Educational Evaluation and Learning Management Journal

https://mac.mans.edu.eg/

ISSN 2812-5312- eISSN 2812-5320 https://mac.mans.edu.eg/

Item analysis of multiple choice questions of a genetic term exam in an Egyptian Veterinary college for a viable question bank

Eman Zahran, Ahmed Mustafa

To cite this article: Eman Zahran and Ahmed Mustafa. Item analysis of multiple choice questions of a genetic term exam in an Egyptian Veterinary college for a viable question bank. Educational Evaluation and Learning Management journal 1:1 (2023) 1-5

To link to this article: https://doi.org/10.21608/eelmj.2023.172846.1001

Published online: 02 May 2023

Educational Evaluation and Learning Management journal 1:1 (2023) 1-5

https://doi.org/10.21608/eelmj.2023.172846.1001

Original Article

Item analysis of multiple-choice questions of a genetic term exam in an Egyptian Veterinary college for a viable question bank

Eman Zahran[1*], Ahmed Mustafa[2]

[1*]Department of Aquatic Animal Medicine, Faculty of Veterinary Medicine, Mansoura University, Mansoura 35516, Egypt

[2]Department of Biological Sciences, Purdue University Fort Wayne, IN 46805, USA

ARTICLE HISTORY

Received: November 05, 2022

Revised: February 4, 2023

Accepted: March 6, 2023

*Corresponding author: Eman Zahran

E-mail address:

emanzahran@mans.edu.eg

Tel: +201558728807

ABSTRACT

Multiple-choice questions (MCQs) are useful for assessing student performance because they objectively cover a wide range of topics. Its dependability and validity are determined by how well it is built. Defective Items detected by item analysis must be checked for and optimized for item writing flaws. This study used item analysis to evaluate the test items for difficulty levels and discriminating power with functional distractors. A total of 623 students took a summative examination in Genetics. It comprised 60 single-response MCQs and 20 true or false (T/F) questions as a part of a 2-h paper for 25 marks. Items were categorized according to their difficulty index, discrimination index, and distractor efficiency. Among 60 MCQs, 32 have zero non-functioning distractors (NFD); 19 had one, 8 had two, and 1 had three. DIFI and DI were in the acceptable range; however, T/F items showed 15% (n=3), 10% (n=2), 5% (n=1) fair, poor, and negative discrimination, respectively. Non-functional distractors (NFD) were found as 53.3 % (n=32) of questions have 0 NFD of 100% DE, 7.92% (n=19) have DE of 66.6%, 3.33% (n=8) have DE of 33.3%. About 0.42% (n=1) have DE of 0%. Therefore, item analysis is a valuable tool for identifying poorly constructed test items, and optimizing these items is an essential step in producing a high-quality question bank.

Keywords: Test items, Difficulty index, Discriminatory index, Distractor efficiency

1. Introduction

Developing good quality Multiple-choice questions (MCQs) to comprehensively evaluate students' knowledge at the end of a semester is a difficult, time-consuming process and requires trained faculty. Therefore, a well-designed MCQ test should assess most of the intended learning outcomes, including the higher-order cognitive process of Blooms Taxonomy (Carneson et al.; Quaigrain and Arhin, 2017), which is considered a significant part of how teachers receive feedback on their instructional practices. However, MCQs are frequently criticized for emphasizing what students can recall rather than gauging their capacity to apply and analyze course-related material (Kaur et al., 2016; Sharif et al., 2014). Moreover, the MCQ format allows students to guess even when unfamiliar with the subject (Biggs and Tang, 2011). Blind guessing, on the other hand, is relatively infrequent on well-written classroom tests, and informed guessing, based on a critical analysis of the question and the available options, offers a valid indicator of student achievement, according to Downing (2005).

There are two main types of multiple-choice tests: criterion-referenced tests (CRTs) and norm-referenced tests (NRTs). In criterion-referenced tests (CRTs), the goal is usually to decide whether or not a student can show mastery in a certain content area and competencies. In norm-referenced tests (NRTs), the goal is usually to rank the whole group of people so that you can compare their performances relative to each other (Hotiu, 2006). For example, a typical MCQ has a stem that is a question and a set of two or more options, usually from 3 to 5 options for the question. The key represents the best or correct response, while the other options are referred to as distractors, which also have a key role in determining the appropriate quality of MCQ (Burud et al., 2019; Cizek and O'Day, 1994).

A good distraction should be able to distinguish between informed and uninformed students (Quaigrain and Arhin, 2017). The reliability of test results is a significant

issue of MCQs. For this reason, item analysis is a very important process, including gathering, compiling, and utilizing data from students' responses to evaluate the quality of the test items (Burud et al., 2019; Quaigrain and Arhin, 2017). Therefore, it is a procedure that gives information about the reliability and validity of a test item. In item analysis, several terminologies are utilized, each of which requires adequate comprehension.

Difficulty index (DIFI) enables us to identify either too hard or too easy items. DIFI (p-value) ranges from 0% to 100%; the easier the question, the higher the percentage of items and the average difficulty level should be between 0.31 and 0.60. (31 % -60 %) (Kheyami et al., 2018). The discriminatory index (DI) tells the difference between students who are doing well and those who are doing poorly. The point-biserial correlation (DI) is a value between -1 and +1. It's +1 when more students in the upper group (the high achievers) get the question right, and it's -1 when more students in the lower group (the low achievers) get the question right. An item with a difficulty of 0 or 1 will always have a DI of 0, and DI is at its highest when DIFI is close to 0.50. It is favorable to have a DI between 0.15 and 0.25 (Burud et al., 2019; Kaur et al., 2016).

Distractors are an integral part of a question and significantly impact students' overall performance (Mehta and Mokhasi, 2014). Distractor efficiency (DE), which indicates whether the questions' distractions were well-chosen or ineffective in preventing students from selecting the correct answer, is therefore crucial (Kheyami et al., 2018). Everything other than the item's key must be close to the key. Functional diversions (FD) are those that are selected by more than 5% of examinees, while non-functional diversions (NFD) are those that are selected by fewer than 5% of examinees (Patil and Patil, 2015; Tarrant et al., 2009). DE is determined by the number of NFDs for each item and ranges from 0% to 100%.

The purpose of the present study was to evaluate the item and test quality and investigate the relationship between difficulty and discrimination indices and distractor performance (DE). Following the revision/disposal of the tested items, the post-evaluation information on items will be used to develop a question bank.

2. METHODS

An internal assessment for a Physiology course was undertaken by 623 undergraduate students of the first semester summative exam of their first year in January 2022. A total of 623 students participated in the test. The test consisting of 60 MCQs and 20 true or false questions, was based on the assessment blueprint. All respondents were first-year undergraduate students pursuing first-semester summative exams on the genetics course. It was conducted at Mansoura University- faculty of veterinary medicine in 2021. All ethical standards were strictly adhered to. Students' responses from the test were analyzed using Microsoft Excel (2016). Both test items were analyzed for their difficulty level, a measure of difficulty index (p-value), power of discrimination measured by the discrimination index (DI), and distractor analysis for all non-correct options. Using the following formulas, the indices DIF I, DI, DE, and non-functional distractor (NFD) were calculated for a total of 80 test items.

Index of Difficulty (DFI) = (UG+LG/N) x 100, where UG represents the upper group of students with high ability who correctly answered the question, while LG represents the lower group of students with low ability who correctly answered the question. N is the total number of students who correctly answered the question. The greater the difficulty index, the more difficult the item is believed to be.

Index of Discrimination (DI) = (UG-LG/N), where the item DI is the biserial point correlation between the item's correct response and the sum of all other item scores. Then, the total number of students in the top 27 percent who received correct responses and those in the bottom 27 percent who received correct responses were tallied. The greater the DI, the more effectively the test item discriminates between students with higher and lower test scores.

Distractor Efficiency (DE) = Non- Functional Distractor (NFD) from the distractor selected by less than 5% of the students (Sharma, 2021).

Table 1. Reference for item statistics.

Difficulty index (DFI) (Ranges)	Interpretation	Discriminatory index (DI) (Range)	Interpretation
< 0.20	Most difficult	Negative	Worst/ defective item
0.20-0.39	Difficult	< 0. 20	Not discriminating item, marginal item
0.40-0.59	Moderately	0.20- 0.29	Moderately discriminating, fair item
0.60-0.79	Moderately	.30- 0.39	Discriminating item, good item
.80-. 89	Easy	≥ 0.40	Very discriminating, very good item
> 0.90	Easiest		

Distractor Efficiency (DE)= Distractor efficiency ranged from 0 - 100% and was determined on the basis of the number of NFDs in an item. Three NFD: DE = 0%; 2 NFD: DE = 33.3%; 1 NFD: DE =66.6%; No NFD: DE = 100%.

3. Ethical Considerations

The study was performed after obtaining the institutional ethics committee's approval.

4. Results

A total of 80 test items and 60 MCQs with 240 distractors were analyzed. The DFI, DI, and DE were analyzed for both test items. Table (2) presented the DIFI and DI for MCQs item analysis.

Table 2: MCQs item analysis: Difficulty index (DFI) and Discriminatory index (DI) (n=60).

Difficulty index (DFI) (Ranges)	Interpretation	No. of item (%)	Discriminatory index (DI) (Range)	No. of item (%)	Interpretation
< 0.20	Most difficult	0	Negative	0	Worst/ defective item
0.20- 0.39	Difficult	1 (2)	< 0. 20	1 (1.67)	poor
0.40- 0.59	Moderately difficult	9 (15)	0.20- 0.29	0	Fair
0.60- 0.79	Moderately easy	39 (65)	.30- 0.39	7 (11.67)	Good
.80-. 89	Easy	11 (18)	≥ 0.40	52 (86.67)	Excellent
> 0.90	Easiest	0			

Distractor Efficiency (DE)= Distractor efficiency ranged from 0 - 100% and was determined on the basis of the number of NFDs in an item. Three NFD: DE = 0%; 2 NFD: DE = 33.3%; 1 NFD :DE =66.6%; No NFD: DE = 100%.

Regarding MCQs' difficulty index, about 18% (n=11) and 65% (n= 39) were easy and moderately easy questions, respectively. About 15% (n=9) were moderately difficult, while 2% (n=1) were the difficult question. Concerning the discriminatory index, (86.67%, n=52) items showed excellent discriminatory index, and 11.67% (n=7) showed good discrimination. Meanwhile, 1.67% (n=1) showed poor discrimination. No negative or fair DI has been recorded. Non-functional distractors (NFD) were found as 53.3 % (n=32) of questions have 0 NFD of 100% DE, 7.92% (n=19) have DE of 66.6%, 3.33% (n=8) have DE of 33.3%. About 0.42% (n=1) have DE of 0% (Table 3).

Table 3: MCQs Percentage of non-functional distractors (n=60).

No. of Nonfunctional Distractors (NFD)	number of Items	DE
0 NFD	32 (53.3)	100%
1 NFD	19 (7.92)	66.66%
2 NFD	8 (3.33)	33.33%
3 NFD	1 (0.42)	0%

Concerning the true or false (T/F) item analysis, about 15% (n=3) and 65% (n= 13) were easy and moderately easy questions, respectively. About 20% (n=4) were the difficult question. For the discriminatory index, 50% (n=10) items showed excellent discriminatory index, 20% (n=4) good discrimination; meanwhile, 15% (n=3), 10% (n=2), 5% (n=1) showed fair, poor, and negative discrimination (Table 4).

Table 4: True or false item analysis: Difficulty index (DFI) and Discriminatory index (DI) (n=20).

Difficulty index (DFI) (Ranges)	Interpretation	No. of item (%)	Discriminatory index (DI) (Range)	No. of item (%)	Interpretation
< 0.20	Most difficult	0	Negative	1 (5%)	Worst/ defective item
0.20- 0.39	Difficult	4 (20)	< 0. 20	2 (10)	poor
0.40- 0.59	Moderately difficult	0	0.20- 0.29	3 (15)	Fair
0.60- 0.79	Moderately easy	13 (65)	.30- 0.39	4 (20)	Good
.80-. 89	Easy	3 (15)	≥ 0.40	10 (50)	Excellent
> 0.90	Easiest	0			

5. Discussion

Classroom instruction must synchronize with the test items to achieve instructional validity, which requires a good test item and analyzing items. In addition to the DIFI and the DI, the DE is extremely important because the distractors' quality substantially affects the DIFI and the DI (Bhat and Prasad, 2021). Items with NFD (5% of examinees chose the distractor) are critical for establishing DE. The NFDs present in an item range from 0% to 100%. DE is indirectly proportional to NFD, and functional distractors increase DE (Hingorjo and Jaleel, 2012; Rehman et al., 2018). Items with high NFDs decrease both the DE and DI while increasing the DIFI; consequently, the item is easy for students but a poor predictor of their performance (Kheyami et al., 2018). DE is expressed as 0%, 33.3%, 66.6%, and 100% depending on the number of NFDs, which is 3, 2, 1, or 0 in each case (Elgadal and Mariod, 2021). DE provides the best direction for selecting or rejecting items for question banks. Items with 0% DE should be discarded, whereas items with varying DE percentages should be revised by replacing the distracters with better options and reusing them in future exams (Adiga et al., 2021).

In the current study, MCQs difficulty index was in the acceptable range, 80% items (P= 40-80), 2% items (P= 20-<40), and 15% items (p= 80- <90). Similarly, DFI for the true

or false items was acceptable also, 65% of items (P=40-80), 15% of items (P= 80- <90), and 20% of items (P= 20- <40). Our results are consistent with previous studies that showed DFI within the same range, Karelia et al. (2013) showed 61% of items in the acceptable range (P = 30–70%), 24% of items (P >70%), and 15% items (P < 30%). Patel and Mahajan (2013) conducted a study on 150 students for an MCQ test with 50 questions 40 (80%) items were in the acceptable range. Item analysis for MCQs test comprising 50 questions in the subject of anatomy done by Mehta and Mokhasi (2014) on 100 students reported DIF I of 31 (62%) items in the acceptable range (P = 30–70%), 16 (32%) items were too easy (P > 70%), and 3 (6%) items were too difficult (P < 30%), further, the P value of 26 (65%) items was in acceptable range (30–70%), 10 (25%) items were easy (P > 70%), and 4 (10%) items were difficult (P < 30%) (Kolte, 2015). Too difficult items (DIF I ≤ 30%) can lead to deflated scores, while the easy items (DIF I > 70%) may result in inflated scores and a decline in motivation (Gajjar et al., 2014). High DIF Items (>90%) should be placed either at the start of the test as "warm-up" questions to boost the confidence of students or discarded, similarly items with low DIF I (<30%) should be either revised or removed altogether (Kaur et al., 2016). The present study's most difficult or too easy questions were not recorded.

DI is a crucial item analysis indicator, distinguishing between skilled and non-skilled students. DI normally ranges from -1 and |+1; the negative DI could be linked either to an ambiguous question or an answer key that was wrongly marked. In this study, 52 MCQs had DI ≥ 40 (excellent), 7 MCQs had DI 30 - <40 (good), meanwhile only on1 MCQ had DI <20 (poor). True or false questions, 10 items had DI ≥ 40 (excellent), 4 items had DI 30 - <40 (good), and 3 items had DI 20 - <30 (fair); meanwhile, only two items and one item had DI <20 (poor), and negative DI. Questions with negative and poor DI should be discarded since it decreases the validity of the test. In the same context, out of a total of 100 items, 24 had DI < 0.2 (poor), 45 had DI ≥ 0.20 and ≤ 0.35 (good), and 31 had DI > 0.35 (excellent) (Patil and Patil, 2015). Six items (30%) with DI < 0.2, 4 (20%) items with DI ≥ 0.20 and ≤ 0.35, and 10 (50%) items with DI > 0.35 were reported by (Singh et al., 2014) on item analysis of 20 MCQs. Further, (Patel and Mahajan, 2013) reported 9 items had DI < 0.2, 21 items had DI ≥ 0.20 and ≤ 0.35, and 20 items had DI > 0.35 for item analysis of 50 MCQs. Mehta and Mokhasi (2014) showed that out of a total of 50 items, 15 (30%) items had DI < 0.2, 9 (18%) items was DI ≥ 0.20 and ≤ 0.35, and 26 (52%) items had DI > 0.35.

The fundamental principle when framing effective MCQs is that the distracters must be plausible, close to the correct response, increasing the likelihood that students will select these diversionary strategies instead of the right response. Unrealistic distractions prevent opportunities to test a learner (Bayir et al., 2011; Tarrant et al., 2009). Herein, out of the 240 distractors, 32 (53.3%) items had 0 NFD, while 19 (7.92%), 8 (3.33%), and 1 (0.42%) items contained 1, 2, and 3 NFDs, respectively. In the same context, Gajjar et al. (2014) conducted item analysis on 50 items with 150 distractors and found 133 were FD and 17 NFD. Mehta and Mokhasi (2014) reported that a total of 150 distractors were used, of which 69 had no response, 28 were FDs, and 53 were NFDs. Additionally, 263 FD and 37 NFD were found in a study of item analysis (Patil and Patil, 2015). Bhat and Prasad (2021) found 25 MCQs have FD; 7 had 1NFD, 5 had 2NFD, and 3 had 3NFD out of 40 MCQs.

Conclusion

Performing item analysis is an important aspect of quality assurance of examinations. It is concluded from the present study that considerable test items were within the recommended values. However, very few true or false items did not meet the requirement of well-designed question items, especially regarding their DI. Hence, these items can be revised or discarded, and a viable question bank can be prepared.

References

Adiga, M.N.S., Acharya, S., Holla, R., 2021. Item analysis of multiple-choice questions in pharmacology in an Indian Medical School. Journal of Health and Allied Sciences NU 11(03), 130-135.

Bayir, A., Sirkecioglu, A.N., Bayir, M., Haliloglu, H.I., Kocaman, E.M., Aras, N.M., 2011. Metabolic responses to prolonged starvation, food restriction, and refeeding in the brown trout, Salmo trutta: Oxidative stress and antioxidant defenses. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 159(4), 191-196. https://doi.org/10.1016/j.cbpb.2011.04.008

Bhat, S.K., Prasad, K.H., 2021. Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: A cross-sectional study. Indian journal of ophthalmology 69(2), 343.

Biggs, J., Tang, C., 2011. Teaching for quality learning at university. McGraw-hill education (UK).

Burud, I., Nagandla, K., Agarwal, P., 2019. Impact of distractors in item analysis of multiple choice questions. International Journal of Research in Medical Sciences 7(4), 1136-1139.

Carneson, J., Delpierre, G., Masters, K., Designing and managing MCQs. Appendix C: MCQs and Blooms taxonomy (Online) 2011 (Cited 2011 Feb2). Available from URLhttp. web. uct. ac. za/projects/cbe/mcqman/mcqappc. html.

Cizek, G.J., O'Day, D.M., 1994. Further investigation of non-functioning options in multiple-choice test items. Educational and Psychological Measurement 54(4), 861-872.

Downing, S.M., 2005. The effects of violating standard item writing principles on tests and students: the consequences of using flawed test items on achievement examinations in medical education. Advances in health sciences education 10(2), 133-143.

Elgadal, A.H., Mariod, A.A., 2021. Item Analysis of Multiple-choice Questions (MCQs): Assessment Tool For Quality Assurance Measures. Sudan Journal of Medical Sciences 16(3), 334-346.

Gajjar, S., Sharma, R., Kumar, P., Rana, M., 2014. Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine 39(1), 17.

Hingorjo, M.R., Jaleel, F., 2012. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. JPMA-Journal of the Pakistan Medical Association 62(2), 142.

Hotiu, A., 2006. The relationship between item difficulty and discrimination indices in multiple-choice tests in a physical science course. Citeseer.

Karelia, B.N., Pillai, A., Vegada, B.N., 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), 41-46.

Kaur, M., Singla, S., Mahajan, R., 2016. Item analysis of in use multiple choice questions in pharmacology. International Journal of Applied and Basic Medical Research 6(3), 170.

Kheyami, D., Jaradat, A., Al-Shibani, T., Ali, F.A., 2018. Item analysis of multiple choice questions at the department of paediatrics, Arabian Gulf University, Manama, Bahrain. Sultan Qaboos University Medical Journal 18(1), e68.

Kolte, V., 2015. Item analysis of multiple choice questions in physiology examination. Indian J of Basic & Applied Medical Research 4(4), 320-326.

Mehta, G., Mokhasi, V., 2014. Item analysis of multiple choice questions-an assessment of the assessment tool. Int J Health Sci Res 4(7), 197-202.

Patel, K., Mahajan, N., 2013. Itemized analysis of questions of multiple choice question (MCQ) exam. Int J Sci Res 2(2), 279.

Patil, V.C., Patil, H.V., 2015. Item analysis of medicine multiple choice questions (MCQs) for under graduate (3rd year MBBS) students. RESEARCH JOURNAL OF PHARMACEUTICAL BIOLOGICAL AND CHEMICAL SCIENCES 6(3), 1242-1251.

Quaigrain, K., Arhin, A.K., 2017. Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. Cogent Education 4(1), 1301013. https://doi.org/10.1080/2331186X.2017.1301013

Rehman, A., Aslam, A., Hassan, S.H., 2018. Item analysis of multiple choice questions. Pakistan Oral & Dental Journal 38(2), 291-293.

Sharif, M.R., Rahimi, S.M., Rajabi, M., Sayyah, M., 2014. Computer software application in item analysis of exams in a college of medicine. ARPN J Sci Tech 4, 565-569.

Sharma, L.R., 2021. Analysis of difficulty index, discrimination index and distractor efficiency of multiple choice questions of speech sounds of English. International Research Journal of MMC 2(1), 15-28.

Singh, J., Kariwal, P., Gupta, S., Shrotriya, V., 2014. Improving Multiple Choice Questions (MCQs) through item analysis: An assessment of the assessment tool. International Journal of Sciences & Applied Research 1(2), 53-57.

Tarrant, M., Ware, J., Mohammed, A.M., 2009. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. BMC medical education 9(1), 1-8.