Distractor Efficiency of an Intermediate English Proficiency Course Final Examination Paper

Kamarul Ariffin Ahmad[1], Nora Liza Abdul Kadir[2], Mohd Zahren Zakaria[2], Muhamad Lothfi Zamri[3], Siti Najihah Abd Rahman[2], Nurhafza Mohamad Mustafa[2]
[1,2]Centre for Languages and General Studies, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak.
[3]Department of English Language and Literature, Faculty of Languages and Communication, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak.
kamarul.a@pbmpu.upsi.edu.my[1]

Received: 1 November 2021; **Accepted:** 10 December 2021; **Published:** 21 December 2021

To cite this article (APA): Ahmad, K. A., Abdul Kadir, N. L., Zakaria, M. Z., Zamri, M. L., Abd Rahman, S. N., & Mohamad Mustafa, N. (2021). Distractor Efficiency of an Intermediate English Proficiency Course Final Examination Paper. *Asian Journal of Assessment in Teaching and Learning*, *11*(2), 103-112. https://doi.org/10.37134/ajatel.vol11.2.10.2021

To link to this article: https://doi.org/10.37134/ajatel.vol11.2.10.2021

Abstract

Language proficiency tests have adopted various types of items which include multiple-choice questions (MCQ). Designing MCQ needs a longer time to be completed despite being perceived as easy to administer and mark. After the administration and marking process are completed, the test needs to be analysed to determine the index of facility, discrimination power and distractor efficiency. This study aimed to scrutinise those indices and determine if there is an association between the indices with the functionality of the distractors in an intermediate English Proficiency course final examination paper. With the use of QUEST software, the data were computed to get the difficulty and discrimination indices, and distractor efficiency.The finding showed that there is no clear association between facility and discrimination indices with the distractor efficiency. This study found that certain items had acceptable indices despite having poor or non-functional distractor and certain items had poor indices even though the distractors were functioning as they were intended to. This study concluded that distractor efficiency may have a trivial association to the item facility and discrimination power. Therefore, it is suggested for a deeper investigation to be conducted in terms of the language used to construct the stem and the options.

Keywords: Distracter, Efficiency, Functional, Facility and Discrimination

INTRODUCTION

English Proficiency 3 is an upper-intermediate course which develops students' ability to construct compound sentence structures with the use of relevant phrases and clauses. Students should be able to express agreements and disagreements with good justification in listening and speaking activities. Students could infer, draw conclusions and predict outcomes from reading texts. The grammar in focus assists students in writing reports, narratives and cause-and-effect essays. The final examination paper for this course consists of 40 multiple-choice questions with the first 20 questions assess certain skills in reading, such as the ability to comprehend main information in the text, scan detailed information, understand the meaning of words and make inferences. Another 20 questions assess students' ability in grammar mastery.

Despite the growing trend in alternative assessment in line with the latest development in learning and language teaching of the 21st century, the use of multiple-choice questions is still popular

in tertiary level due to its advantage in administering examination for large number of students. Although multiple-choice questions are commonly used at university level and other levels of language instruction in Malaysia, there has not been enough evidence about index of facility, discrimination power and distractor efficiency in the literature. Additionally, such analyses are needed to facilitate the development of new items in the future as well as to serve an empirical evidence to be kept for item banking and for future researches. Therefore, this study attempts to fill this gap by answering the following research questions:

i. To calibrate the item facility index and discrimination power index of the items
ii. To determine the association of the indices with distracters efficiency.

Achievement test

The English Proficiency test that was analysed in this study was considered as an achievement test. The outcome or result of the achievement test is useful in determining the output of the hard word invested in improving the education system and make it clear on the next pathway to take in educating students (Bichi et al., 2016). Bichi et al. (2016) stated that sharing the students' achievements, the enhancement in the instructional plan, motivating students and student selection are among the purposes of an achievement test. Achievement tests are useful in understanding how much the students have mastered the learning goals of the course they have learned. Therefore, the instrument of the test is deemed useful for the improvements of the course thus ensuring its quality would be vital.

Multiple Choice Questions (MCQ)

Three reasons were stated by the Central Queensland University, Australia, that cause the possibility of students guessing as high as 50%, 1) the actual experience in the working world is not rightly represented in the MCQ tests, 2) the possibility of poor distractors to lead students to selecting the wrong answer and 3) the students may not clearly understand the MCQ items, that lead them to take a drastic step in not allowing MCQ test to be used to assess students (Ibbett & Wheldon, 2016). From the literature review done, Ibbett and Wheldon (2016) were able to identify ten MCQ errors that could cause 'positive clueing signals to students', which are; 'all of the above, none of the above, and complex format', 'specific determiners', 'grammatically incorrect stem and distracters', 'similarity of wording in stem and distracters', 'inconsistency in length of distracters', 'pairs of options', 'implausible distracters', 'unfortunate coincidence', 'numerical order' and 'option bias'. When a tester plans to use the MCQ in his test, the purpose of the test will affect the development of the questions in a few aspects; stem formation, the options of answers, as well as the distractors (Iwintolu and Afolabi, 2018). According to Wilson and Masters (1993), as cited by Iwintolu and Afolabi (2018), one of the drawbacks of MCQ test papers is where the testees are limited to finding the only correct answer, and this will put the low-performing students at a disadvantage because they were not awarded marks for their efforts to find the answer. One of the contributing factors of poor quality of MCQ test paper is when the teachers or testers have little or no knowledge in developing MCQ; even the culture of using MCQ is not that prevalent in the institutions, as illustrated by Karkal and Kundapur (2016). Karkal and Kundapur (2016) mentioned a situation discovered in a study by Tarrant and Ware (2008) where students who display good performance in general will be negatively affected by poorly made MCQ items as compared to weaker students. According to Iqbal, Irum and Yousaf (2017) in medical colleges, MCQ is one of the most frequently used method to test students of their knowledge. It is strongly held belief that any questions from the low order thinking skill (LOT) and high order thinking skill (HOT) levels in Bloom's Taxonomy can be produced using MCQs (Iqbal, Irum and Yousaf. 2017). An excellent MCQ should have very good distractors, which for each distractor; the answer can be thought to be the correct one.

This study deployed Rasch analysis where facility index, discrimination index and distractor efficiency will be determined with QUEST software. Rasch analysis with QUEST provides useful information to help researchers or item developers in performing item analysis. Rasch analysis, not only providing the indices stated above, will compute the data into logits where the items and the testees could be compared directly on an interval scale.

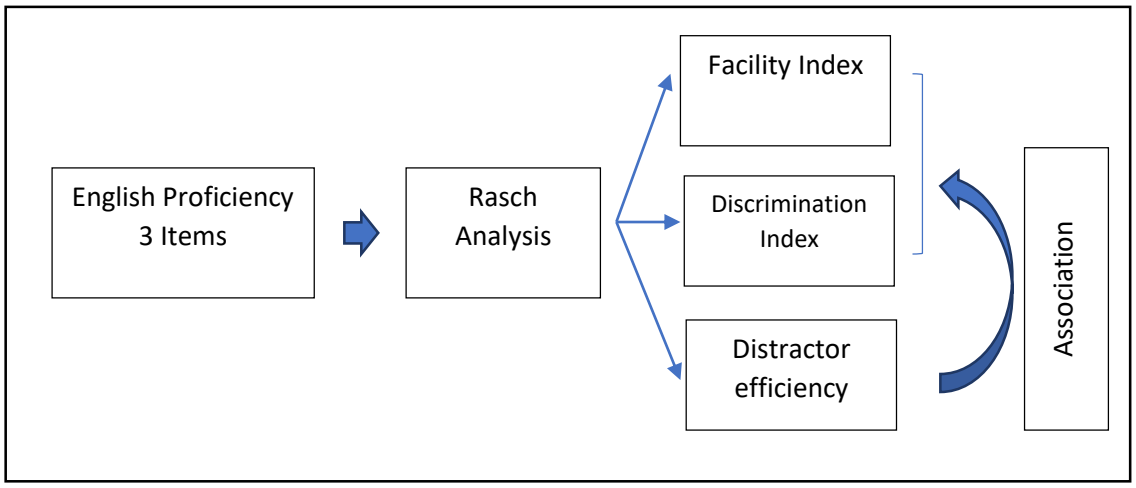

Figure 1. Conceptual Framework

The following subtopics will discuss in details of the indices that would be the indicators of the variables. According to Rao et al (2016), a good MCQ is ought to have good difficulty and discrimination indices with three functioning distractors; in case of four-option MCQ. On the other hand, Puthiaparampil and Rahman (2021), added that 50% of distractor efficiency would be optimum to ensure the quality. The baseline of these ideas is that an MCQ must consist of functioning distractors in order to retain its quality and to be kept for future use.

Item facility index

This index is crucial to help test developers to arrange the items according to the difficulty from the easiest to the most difficult (Crisp & Palmer, 2007). They further iterated that the use of these indices is not to eliminate items from the test but to understand how each testee responds to the item. This may not be appropriate and not applicable in this test since this test is divided into two sections where both sections are testing different knowledge and ability.

Nevertheless, according to the test specification table, the distribution of the items according to Bloom taxonomy shows that majority of items fell on average level of the taxonomy. Generally, the facility should fall between the average and according to Jonm in MSDLT University of Oxford (2012) the average index of facility should fall slightly higher than the passing marks in order to ensure the items would discriminate the testees well.

Item discrimination index

From their research on the Northwest University (NWU) Post-UTME Economics test items, (Bichi et al., 2016), they found that nearly three quarter of the items (71%) have low or insignificant discriminating ability, and therefore these items have to be rechecked or discarded, but 29% of the items are moderate or significant in differentiating students. Discrimination index helps teachers to identify items that can tell apart between students of different abilities; hence, it means a test paper which items have high discrimination index would indicate that the students are displaying a wide difference in the levels of performance. Different performing students with differing abilities can be identified by the discrimination index (DI), and the students are normally placed within the range of 1 and 0, but there are instances where poor performing students managed to get the answer correct compared to better achieving students, and here the DI is less than 0 (Kaur, Singla and Mahajan, 2016). Koçdar, Karadağ and Sahin (2016) described the DI as the index number of test items that are able to categorise students who were given the opportunity to display their performance, and the further is the index from 1 meaning the item is not able to differentiate the students' level. Koçdar, Karadağ and Sahin (2016), citing a research by Kim et al. (2012), mentioned that the difficulty indices of items testing the HOTs levels are lower than those testing LOTs, and this is found in the MCQ test paper in pharmacy studies, but on the other hand, the discrimination indices for 2 of the HOTs level are more significant compared

to that of the LOTs for the same test papers. The research to see the connections between cognitive levels and the index for difficulty and DI in examination papers are not many, said Koçdar, Karadağ and Sahin (2016), but they stated that different subject and contexts of the examination papers would have differing relationship between the indices and the cognitive levels, as what has been shown by Momsen et al. (2013). Büyükturan and Şireci (2018) defined item DI as "the power of the item to distinguish between individuals with or without the tested qualities, or in other words, the individuals who have or have not acquired the desired quality". They also stated that the test scores of the test takers are used as the criterion to categorize students into different groups such as by the accuracy of the answers given, the quality they want the students to have and the low and high performing students. According to Büyükturan and Şireci (2018), the research done by Crocker and Algina (1986), Baykul (2000) and Kilmen (2014) disagree to the practice in which testers would refer to the students' performance: the low and high performing scores, and use it to get their item DI, because they said that the "significant part of the group" of test takers. But other methods are mentioned as well by Büyükturan and Şireci (2018), which are "methods based on the correlation between item score and test score". The validity of the test items can be determined by using the item DI, and it is done by observing if the item discrimination is consistent with an external (or outside of the items) criterion., and this has been explained by Demir's (2011) and Kızılkaya ve Aşkar (2009), when they did the scale development studies (Büyükturan and Şireci, 2018). Hassan and Hod (2017) gave a straightforward explanation for DI, where they stated that item DI is the difference between the total number of high performing students and the total number of low performing students who got the correct answer. Echoing what Kaur, Singla and Mahajan (2016) has stated about the value of the DI, Hassan and Hod (2017) mentioned that a good index should be between 0 and 1, which is always sought for, however, there would be instances where the index value is in the negative.

Development of distractor

There are several characteristics of a good distractor in an MCQ test, as listed by Boland et al. (2010), in which among them are where distractors should not be related in any way, straight to the point, does not have the same content as with other distractors, distractors are arranged in an acceptable manner and not much difference in length, grammar and content. Jonick, Schneider and Boyland (2017) have illustrated this from the question paper they selected, where the accounting question asked only about the policy and the distractors contain statements of facts and the answers is a sentence from the policy. Another sample cited in the research done by Jonick, Schneider and Boyland (2017), in which the study done on Certified Public Accountant (CPA) exam paper, shows how poor distractors can lead students to answer wrongly, which in turn will not be usable in evaluating students learning. Koçdar, Karadağ and Sahin (2016) held the belief that one of the possible reasons of items testing LOTs level performed well is because low performing students could guess the answers due to the fact that the distractors used in the items are weak. A good functioning, acceptable distractor, which would be somewhat correct and related to the item, is one that would be chosen as their answer by more than 5% of the test takers, and instead, a poorly made distractor would be chosen by less than 5% of the test takers, and this is a normal practice done when using Single Best Answer (SBA) items (Hassan and Hod, 2017). Mark J. Gierl, Okan Bulut, Qi Guo, and Xinxin Zhang (2017) recommended these in developing distractors after reviewing literature particularly by Frey et al., 2005; Haladyna & Downing, 1989; Haladyna et al., 2002; Moreno et al., 2006, 2015:

(a) use plausible distractors in multiple-choice items,
(b) place distractors in logical order,
(c) keep the content within the distractors independent of one another,
(d) none-of-the-above and all-of-the-above should be used carefully,
(e) avoid providing inadvertent clues to the correct option in the distractors,
(f) incorporate common errors of students in distractors,
(g) keep distractors homogeneous in content and grammatical structure, and
(h) phrase distractors positively

In addition, Mark J. Gierl, Okan Bulut, Qi Guo, and Xinxin Zhang (2017) also mentioned that two distractors are more effective than three or more as it is optimal for the students to answer all questions within the stipulated time. In term of how distractors should be positioned, researchers have different views including randomize ordering to increase assessment security and reduce the probability of dishonesty (McLeod et al., 2003; Mosier & Price, 1945; Schroeder et al., 2012). Distractors can also be arranged in ascending or descending order but this is mostly applicable in content areas like mathematics (Haladyna et al., 2002; Huntley & Welch, 1993). The quality of the items must be reviewed once the test has been administered in order to assess the effectiveness of the distracters. The common way of doing this is by identifying "low-frequency distractors" (Haladyna & Downing, 1993). These items are either revised or removed by the content expert. In traditional method, every distractor is constructed by the content specialist and it can be daunting. Two alternative approaches to overcome this challenge are recommended in the article. High quality distractors can be developed in large volumes using key features and content similarity.

MATERIALS AND METHODS

The material analysed in this research is the final examination questions used for English Proficiency 3 (EP3) students. The question paper consists of 40 MCQs where they range from C2 until C4 in the cognitive scale of Bloom's Taxonomy. Based on the Test Specification Table, eight questions fall under C2 scale, 26 questions for C3 and six questions for C4 and one mark is awarded for each correct answer. The paper comprises of two sections, A and B, where the former encompasses of two reading passages, followed by 20 MCQs while the latter contains two cloze passages, also tested with 20 MCQs.

The materials used in the paper are adapted from local tourism website, News Straits Time (NST) online as well as Wikipedia where it featured an article on Malaysian delicacies. This is to ensure that the passages used are local-based and are relatable to the students. In Section A, students are tested on their understanding of the passages on both explicit and implicit ideas. Certain questions also tested students' vocabulary level by asking the antonyms and synonyms of selected words. The cloze passages in Section B mainly focus on testing the vocabulary and grammar aspects such as tenses, prepositions, determiners, pronouns, linkers and Wh-questions. The question paper is analysed to determine the (1) index of facility, (2) discrimination power and (3) distractor efficiency.

Threshold

The software QUEST is used in this research to find out the positions of test takers and item difficulty index. The data are then compared to predict the difficulty level and to determine the number of test takers with average ability ($\theta = 0$) in answering certain items correctly. A test with good quality will have a balanced of easy-difficult item and the range of threshold; the logit which represents the difficulty index of the item, must be within =2 and -2 (Lord and Novick, 1970; as cited in Bond and Fox, 2007).

Discrimination power

A test is considered good when it has a good discrimination index; in between .2 and .8. According to Singh (2009), when a test has a negative logit for discrimination, it is regarded as a test with low discrimination index. This signifies that students from low achiever group are also able to answer to questions correctly. Among other reasons, poor distractors and guessing may be some of the possible cause (Kamarul, 2018).

Point-biserial

According to Kamarul (2018), point-biserial must be positive for the key and negative for the distractors as positive logits signifies weak distractor. When the key shows negative logit, it indicates that even weak students are able to choose the correct answers. This could also mean that there could be more than one acceptable answers in the item.

RESULTS AND DISCUSSION

Multiple choice items which undergone item analysis will help the test developers to identify the standard quality of the items and the procedure is helpful in many ways. Therefore, the procedure provides the test developers with statistical value in which not only useful for understanding the behaviour of the items and the test takers, but also in revising, storing and discarding poor items. Some researchers (Si-Mui Sim & Raja Isaiah Rasiah, 2006) found that items with good discrimination power would have average index of facility and few others (Kaur, Singla, Mahajan, 2016) would find that there is no significant difference when it comes to the relationship between both indices.

The results of this study show that there are several items in the test that need attention with either having poor discrimination power (DP), unacceptable facility index (FI) or low-efficiency distractors; in which these may pose some threats to the reliability and the validity too. The portion of the items according to the DP and FI is represented in the pie charts below.

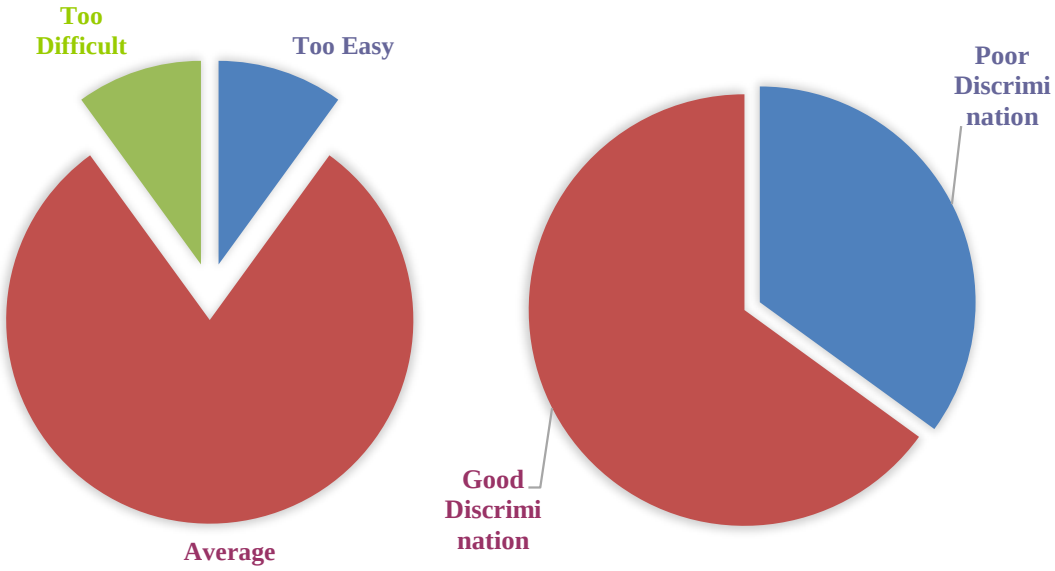

Figure 2. Distribution of Item Facility and Discrimination Power of the Test Paper

From the pie charts, it could be seen that the instrument consisted of majority average difficulty items and 14 items with poor discrimination power.

Table 1. Facility Index and Discrimination Power of Items in the English Proficiency Test

No	Descriptions	Item No.	No. of Items
1	Poor Discrimination Power	1, 2, 3, 5, 17, 18, 19, 22, 24, 27, 28, 32, 37, 40	14
2	Unacceptable Facility Index	2, 3, 5, 18, 19, 22, 24, 27	8

This study found that the items which have unacceptable facility indices are all having poor discrimination power. This supports the findings of Si-Mui Sim and Raja Isaiah Rasiah (2006) where items which are either too difficult or too easy will have poor discrimination power. Six other items which are Item 1, 17, 28, 32, 37 and 40 are having poor discrimination power despite having average facility indices.

Table 2. Items with Problematic Distracters

No	Descriptions	Item No.	No. of Items
1	Items with non-functional distracter(s)	1, 2, 3, 5, 10, 23, 28,	7
2	Items with low-efficiency distracter(s)	13, 16, 17, 18, 19, 22, 24, 27, 32, 37, 40	11
Total			18

Based on Table 2, this study found that there are four items (Item 10, 13, 16 and 23) which have good discrimination power and average facility index even though they are having non-functional or low-efficiency distractors. This is consistent with the result found in Toksoz and Ertunc (2017) where they found few items with good discrimination power will still have low-efficiency distractors. Items with acceptable indices yet having non-functioning or low-efficiency distractor need to be revised and studied. The non-functioning distractors, which are also referred to implausible distractors, are common and could be found in many tests and the frequency would defer from one to another (Zafar Iqbal, Irum & Sohaib Yousaf, 2017) thus making item revision a crucial step in developing a test with high validity and reliability.

Table 3. Item Analysis for Item 1

Discrimination Index				.15
Facility Index				-.94
Options	A*	B	C	D
Count	100	0	22	10
Percentage	75.8	0	16.7	7.6
Pt. Biserial	.15	*Error*	-.18	.00

*Correct answer

Based on Table 3, Item 1 has shown considerable errors in item development. Despite having poor discrimination power (*.15*), one of the distractors (distractor B) is not functioning and none of the students chose that option. Nevertheless, this item is considered as quite easy (*-.94*). As Battisti, Hanegan, Sudweeks and Cates (2010) iterated in their study, distractors should function to demonstrate the "different level of understanding" (p.84) among test takers, distracter B in Item 1 definitely failed to serve that purpose therefore it should be discarded and replaced.

Table 4. Item Analysis for Item 5

Discrimination Index				.17
Facility Index				-3.30
Options	A*	B	C	D
Count	128	0	4	0
Percentage	97	0	3	0
Pt. Biserial	.17	*Error*	-.17	*error*

*Correct answer

Item 5 (as in Table 4) has two non-functioning distractors (distractor B and D). This item is also having poor discrimination power and an outlier (too easy) in terms of item facility index. Distractors play a very vital role to ensure the reliability of the score in a test. According to Vegada, Shukla, Khilnani, Charan dan Desai (2016), test takers who possess the ability to eliminate distractors will reduce the reliability of the test as well. Aside from having poor discrimination and unacceptable facility index, with the non-functioning distractors, this item also needs to be discarded and replaced.

Table 5. Item Analysis for Item 13

Discrimination Index				.22
Facility Index				.31
Options	A	B	C	D*
Count	15	36	16	65
Percentage	11.4	27.3	12.1	49.2
Pt. Biserial	-.28	-.05	.00	.22

*Correct answer

Despite having good discriminating power (*.22*) and average difficulty (FI = *.31*), Item 13 has one low-efficient distractor which is distractor C. Point Biserial (*.00*) indicates that half of those who chose option C were the high-achievers. Papenberg and Musch (2017) in their study concluded that items with good quality distractors yielded more reliable score. In their study, they did a comparison between good distractors and the number of options where the study concluded that, an item with more

options will increase the item facility index without affecting the reliability. Therefore, items with good facility index and good discrimination power may have problematic distractors.

Table 6. Item Analysis for Item 16

Discrimination Index				.24
Facility Index				.54
Options	A	B*	C	D
Count	25	58	8	41
Percentage	18.9	43.9	6.1	31.1
Pt. Biserial	-.32	.24	.04	-.01

*Correct answer

Similar to Item 13, Item 16 is having a better discrimination power (*.24*) and average difficulty (FI = *.54*) yet there is one distractor which can be classified as low-efficient (distractor C). Point Biserial (*.04*) indicates that out of eight students who chose option C, more than half chose option C as the correct answer were the high-achievers. In relation to Papenberg and Musch (2017), Pawade and Diwase (2016) also suggested that low-efficiency distractor should be discarded or rephrased to ensure the overall quality of the items.

Table 7. The Association between Facility Index and Discrimination Power with Distractor Efficiency

	Facility Index	**Discrimination Power**
Distractor Efficiency (Correlation)	.39	-
p-value	.08	-

The correlation study suggests that there is no significant association between the efficiency of the distracter with the discrimination power of the items. Interestingly, there is a non-significant correlation between distractor efficiency with the facility, $r(11) = .39$, $p = .08$. The result of this study is inconsistent with Mahjabeen et. al. (2017) where their study found that there is a significant correlation between distractor efficiency. Nevertheless, their study found that the correlation between distractor efficiency with the discrimination power is non-significant.

CONCLUSION

This study calibrates item facility index and discrimination power of an English Proficiency final examination paper and these indices association with the distractor efficiency. The aim was to investigate whether the efficiency of the distractors may influence the facility index and item discrimination power. A total of 132 students were randomly selected from the population and the examination paper consists of 40 MCQs; reading comprehension and grammar items.

The results bring two crucial information to conclude the study, thus providing valuable reminders in item building process. Firstly, this study found out that discrimination power may not be influenced by the index of facility. This study found that there are more items with poor discrimination power than the ones with unacceptable facility index. Nevertheless, all items with unacceptable facility index are having poor discriminating power. Therefore, this study suggests that a continuous process of developing and analysing items must be performed to ensure these indices are catered thus yielding a more reliable and valid results.

Secondly, even though this study found out that there is no significant relationship between these indices with the efficiency of the distractor, the first analysis with Rasch suggested a different motion. Most items with either low-efficiency or non-functioning distractor, are having either unacceptable facility indices or poor discrimination power or both. Interestingly, there are also few items with good indices with problematic distracters. Therefore, this study suggests that a careful item development process should be carried out in terms of the wordings as well as the structure of both stem and options.

In conclusion, this study found out that there may be a trivial association between facility index and discrimination power with the distractor efficiency and therefore this suggests for a deeper analysis to the linguistic level of the items. This also indicates that the development of items or instruments

requires a thorough pre and post analysis to ensure the quality.

REFERENCES

Bichi, A. A., Hafiz, H., & Bello, S. A. (2016). Evaluation of Northwest University, Kano Post-UTME Test Items Using Item Response Theory. International Journal of Evaluation and Research in Education, 5(4), 261-270.

Battisti, B., Hanegan, N., Sudweeks, R., & Cates, R. (2010). Using Item Response Theory to Conduct a Distracter Analysis on Conceptual Inventory of Natural Selection. *International Journal of Science & Mathematics Education*, 8(5), 845–868. DOI:https://doi.org/10.1007/s10763-009-9189-4.

Bond, T.G. & Fox, C.M. (2007). *Applying The Rasch Model: Fundamental Measurement in the Human Sciences* (2nd Ed.). London: Lawerence Erlbaum Associates, Publishers.

Crisp, G.T., & Palmer, E.J. (2007). Engaging Academics with a Simplified Analysis of their Multiple Choice Question (MCQ) Assessment Results. *Journal of University Teaching and Learning Practice, 4*(2), 88-106.

Büyükturan, E. B., & Sireci, A. (2018). Using Classroom Observation Scores Instead of Test Scores as Criterion in the Estimation of Discrimination Index. Journal of Education and Training Studies, 6(7), 55-62. DOI: https://doi.org/10.00004/jets.v6i7.3191.

Hassan S., & Hod, R. (2017). Use of item analysis to improve the quality of single best answer multiple choice question in summative assessment of undergraduate medical students in Malaysia. Education. *Medicine Journal*, 9(3), 33–43. DOI:https://doi.org/10.21315/eimj2017.9.3.4.

Ibbet, N. L., & Wheldon, B. J. (2016). The Incidence of Clueing in Multiple Choice Testbank Questions in Accounting: Some Evidence from Australia. *e-Journal of Business Education & Scholarship of Teaching*, 10(1), 20-35.

Iqbal, M. Z., Irum, S., &Yousaf, M.S. (2107). Multiple choice questions; Developed by thefaculty of a public sector medical college. *Professional Med J*, 24(9), 1409-1414.

Iwintolu, R. O., & Afolabi, E. R. I. (2018). Construction and Validation Of Polytomously-Scored Multiplechoice Items in Mathematics. *Bulgarian Journal of Science and Education Policy*, 12(2), 359-379.

Jonick, C., Schneider, J.,& Boylan, D. (2017). The effect of accounting question response formats on student performance. *Accounting Education26*(4), 291-315. DOI://doi.org/10.1080/09639284.2017.1292464.

Jonm. (2012). *Question Analysis – Difficulty/Facility and Discrimination*. Retrieved from: https://learntech.imsu.ox.ac.uk/blog/question-analysis-difficultyfacility-and-discrimination/.

Kamarul, A. A. (2018). *Item Analysis and Score Dependability of English Proficiency 1 Examination in Public University* (Master dissertation). University of Malaya, Kuala Lumpur, Malaysia.

Karkal, Y. R., & Kundapur, G. S. (2016). Item analysis of multiple choice questions of undergraduate pharmacology examinations in an International Medical School in India. *Journal of Dr. NTR University of Health Sciences*, 5(3), 183-186.

Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. *International Journal of Applied and Basic Medical Research*, 6(3), 170-173.

Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2017). Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions. *Annals of Pakistan Institute of Medical Sciences*, 13(4), 310–315.

Papenberg, M., & Musch, J. (2017). Of Small Beauties and Large Beasts: The Quality of Distractors on Multiple-Choice Tests Is More Important Than Their Quantity. *Applied Measurement in Education*, 30(4), 273–286. DOI:https://doi.org/10.1080/08957347.2017.1353987.

Pawade, Y. R., & Diwase, D. S. (2016). Can Item Analysis of MCQs Accomplish the Need of a Proper Assessment Strategy for Curriculum Improvement in Medical Education? *Journal of Educational Technology*, *13*(1), 44–53.

Puthiaparampil, T., & Rahman, M. (2021). How important is distractor efficiency for grading Best Answer Questions?. BMC medical education, 21(1), 1-6. DOI:https://doi.org/10.1186/s12909-020-02463-0

Rao, C., Prasad, H. K., Sajitha, K., Permi, H., & Shetty, J. (2016). Item analysis of multiple choice questions: Assessing an assessment tool in medical students. *International Journal of Educational and Psychological Researches*, 2(4), 201-204.

Serpil KOÇDAR, Nejdet KARADAĞ and Murat Doğan ŞAHIN. (2016). Analysis of the Difficulty and Discrimination Indices of Multiple-Choice Questions According to Cognitive Levels in an Open and

Distance Learning Context. *The Turkish Online Journal of Educational Technology 15*(26), 291–315. DOI:https://doi.org/10.1080/09639284.2017.1292464

Si-Mui Sim & Raja Isaiah Rasiah (2006). *Relationship Between Item Difficulty and Discrimination Indices in True/False-Type Multiple Choice Questions of a Para-clinical Multidisciplinary Paper*. Retrieved from: http://www.annals.edu.sg/pdf/35VolNo2200603/V35N2p67.pdf

Singh, A. K. (2009). *Tests, Measurements and Research Methods in Behavioural Sciences*. New Delhi: Bharati Bhawan Pub. & Dis.

Toksöz, S., & Ertunç, A. (2017). Item Analysis of a Multiple-Choice Exam. *Advances in Language and Literary Studies*, *8*(6), 141–146.

Vegada, B., Shukla, A., Khilnani, A., Charan, J., & Desai, C. (2016). Comparison between three option, four option and five option multiple choice question tests for quality parameters: A randomized study. *Indian Journal of Pharmacology*, 48(5), 571–575. DOI:https://doi.org/10.4103/0253-7613.190757.

Zafar Iqbal, M., Irum, S., & Sohaib Yousaf, M. (2017). Multiple Choice Questions; Developed by the Faculty of a Public Sector Medical College. *Professional Medical Journal*, 24(9), 1409–1414. DOI:https://doi.org/10.17957/TPMJ/17.3989.