University

Dr. Jagannath K. Dange[2]
Assistant Professor
Education Department, Kuvempu University

Mir Mohsin[3]
Ph.D Scholar
Institute of Management Studies,Kuvempu University

Abstract

Studies have shown that the best way to test the students ability in reading comprehension is the Multiple-choice questions for its validity and reliability. The efficiency of MCQs as an efficient tool for evaluation solely rests upon their quality which is best assessed by item and test analysis. This paper tries to assess item and test quality and in order to explore the relationship between difficulty index (p-value) and discrimination indices (DI) with distractor efficiency (DE). The study was conducted among 134 second year Yemeni EFL students in Sana'a University, Yemen. Twenty MCQs administered, after checking its reliability and validity, were analysed for p-value, DI and DE. Results indicate that the mean score was 9.49 with S.D 2.82. Internal consistency reliability of the test as per KR20 was 0.7. Mean p value and DI were 61.92 ± 25.1% and 0.31 ± 0.27, respectively. DI was noted to be maximum at p value range between 40% and 60%. Combining the two indices, 19 items could be called 'good' having a p-value from 20% to 90%, as well as a DI ≥ 0.40. Overall 75% items had 2 non-functional distractors (NFDs), while 20% items had 3 functional distractors and 5% had only 1 functional distractor. Mean DE was 80.00 ± 33.00% with good and marginal levels. Excellent discrimination (DI = 33.00) was achieved with 12 items having two NFD respectively while good discrimination was achieved with only 1 item with one NFD had lower DI (33.33). Items having average difficulty and high discriminating power with functional distractors should be incorporated in future to improve the quality of the test.

Keywords: *Difficulty Index, Discrimination Index, Distractor Efficiency, Item Analysis, Multiple Choice Questions, Non-functional Distractor (NFD)*

1. Introduction:

Reading is to get information from written texts. It involves decoding words and identifying the sound that must accompany the printed word (Das, 2009). Comprehension is a part of the communication process of getting the thoughts that were in the author's mind into the reader's mind (Fry, 1963). Simply, reading comprehension is the ability to read text, process it, and understand its meaning. Das (2009) explains that reading comprehension requires four elements which are sentence parsing, words class, knowledge gained in school and past experience & culture context.

Many studies found that Yemeni Learners of English as a foreign language face a number of difficulties in reading comprehension. To investigate the reading comprehension ability of the Yemeni EFL learners, a multiple-choice question test was arranged. This paper attempts to assess item and test quality of the reading comprehension test administered to the Yemeni EFL learners in tertiary level as well as to explore the relationship between difficulty index (p-value) and discrimination indices (DI) with distractor efficiency (DE) of a multiple choice-question reading comprehension test.

2. Reading Comprehension:

To understand a theoretical basis and guidance for learning and teaching reading, it is worthy to have further understanding to the three reading models to comprehend the nature of reading.

A. Bottom-up Approach:

It assumes that a reader constructs meaning from letters, words, phrases, clauses and sentences by processing the text into phonemic units that represent lexical meaning and then builds meaning in a linear manner (Hudson, 2007). Grabe and Stoller (2002) suggest that in this model all reading follows a mechanical pattern in which the reader creates a piece-by-piece mental translation of the information in the text, with little interference from the readers' own background knowledge.

B. Top-down Approach:

It assumes that a reader approaches a text with conceptualization above the textual level already in operation and then works down to the text itself. Here, the reader does not necessarily read each word in the text as it is assumed in the bottom-up approaches. Memory capacity and mental limitations on the speed of information-processing can be working together (Hudson, 2007:33). Grabe & Stoller (2002) explain that top-down model highlights inference and interaction of all processes (lower and higher-level processes) under the control of a central processes (p.32).

C. Interactive Approach:

It is an approach which takes an interactive view of reading purpose. In this model, reading is seen as a bidirectional in nature involving the application of higher order mental processes and background knowledge as well as features of the text itself. In this model, some features are hypothesized such as 1) vocabulary knowledge and sight word recognition 2) phonetic decoding skills 3) relational knowledge and prediction form context and 4) comprehension skills (Carr, 1982, as cited in Hudson, 2007).

3. A Multiple-Choice Question (MCQ) Test:

A multiple-choice question (MCQ) comprises of a stem with a question line underneath it, followed often by a number of 3 to 5 alternatives. Cizek & Oday (1994) explains that one of the alternatives is the correct or appropriate response known as the key, while the others are described as distractors. A salient characteristic of distractors is that all options shall present credible answers and if possible none shall be incorrect (Saudi Commission for Health Specialties, 2015). Distractors are set to attract students who do not know the correct answer while students who know the correct answer are supposed to ignore them. Tests using MCQs can be used to examine student difficulties if the incorrect options are designed to reveal common misconceptions and they can provide a more comprehensive sampling of the subject material because of wider coverage. They are objective and easily adapted for computer delivery. Moreover, this type of test is often more valid and reliable than essay tests because discrimination between performance levels is easier to determine and scoring consistency is virtually guaranteed when carried out by machine (Hotiu, 2006). However, some instructors believe that MCQs are "multiple-guess" items or that MCQs are only capable of testing factual information and so are less appropriate for testing higher-order cognitive skills. But this type of test is now accepted if well-constructed multiple-choice items have been prepared in order to test many of the higher cognitive skills of Bloom"s taxonomy such as knowledge, application, analysis and synthesis. An item in a MCQs is a single test element, which might be a multiple-choice question (University of Washington, 2015). Multiple-choice question (MCQ) is an efficient tool for evaluation; however, this efficiency solely rests up on the quality of MCQ which is best assessed by item and test analysis. Item analysis is a statistical process which examines student responses to individual test items in order to identify the effectiveness of their test items and of the test as a whole. Item analysis can help in identifying potential mistakes in scoring, ambiguous items, and alternatives (distractors) that don't work. When performing item analysis, the following important statistical information are analyzed; difficulty index, discrimination index, distractors analysis & reliability.

4. Multiple Choice Questions Adapted to Assess the Reading Comprehension Ability of Yemeni EFL Learners

A. Development of the Reading Comprehension Test:

The researcher adopted a complete version of the TOEFL test. The test was taken from TOEFL Practice Tests (2015). The test contained two passages with multiple-choice questions. The first passage was a bit longer followed by 17 questions. The second passage was shorter followed by 14 questions, too. Every question aimed to test one of the students' skills in reading comprehension. To answer each question, the students had to choose one option among four alternatives. The scoring of the reading comprehension test was done in the following way. One mark was awarded for each correctly chosen answer and zero for the wrong answer.

B. Preliminary form of the test:

The entire test was arranged and sent to 5 educational and English professors who are experts in the field. The juries were requested to emend and refine each question along with its multiple-choice answer from different perspectives such as the right grammatical correctness, structure of the statements, distractors, appropriateness for students. By considering the suggestions made by those experts, the two passages were agreed to be considered with a number of 13 questions for the first passage and 12 for the second. (Two) items in the two selected passages were dropped out as they are writing-questions which are not needed to test the students' reading comprehension. Thus, the two passages finally comprise only 23 questions as a total were retained and considered for pilot paper.

C. Try out the test:

The pilot paper was carried on a number of 42 students to make sure of the consistency among the items to all students as well as the exact required time to finish answering all items of the test.

Instructions were given twice to them to make it clear. After inserting pilot study data in SPSS software program, reliability found to be 0.88 which means the items are consistent. However, it has found that students hardly did the test in the allocated time in which they require more than two hours.

D. Modification of the test:

It was noticed that the students could not do the whole test due to the test length. They hardly had to do with two passages only in one hour and a half. So, the questions in the two passages were agreed to be eliminated to 10 questions for each taking into account the rigor system of the test a well as considering the limited time.

5. Settings & test scoring

The study was conducted in the department of English Department of Sana'a University for second year EFL students 2015-2016. The test comprising twenty MCQs was administered to 134 EFL students. The time allocated was one hour and a half (90 minutes). After each passage, the items comprised of single stem and four answer options, having a single stem and four answer options, one of them being correct and the other three being 'distractors'. The students were required to darken the correct choice. Each correct response was awarded 1 mark and each incorrect response was awarded 0, range of total score being 0 to 20.

6. Statistical Analysis

Scores of 134 students were entered in order of merit in MS Excel and simple proportions, mean, standard deviations were calculated. Items were categorized according to their difficulty index (p-value), discrimination index (DI) and distractor efficiency (DE) and actions such as discard/ review /revise and store were proposed. Reliability of the test was checked using Kuder-Richardson 20 coefficient (KR20).

a. Item Analysis

To assess the MCQs and test its quality, the difficulty and discrimination indices are among the tools which are used for this purpose. Another tool used for further analysis is the distractor efficiency which analyses the quality of distractors and is nearly associated with difficulty and discrimination indices. Reasons for negative DI can be wrong key, ambiguous framing of questions or generalized poor preparation of students. Items with negative DI decrease the validity of the test and should be removed from the collection of questions. Difficulty index and discrimination index are often inversely related except for extreme situations where the difficulty index is either too high or too low. It has been seen that the relationship between them is not linear, but predicted as dome shaped (Karelia, Pillai, Vegada,2013).

Analyzing the distractors is done to determine their relative usefulness in each item. If students consistently fail to select certain multiple choice alternatives, it may be that the options are probably totally implausible and therefore of little use as traps in multiple choice items. Therefore, designing of plausible distractors and reducing the NFDs is important aspect for framing quality MCQs (Haladyna & Downing, 1989).

The idea behind using three techniques to carry out item analysis is to assess the performance of students with greatest precision and develop a test paper that serves the international standards and quality. It should be kept in mind if there is a conflict among the three techniques of item analysis, the preference should be given to the discrimination index due to its effectiveness in discriminating on the basis of performance among the students. This inculcates the meritocracy among the students or academicians rather than bias and favoritism.

Thus, item wise analysis was conducted using the following procedures.

b. Difficultly Index:

Difficulty index (**p**) is expressed as the proportion of the students who answer the items correctly. The term actually is a misnomer as it should have been easiness index. The formula for computing difficulty index is given below.

$$Difficulty\ \text{Index} = \frac{\text{Students with correct answers}}{\text{Total Number of Students}} \times 100$$

The p-value statistics ranges from 0-1 or 0-100%. The higher p-value, the easier is the question. As a general convention, items with p-value between 20-90% are considered good and acceptable. Whereas the p-value between 40-60 are considered excellent and items with p-value less than 20% are considered difficult. Finally, the items with p-value more than 90% are considered easy and might need modification or elimination.

c. Discrimination Index (DI):

It is the major of effectiveness of an item in discriminating between high and low scores. As a general convention, in order to compute discrimination index, the test takers are divided into two group; high and low. The formula for calculating discrimination index is given below.

$$\text{Discrimination Index} = \frac{\text{H - L}}{27\%\ of\ Total\ \text{Students}} \times 100$$

H= Number of correct answers from top 27% of the students

L= Number of correct answers from bottom 27% of the students

The values of item discrimination index ranges from -1 to +1. The higher the value of DI, the more effective the item is. When DI is 1, all test takers in the upper group and no test takers in the lower group answered the item correctly. Conversely, if none of the high group but all of the low group answered an item correctly; the DI value would be -1.00. In general, the DI value (0.40) and greater are considered excellent items; Items with DI (between 0.30 to 0.39) is considered reasonably good but possibly subject to improvement; those with DI (0.20 to 0.29) are considered marginal items and should be reviewed while those with DI (below 0.19) are considered poor items and should be eliminated.

d. Distractor Analysis:

Distractor analysis is a statistical technique used to check the quality of each option in a multiple choice question and describes the degree of attraction of examinees or test-takers towards each option. Distractor analysis can be a useful tool in evaluating the effectiveness of these distractors in which they reveal if students do guessing and not really know the right answer. There is a greater possibility that students will be able to select the correct answer by guessing as the options have been reduced. There are two types of distractors; non-functional and functional. Non-functional distractors (NFDs) are options that are selected infrequently (<5%) by examinees and functional or effective distractor is the option selected by 5% or more students. As such, NFDs should be revised, removed or be replaced with a more plausible option.

Distractor efficiency (DE) is determined for each item on the basis of the number of NFDs in it and ranges from 0 to 100%. If an item contains three or two or one or nil NFDs, then DE will be 0, 33.3, 66.6, and 100%, respectively. The formula for calculating distractors efficiency is given below.

$$Distractor\,Efficiency(DE) = \frac{Total\ \text{number of distractors}\ (TD) - \text{Number of nonfunctional Distractors}\ (NFD)}{Total\ \text{number of distractors}\ (DT)} \times 100$$

$$DE = \frac{TD - NFD}{TD} \times 100$$

Non-Functional Distractors NFD= options which are selected infrequently (<5%)by students
FD= options which are selected by 5% of the students
TD= Total number of distractors

e. Test of Reliability:

In order to check the internal consistency of the given test, Kuder-Richardson 20 coefficient was applied. The formula for computation of this test is given below;

$$r = \frac{k}{k-1}\left[1 - \frac{\sum_{i=1}^{k} pq}{\sigma^2 x}\right]$$

where, *pi* is the proportion of correct responses to test item i, qi is the proportion of incorrect responses to test item i (so that pi + qi = 1), and the variance for the denominator is

$$\sigma_x{}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

where, n is the total sample size, Xi is the score of individual students and X is the mean total score. The value of KR20 can range from 0 to 1, with numbers closer to 1 reflecting greater internal consistency indicating that the items are all measuring the same thing or general construct. The widely-accepted cut-off value of KR is greater than or equal to 0.7.

7. Discussion & Results:

To access and analyze the reading comprehension ability of Yemeni EFL learners, difficulty index, discrimination index and distractors efficiency statistics were conducted and results show the following;

Total 20 MCQs and 60 distractors were analyzed. The scores of the 134 students ranged from 2 to 19 marks out of total 20 marks. The mean score achieved is 9.49 with S.D is 5.03. The mean score according to the groups i.e. top 27% (36) is 16.08 with S.D 2.82 and bottom 27% (36) is 3.33 with S.D 1.96 respectively. Means and Standard deviations for difficulty index, discrimination index and distractor efficiency were (M=47.43, S.D=16.96),(M=0.61, S.D=0.46) and(48.33,27.51) respectively. The results are shown in following table1.

Table (1):

Parameter	**Mean**	**SD**
Difficulty index	47.43	16.96
Discrimination index	0.61	0.46
Distractor efficiency	48.33	27.51

The descriptive statistics of distractor efficiency is provided in the following tables:

Table (2): Distribution of selection of various options by examinees in high & low groups in corresponding items

Item No.	Option A	Option B	Option C	Option D
Item1	1	11	59	1
Item2	36	33	2	1
Item3	32	2	36	2
Item4	1	36	2	33
Item5	36	1	33	2
Item6	33	36	1	2
Item7	33	2	1	36
Item8	36	1	1	34
Item9	34	36	1	1
Item10	34	1	1	36
Item11	35	2	33	2
Item12	8	26	32	6

Item13	16	20	34	2
Item14	10	7	48	7
Item15	2	1	33	36
Item16	33	2	36	1
Item17	35	2	33	2
Item18	36	25	5	6
Item19	20	7	36	9
Item20	40	1	29	2
Total	**511**	**252**	**456**	**221**
Average	**25.55**	**12.6**	**22.8**	**11.05**

It is evident from the above table that majority of the students from both groups (top & bottom) selected option (A) average 25.55. The average selection of options A,B,C and D by the Yemeni EFL learners is 25.55, 12.6, 22.8 and 11.05 respectively. The proportionate usage of all the options is provided in the below table.

Table (3): Proportionate distribution of selection of various options
by high & low examinees in corresponding items

Option	**A**	**B**	**C**	**D**
Item1	0.01	0.15	0.82	0.01
Item2	0.50	0.46	0.03	0.01
Item3	0.44	0.03	0.50	0.03
Item4	0.01	0.50	0.03	0.46
Item5	0.50	0.01	0.46	0.03
Item6	0.46	0.50	0.01	0.03
Item7	0.46	0.03	0.01	0.50
Item8	0.50	0.01	0.01	0.47
Item9	0.47	0.50	0.01	0.01
Item10	0.47	0.01	0.01	0.50
Item11	0.49	0.03	0.46	0.03
Item12	0.11	0.36	0.44	0.08
Item13	0.22	0.28	0.47	0.03
Item14	0.14	0.10	0.67	0.10
Item15	0.03	0.01	0.46	0.50
Item16	0.46	0.03	0.50	0.01
Item17	0.49	0.03	0.46	0.03
Item18	0.50	0.35	0.07	0.08
Item19	0.28	0.10	0.50	0.13
Item20	0.56	0.01	0.40	0.03

It is evident from table (3.16) that maximum percentage of Yemeni EFL learners selected in items 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20 options C,A,C,B,A,B,D,A,B,D,A,C,C,C,D,C,A,A respectively.

Table (4):
The distribution of difficulty and discrimination indices of the items and their corresponding Distractor Efficiency

Items	Difficulty Index	Discrimination Index	Distracter Efficiency
Item1	90.30	0.36	0.33
Item2	70.15	0.95	0.33
Item3	50.75	0.96	0.33
Item4	40.30	0.91	0.33
Item5	40.30	0.92	0.33
Item6	40.30	0.92	0.33
Item7	40.30	0.94	0.33
Item8	40.30	0.93	0.33
Item9	40.30	0.95	0.33
Item10	50.75	0.94	0.33
Item11	35.07	0.92	0.33
Item12	37.31	0.28	1.00
Item13	37.31	0.28	0.67
Item14	**68.66**	**-0.67**	**1.00**
Item15	**84.33**	**-0.08**	**0.33**
Item16	35.07	0.92	0.33
Item17	35.07	0.92	0.33
Item18	37.31	0.28	1.00
Item19	37.31	0.28	1.00
Item20	37.31	0.28	0.33

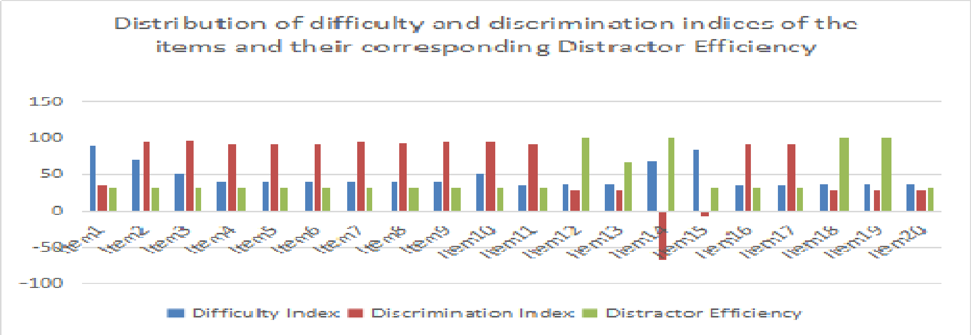

Figure 1: Diagrammatic representation of the distribution of difficulty and discrimination indices of the items and their corresponding Distractor Efficiency

The descriptive statistics of using all the three test statistics are represented in Table 4:

Table (5): Descriptive statistics using all the three test statistics

Difficulty Index	**Interpretation**	**No.of Items(%)**	**Distractor Efficiency**	**Proposed Action**
20-90	Good	19	48.89	Store
<20	Too Difficult	0	0	-
>90	Too easy	1	33.33	Store/review
Discrimination Index				
>=0.40	Excellent	12	33.00	Store
0.30-0.39	Good	1	33.33	Store/Review
0.20-0.29	Marginal	5	80.0	Store
<=0.19	Poor	2	66.66	Discard

The table reveals that out of total 20 items, 19 have acceptable level of difficulty with p-value within the range of 20% to 90% whereas one item among them had p value >90%. Distractor efficiency values corresponding to difficulty index are 48.89, 0 and 33.33 respectively. The Difficulty Index was noted to be maximum at p value range between 40% and 60%. Combining the two indices, 19 items could be called 'good' having a p-value from 20% to 90%, as well as a DI ≥ 0.40. Overall 75% items had 2 non-functional distractors (NFDs), while 20% items had 3 functional distractors and 5% had only 1 functional distractor. Mean DE was 80.00 ± 33.00%. Excellent discrimination (DI = 33.00) was achieved with 12 items having two NFD respectively while good discrimination was achieved with only 1 item with one NFD had lower DI (33.33). Two items with p-value ranged <0.19 have 1 FD and another with 2 NFD are found to be poor because they were answered correctly by the low group while they were not by the high group.

Similarly, majority of items (12) have excellent discrimination indexes (DI≥0.40), one item has good discrimination indexes (DI between 0.30 to 0.39), 5 items have marginal discrimination indexes and 2 items having poor DI respectively. Likewise, the Distractor efficiency corresponding to this discrimination index is 33.00%, 33.33%, 80.00% and 66.66% respectively. Combining all the three item analysis statistics, it can be inferred that item no. 14 & 15 should be discarded.

8. Conclusion:

The study concludes after assessing all items in the given question paper using three test statistics namely difficulty index, discrimination index and distractor efficiency that out of 20 questions, 18 are reliable whereas two items are statistically unreliable and can be discarded. The same has been also confirmed by the Kuder-Richardson 20 coefficient. However findings of the present study have to be interpreted cautiously in the light of certain limitations; the number of items in this test was less and other semester students were not included. Future studies with larger number of items having average difficulty and high discrimination with functioning distractors administered to a bigger sample will add to the findings of this study.

References:

1. Cizek GJ, O'Day DM. (1994). Further investigations of nonfunctioning options in multiple-choice test items. EducPsycholMeas; 54(4):861-72.
2. Das, J.P. (2009). Reading Difficulties and Dyslexia: An Interpretation for Teachers. SAGE Publications: India Pvt. Ltd.
3. Designing and managing MCQs (2015).: Appendix C: MCQs and Blooms taxonomy. [Internet] [cited 2015 June 15].Available from:http://www.u.arizona.edu/~jag/POL602/Designing-Managing-MCQs.pdf
4. Educational Development Center (2016). Item Analysis. Scantron Guides. Available from https://carleton.ca/edc/wp-content/uploads/Item-Analysis.pdf
5. Fry, Edward. (1963). Teaching Faster Reading: Cambridge University Press, Cambridge. Print. Page 24.
6. Grabe, William & Stoller, Fredricka (2002). Teaching and Researching Reading. England. Pearson Educational Limited.
7. Hudson, Thom. (2007). Teaching Second Language Reading. Oxford. Oxford University Press.
8. Haladyna TM (1989). Downing SM: Validity of taxonomy of multiple choice item-writing rules. ApplMeasEduc ; 2(1):51-78
9. Hotiu A. (2006).The relationship between item difficulty and discrimination indices in multiple-choice tests in a Physical science course [MSc thesis]. Boca Raton, Florida: Florida Atlantic University; [cited 2015 June 14]. Available from:http://www.physics.fau.edu/research/education/A.Hotiu_thesis.pdf
10. Karelia BN, Pillai A, Vegada BN (2013). The levels of difficulty and discrimination indices and relationship between them in four response type multiple choice questions of pharmacology summative tests of year II MBBS students. IeJSME ; 7(2):41-6. [20].
11. Mukherjee, Poulomi & Saibendu, Lahiri. (2015). Analysis of Multiple Choice Questions (MCQs): Item and Test Statistics from an assessment in a medical college of Kolkata, West Bengal. IOSR Journal of Dental and Medical Sciences (IOSR-JDMS), e-ISSN: 2279-0853, p-ISSN: 2279-0861.Volume 14, Issue 12 Ver. VI (Dec. 2015), PP 47-52 www.iosrjournals.org
12. Saudi Commission for Health Specialties, (2015). Item writing manual for multiple-choice questions. [Internet] [cited 2015 June 12]. Available from:http://www.scfhs.org.sa/education/HighEduExams/Guidlines/mcq/Documents/MCQ%20 Manual.pdf .
13. University of Washington. (2015). Understanding item analysis reports. [Internet] [cited 2015 July 4]. Available from: http://www.washington.edu/oea/services/scanning_scoring/scoring/item_analysis.html