ORIGINAL RESEARCH

Test Item Analysis and its Relevance to a Dental Foundation Course at King Saud University

Khaled M Alqahtani

Received on: 12 July 2022; *Accepted on:* 17 August 2022; *Published on:* 01 October 2022

Abstract

Aim. To conduct an item analysis of a dental foundation course (biostatistics) taught in the first semester of the foundation year at King Saud University (KSU), Saudi Arabia.

Materials and methods: A total of 32 students completed the final examination for a dental foundation course (biostatistics) that is taught in the first semester of the foundation year at KSU. This exam consisted of 30 test items. The test items were evaluated for their level of difficulty, the measure of difficulty index (*p*-value), power of discrimination as measured by the discrimination index (DI), and distractor analysis. In the data analysis, the test reliability of inter-item consistency, or how strongly the test items are connected was determined using the Kuder–Richardson formula.

Results: The average test score was 18; the standard deviation was 6.9, and the standard error of the mean was 2.2. The skewness for the scores was 0.31, which indicates that the distribution was positively skewed. The kurtosis was 1.82, which indicates that the distribution was almost normal. No correlation was found between item DI and the item difficulty level.

Conclusion: The item DI and item difficulty level had no relationship, indicating that the test items lacked practical and excellent discriminating power.

Clinical significance: Item analysis is a valuable test for determining the accuracy and quality of multiple choice items, and it should be utilized when creating exams and assessments for dentistry students.

Keywords: Classroom assessment, Dental courses, Difficulty index, Discrimination index, Distractor analysis, Item analysis.

World Journal of Dentistry (2022): 10.5005/jp-journals-10015-2141

Introduction

Classroom assessment is a crucial component of the learning and teaching process.[1] From an analytical and quantitative standpoint, researchers have established valuable theories and learning taxonomies for assessing the academic skills, intellectual growth, and cognitive capacities of pupils.[2] In different situations, different types of evaluations are suitable. Item analysis is a useful tool for academic teachers can utilize to guide and enhance the quality of their training.[3]

Item analysis is a method of evaluating pupil responses to each test question (item) to measure the test's overall quality.[4] It is a valuable tool as it incorporates a number of several measures to determine the validity of the test as well as to improve the accuracy and quality of multiple choice items so they may be utilized in future tests. These statistical tools can be used to assess both the performance of examinees and the effectiveness and fairness of tests.[5] The test statistics summary and frequency table are one technique to analyze multiple choice tests which describe the distribution of test scores.[6,7] Another method is to employ item analysis statistics, which consider the class's overall performance on each test item.[4] It is crucial to know how to develop a well-designed exam as it is to know how to assess and apply data based on student test scores. It is critical to use test feedback to guide and enhance training which is an important step in the process. This may include the use of certain statistical indices such as item discrimination, item difficulty, reliability, and distractor analysis.[3]

The present study was carried out to conduct an item analysis of a dental foundation course (biostatistics) that is taught in the first semester of the dental foundation year at KSU and evaluate its validity and reliability.

Department of Mathematics, College of Science & Humanity Studies, Prince Sattam Bin Abdulaziz University, Al-Kharj, Kingdom of Saudi Arabia

Corresponding Author: Khaled M Alqahtani, Department of Mathematics, College of Science & Humanity Studies, Prince Sattam Bin Abdulaziz University, Al-Kharj, Kingdom of Saudi Arabia, Phone: 00966544455222, e-mail: k.alqahtani@psau.edu

How to cite this article: Alqahtani KM. Test Item Analysis and its Relevance to a Dental Foundation Course at King Saud University. World J Dent 2022;13(S-1):S72–S75.

Source of support: This project was supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project No. 2018/01/8321.

Conflict of interest: None

Materials and Methods

A cross-sectional study was carried out at the Faculty of Dental Medicine of KSU, Saudi Arabia. After a dental foundation course (biostatistics) taught in the first semester of the dental foundation year, 32 students took the MCQs test comprising 30 questions with a single best response. The test duration was 2 hours. There were four possibilities for each item, one of which was the correct answer while the other three were distractions. One point was given for each right answer. For erroneous responses, there were no negative marks. The test was evaluated by the authors. The study

was conducted in July 2021. All the students who completed the course were included in the study.

Item Analysis Procedure[9]

The mean item difficulty (or mean item ease), difficulty index, reliability coefficient, and item DI were calculated using the results of the student's exam performance.

The percentage of respondents (examinees) choosing the correct response for items (multiple choice items) with one correct alternative worth one point is known as the item difficulty and is given by

$$p = \frac{c}{n}$$

Where p = the difficulty factor, c = the number of respondents selecting the correct answer to an item, and n = the total number of respondents.

Item difficulty plays a crucial role and determines whether or not pupils have understood the topic being evaluated. It also aids in determining if an item can distinguish between students who know and do not know the topic being examined.[3]

The ideal (or moderate or desirable) item difficulty level (p_m), is defined as a point halfway between the probability of success (p_s), of correctly answering the multiple choice item and a perfect score for the item to maximize item discrimination, and is given by

$$p_m = p_s + \frac{1 - p_s}{2}$$

(i) Using the formula above, p_m for multiple choice questions may be quickly estimated, as shown in the Table 1.[10]

Item DI is a fundamental measure of an item's validity. It is described as an item's capacity to distinguish between high and low achievers based on the same criteria, that is, (1) internal criterion, for example, the test itself; and (2) external criterion, for example, achievement or intelligence test. Furthermore, the item DI calculation presupposes that the test scores distribution is normal and that the wrong or right dichotomy of a pupil's performance on an item is based on a normal distribution.[11]

Let the pupils' test scores be ranked ordered from lowest to highest.

$$p_u = \frac{\textit{Number of students in upper } 25\%\text{: } 30\% \textit{ group answering the item correctly}}{\textit{Total number of students in upper } 25\%\text{: } 30\% \textit{ group}}$$

and

$$p_l = \frac{\textit{Number of students in lower } 25\%\text{: } 30\% \textit{ group answering the item correctly}}{\textit{Total number of students in lower } 25\%\text{: } 30\% \textit{ group}}$$

According to some, the analysis group is made up of the lower 27% and the upper 27%. The item discrimination index (D) is calculated as follows:

$$D = p_u - p_l$$

Table 2 provides a guideline for the value of an item discriminating index.[12]

The average DI for all test items combined is the mean item DI. A significant positive number (>0.30) shows that the upper and lower performing pupils are well separated. Tests that don't discriminate well aren't particularly dependable, and they should be re-evaluated.

The test items were evaluated for their level of difficulty, the measure of difficulty index (p-value), power of discrimination as measured by the DI, and distractor analysis. Data were analyzed with simple proportions, mean, standard deviations, skewness, and kurtosis. The Kuder–Richardson method was utilized to determine the test reliability of inter-item consistency, or how well the test items are associated. The Kuder–Richardson formula 20 was used to analyze the internal consistency reliability of the test findings for our test, which had multiple choice questions that were scored incorrect or correct and were administered just once.[8] We created an R package that can be used to analyze any test item. The data is first imported as an excel file in this package. However, the package can also import any type of file extension. The R code is available upon request.

Results

The exam comprised 30 items. The scores of 32 pupils ranged from 29 to 10. The mean test score was 18; the standard deviation was 6.9; and the standard error of the mean was 2.2 (Table 3). The score skewness was 0.31, indicating that the distribution was positively skewed. The kurtosis of the distribution was 1.82, indicating that it was almost normal.

Based on the statistical evaluations, the middle value of KR20 = 0.88 for this biostatistics course implies higher reliability. The standard error of measurement was 2.2 and an ideal mean value was 18 (>15) for the test. These findings are depicted in the scatter plots and bar charts created using R for different test item data, such as item difficulty (p), item DI, and so on (D), which are presented below in Figures 1 and 2. Figure 1 shows the distribution of item difficulty for the thirteen questions. It ranges from 0.2 to 1, where most of the item difficulty lies between 0.4 and 0.6. Whereas Figure 2 shows the distribution of the DI for the 30 questions. It ranges from 0 to 1, where most of the discrimination index lies between 0.2 and 0.8.

The relation between the difficulty index and the DI is seen in Figure 3. It shows that there is no link between the item DI and item difficulty. This indicates that the test items lacked excellent and effective discriminating power, demonstrating that item DI and item

Table 1: Item difficulty for multiple choice questions

Number of alternatives	*Probability of success* (p_S)	*Idea item difficulty level* (p_M)
Two	0.50	0.75
Three	0.33	0.67
Four	0.25	0.63
Five	0.20	0.60

Table 2: Item discriminating index's guideline

Item discrimination index, D	*Quality of an item*
$D \geq 0.5$	Very good item; definitely retain
$0.40 \leq D \leq 0.49$	Good item; very usable
$0.30 \leq D \leq 0.39$	Fair quality; usable item
$0.20 \leq D \leq 0.29$	Potentially poor item; consider revising
$D \leq 20$	Potentially very poor; possibly revise substantially, or discard

Table 3: Summary of test parameters

Exam	*Reliability*	*Mean*	*SD*	*SEM*	*p < 0.3*	*0.3 < p < 0.7*	*p > 07*	*D > 0.2*
109 STAT	0.99	18	6.9	2.2	1	22	7	28

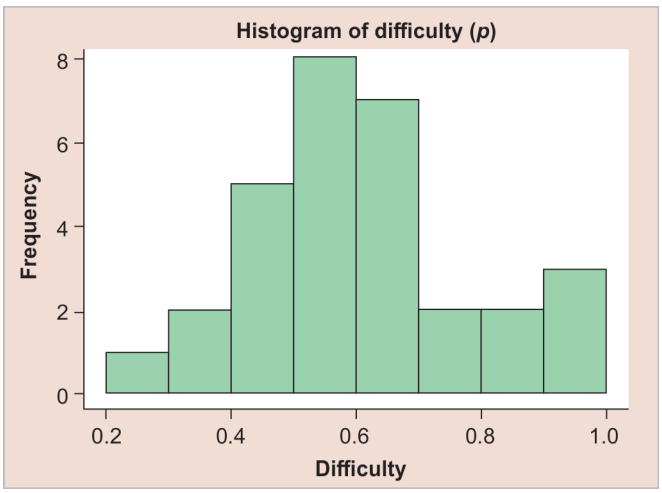

Fig. 1: Distribution of item difficulty for the thirteen questions

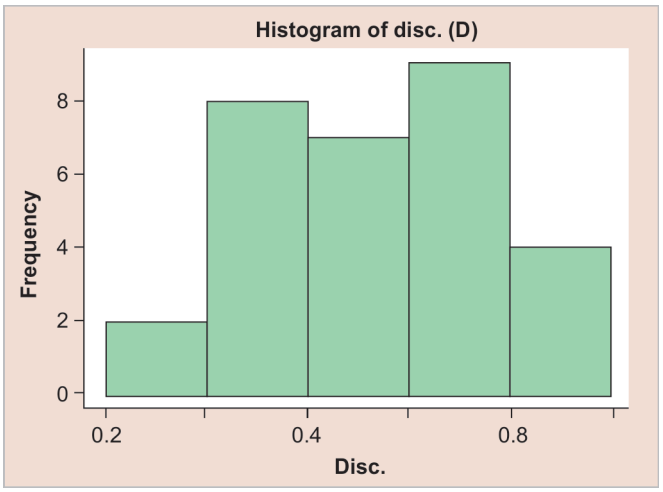

Fig. 2: Distribution of discrimination index for the thirteen questions

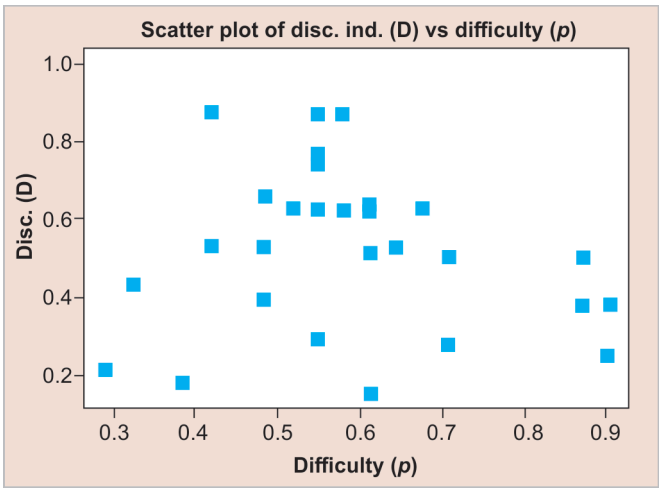

Fig. 3: Scatter plot showing relationship between difficulty index (*P*) and discrimination index of items

difficulty level had no relationship. Where the test is not legitimate, a good test should have a substantial positive connection between item difficulty and item DI.

Discussion

There should be synchrony of classroom instruction with test items to obtain instructional validity, which requires the development of good test items. Our statistical studies show how important it is to understand and use statistical analysis of test materials to enhance test construction and design This study might be useful in identifying the most important aspects of test item data and determining whether or not the test item needs to be revised. This paper's methodology may be used to demonstrate why test item analysis is crucial in classroom exams. These approaches may be used to assess, characterize, and enhance tests or surveys like college mathematics placement examinations, attitude surveys, mathematics study skills, information literacy, other general education learning outcomes, and test anxiety, among others.

Finally, this study endeavor has provided us with fresh insights into the biostatistics course's requirements. It has given us insight into the pupils' learning methods, individual characteristics, and skills. It has also provided us with the knowledge necessary to design legitimate and reliable assessments and examinations to increase student performance and accomplishment in the classroom. The difficulty and discrimination indices can both be used by instructors to assess the quality of test items. The average item difficulty in this study was 0.6. The DI was 0.5 on average. Distractors must be analyzed to identify the relative usefulness of each test item. If certain options are consistently not selected, they are likely to be implausible and not of much use as test items. Hence, it is necessary to appropriately design plausible distractors to improve the quality of tests.[3]

Odukoya et al. conducted an Item analysis of university-wide multiple choice objective examinations in a Nigerian Private university and found that majority of the items did not meet the psychometric standard in terms of difficulty and distractive indices and recommended that these should be part of statutory quality assurance procedures.[13] Ingale AS et al. conducted research to assess the item and test the quality of MCQs and deal with the learning difficulties among medical students in India. They concluded that items having average difficulty and high discriminating power with functional distracters should be incorporated into tests to make them more robust.[14] Rehman et al. conducted an item analysis to evaluate the quality of MCQs being used to assess final-year dental students in the subject of Prosthodontics in Pakistan and opined that posttest item analysis of all summative and formative assessments should be conducted to improve the quality of healthcare education.[15]

This study used item analysis to determine the validity of the biostatistics final examination that is administered in the first year as a foundational course for dental students at KSU to determine the accuracy and quality of multiple choice questions in this examination. The test items in this study did not have strong and

effective discriminating power since there was no link between item DI and item difficulty level. As a result, item DI and item difficulty level had no relationship. As a result, the effectiveness of the exam's test components must be evaluated. Assessment in the classroom is an important part of the teaching and learning process. Item analysis is a useful statistical technique to evaluate student responses to each test question (item), to determine the quality of each item and the test.

Conclusion

We applied the item analysis to a dental foundation course (biostatistics) taught in the first semester of the dental foundation year at KSU. As there was no correlation between item DI and the item difficulty level, the test items in this study did not have an effective and good discriminating power. Hence, there was no relation between item DI and item difficulty level. Hence, there is a need to evaluate the effectiveness of the test items.

Acknowledgments

This project was supported by the Deanship of Scientific Research at Prince Sattam Bin Abdulaziz University under the research project No. 2018/01/8321.

References

1. McMillan JH. Classroom Assessment. In: Wright JDBT-IE of the S& BS (Second E, ed. Oxford: Elsevier; 2015:819–824. DOI: 10.1016/B978-0-08-097086-8.92074-9
2. Adams NE. Bloom's taxonomy of cognitive learning objectives. J Med Libr Assoc 2015;103(3):152–153. DOI: 10.3163/1536-5050.103.3.010.
3. Quaigrain K, Arhin AK. Using reliability and item analysis to evaluate a teacher developed test in educational measurement and evaluationKing Fai Hui S, ed. Cogent Educ. 2017;4(1):1301013. DOI: 10.1080/2331186X.2017.1301013
4. Lange A, Lehmann IJ, Mehrens WA. Using item analysis to improve tests. J Educ Meas 1967;4(2):65–68. http://www.jstor.org/stable/1434299. DOI: 10.1111/j.1745-3984.1967.tb00572.x
5. Gajjar S, Sharma R, Kumar P, et al. Item and test analysis to identify quality multiple choice questions (MCQs) from and assessment of medical students of Ahmedabad, Gujarat. Indian J community Med 2014;39(1):17–20. DOI: 10.4103/0970-0218.126347
6. Agresti, A. Finlay B. Statistical Methods for the Social Sciences. Prentice Hall, Upper Saddle River, NJ. 1997.
7. Tamhane, A. C. Dunlop DD. Statistics and Data Analysis from Elementary to Intermediate. Prentice Hall, Upper Saddle River, NJ. 2000.
8. Saudi Commission for. Health Specialties. Item writing manual for multiple choice questions. Available at: http: //www.scfhs.org.sa /education /HighEduExams/Guidlines/mcq/Documents/MCQ% 20Manual.pdf. Accessed in September 2021
9. Shakil M, Campus H. Using beta-binomial distribution in analyzing some multiple choice questions of the final exam of a math course, and its application in predicting the performance of future students POLYGON-A Web-Based, Multi-Disciplinary Publication, Miami-Dade College, Hialeah Campus. 2009 Mar 5
10. Lord F. The relation of the reliability of multiple choice tests to the distribution of item difficulties. Psychometrika 1952;17(2):181–194. DOI: 10.1007/BF02288781
11. Panjaitan RL, Irawati R, Sujana A, et al. Item validity vs. item discrimination index: a redundancy? J Phys Conf Ser 2018;983:12101. DOI: 10.1088/1742-6596/983/1/012101
12. Chase C. Contemporary Assessment for Educators. January 1999.
13. Odukoya JA, Adekeye O, Igbinoba AO, et al. Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university. Qual Quant 2017;52(3):983–997. DOI: 10.1007/s11135-017-0499-2
14. Ingale AS, Giri PA, Doibale MK. Study on item and test analysis of multiple choice questions amongst undergraduate medical students. Int J Community Med Public Health 2017;4:1562. DOI: 10.18203/2394-6040.ijcmph20171764
15. Rehman A, Aslam A, Hassan SH. Item analysis of multiple choice questions. Pakistan Oral & Dental Journal 2018;38(2):291–293. DOI: 10.53350/pjmhs221621120