

Item Analysis as a Tool for Educational Assessment as Compared to Students, Evaluation to lectures

K.Mohammed Sowdani *, Huda Jaber **, Ara Murad Thomas **.

*Department of of Pharm. Chem., College of Pharmacy, Uni. Al-Mustansyriah

** Department of Clinical Lab. Sciences, College of Pharmacy, Uni. Al-Mustansyriah.

qasimsowdani1@gmail.com

DOI: <https://doi.org/10.32947/ajps.18.02.0380>

Abstract:

Item analysis is an effective method for assessing not only the test but can also reveal important issues of the whole educational process from curriculum design, implementation, assessment and evaluation. The method is suitable for MCQ type questions. After checking the reliability of the test, each question, item, is assessed using the criteria of Facility Value (FV) which represents the ratio of correct answers of each item and the discrimination index (DI) which takes into account the number of students in the upper quartile and that in lower quartile who answered the item correctly. An unacceptable question falls beyond the educationally acceptable limits of, $0.15 < FV < 0.85$ and $DI > 0.1$.

Unacceptable questions were identified and used to assess the teaching process of this subject material and as indicators of lecturers' performance, from the number of unacceptable questions and the average of FV and DI. The smaller the number of unacceptable questions, the better the teaching process of the subject material under study and the higher the performance of the lecture will be.

Method: Random (using six-sided die) and systematic (cumulatively adding the quotient of population to sample size) sampling methods were used and the reliability of the MCQ part of the test as well essay type was checked and found to be satisfactory with $R^2 > 0.7$ (the closer to unity the stronger the relationship and the higher the reliability). Results: For the subject material understudy, Pharmaceutical Organic Chemistry, for lecturer A, the first 13, items out of the 25 MCQ questions tested were found to satisfy the conditions set for FV and DI of acceptable questions. Conclusion: Acceptable items were identified and rated for further improvement in the stem or the distractors especially those near the border limits. For further improvement, item distractors need to be analyzed in detail. This method is effective in quantitatively rating lecturers' abilities in setting effective questions in relation to teaching objectives, the smaller the number of unacceptable items the better the performance of the lecturer. Unacceptable questions can be excluded or subject to future revision.

Key words: Item Analysis, Facility Value, Discrimination Index, Reliability

التحليل الانفرادي كأداة لتقييم التعليم مقارنة مع تقييم الطلبة للأستاذ الجامعي

قاسم محمد السوداني*

هدى جابر وحيد**

آرا مراد توماس**

*فرع الكيمياء الصيدلانية، كلية الصيدلة – الجامعة المستنصرية

**فرع العلوم المختبرية السريرية، كلية الصيدلة – الجامعة المستنصرية

الخلاصة:

يوفر التحليل الانفرادي للأسئلة طريقة فعالة ليس لتقييم الاسئلة فقط ولكن لاضهار جوانب مهمة في العملية التعليمية ابتداء من تصميم المنهج وتطبيقه والاختبارات ثم تقييم النتائج. هذه الطريقة ملائمة لتحليل الامتحانات ذات الاجابة الواحدة- أم سي كيو-. فبعد التأكد من ثبوتية الاختبار يتم فحص كل سؤال بحساب مواصفات قيمة السهولة والتي تمثل نسبة الاجابات الصحيحة لذلك السؤال والمواصفة الثانية هي دليل التمييز السؤال وبحسب من الفرق بين عدد الاجابات الصحيحة في الربع

الاسفل والربع الاعلى مقسمه على نصف عدد الاجابات الكلية لذلك السؤال. الاسئلة الغير مقبولة تقع خارج الحدود المقبولة عالميا لهاتين الصفتين وكالتالي:

$$0.1 > \text{قيمة السهولة} \quad \text{و} \quad 0.85 > \text{دليل التمييز} > 0.15$$

تمثل النسبة العالية لعدد الاسئلة الغير مقبولة مؤشرا سلبيا لمجمل الاعملية التعليمية للمادة الدراسية المراد تقييمها وبضمنها اداء التدريسي لتلك المادة.

طريقة الدراسة: تم انتقاء عينة 30 من العدد الكلي لطلبة المرحلة والبالغ 192 بطريقة الاختيار العشوائي باستعمال النرد السداسي الالوان ومن ثم اكمال الحصول على باقي النموذج الاختيار التسلسلي وذلك باضافة حاصل قسمة العدد الكلي 192 على عدد النموذج الى ناتج النرد الاول وعلى التوالي. وبعد ان تم فحص درجة الثبوتية للاسئلة ذات الاجابة الواحدة أم سي كيو MCQ والثبوتية للاسئلة الاخرى، كانت الثبوتية في الحالين مقبولة وذلك حسب مقياس درجة التشنت 2 وكان اعلى من 0.7 مقارنة مع الحالة المثالية للعلاقة الخطية وهي 1.0.

النتائج: المادة الدراسية تحت الدراسة الحالية هي الكيمياء الصيدلانية العضوية للمرحلة الثالثة للمدرس أ وبمشاركة المدرس ب بنسبة 50% من الاسئلة من رقم 1 الى رقم 13 من بين 25 - كانت ضمن المواصفات المقبولة من قيمة السهولة ودليل التمييز

الاستنتاجات كانت النتائج للاسئلة التي تم فحصها بطريقة الفحص الانفرادي للاسئلة وحسب المواصفات اعلاه من - قيمة السهولة ودليل التمييز - عالية الجودة حسب النسبة العالية جدا من الاسئلة المقبولة المواصفات الخاصة بقيمة السهولة ومعامل التمييز. كما تم وضع الاسئلة القريبة من الحدود المقبولة للمراجعة والتحسين المستمر لنوعية الاسئلة وتحسين الاداء بشكل عام.

الكلمات المفتاحية: التحليل الانفرادي، قيمة السهولة، دليل التمييز، الثبوتية

Introduction:

There is always a need for a more detailed method for evaluating exam results. Many educationalists criticize student's evaluation to lecturer performance. Most students are concerned mainly about examination results while lectures main concern is about the effective implantation of the curriculum and the criteria set to high quality education to produce graduates of high standards. Another factor of getting valid student evaluation is the setting, conduction and analysis of results. However, it is always important of to have feedback from the students during, informative, and at the end of the course, summative, as to review the whole process for continuously improving the learning outcomes. Therefore, some lectures might not get a fair students' evaluation.^[1]

An attempt to address this problem is to resort to item analysis, IA.^[2] In this evaluation technique, each question (item) is individually assessed using criteria including; facility value and discrimination index. In addition to that the reliability of test, which is normally an MCQ, has to be checked.

The main aim of item analysis (IA) is to assess a written examination, especially the

objective tests in particular the Multiple-Choice Questions. The latter is widely used in the medical field as large proportion of the immense subject material has to be addressed with the objective of improving the validity of the test. In addition to that MCQs are easy to mark if large number of students is to be assessed. However, writing such questions is time consuming and needs a lot of skills. MCQs are made of a clearly written short and unambiguous stem with options (usually 4 or 5) one of them is the key answer and the others are the distractors^[3].

Test evaluation through inspecting each item, question, is an effective tool for revealing not only positions of weaknesses and strengths but will reveal details about the effectiveness of curriculum implementation of which the lecturer plays the main role.^[4]

The main features of the individual items including facility value (FV) which tells about the difficulty of the item. It is calculated as ratio of the number of students who answered the item to the number of the students taking the test. The other important feature of the item is the discrimination index which is calculated

from the difference between the numbers of students in the upper quarter who answered the item correctly to the number of in the lower quarter who answered correctly then the difference is divided by half the total of students taking the test. Limits are usually set for these two criteria. Acceptable questions are those with $0.15 < FA < 0.85$ and $DI > 0.1$

However, some researcher set different limits. Unacceptable question is subjected to further analysis including the suitability of the stem and the effectiveness of each distractor in the item. Detailed study of each distractor will give a feedback for better curriculum implementation. However, IA offers vital tool to lecturer performance. A large number unacceptable item in a test is sign of ineffective curriculum implementation and has to be reviewed through more student's participation and more effective teaching methods and planning.^[5]

Method

In this study, 3rd year students in our College of Pharmacy took a mid-term exam which is set equally by two lecturers. The examination is of two parts; an objective MCQs (50%) and a subjective (essay) type questions form 50%. The examination is administered by the Central Examination Committee and closely monitored by the higher administration in the college. MCQ scores were done electronically.

A sample of 32 out of the population 192 from the list of marked papers scores produced electronically and supplied by the Examination Committee. The representative sample was chosen first by the random method, throwing a six-sided dice. The number 3 was shown and then

systematic method (adding 6 which is the quotient of the population divided by the sample size i.e. $192/32$) systematically produced a sample of size 32 numbers with items from the official class list 3,9,15,21 ...,192. Then the sample scores were placed in descending order in order to rank the upper and lower quarters. The Item criteria, Facility Index (FV) = number of students who answered a particular item correctly divided by the total number of students. The other main criteria of the item are the Discrimination Index (DI) was calculated from the ratio below:

$$DI = (UQ_{corr} - LQ_{corr}) / 1/2 \text{ Total}$$

Where UQ_{corr} is the number of students in the upper quarter who answered the item correctly, LQ_{corr} represents the number of students in the lower quarter who answered the specific item correctly.

Before subjecting an MCQ set to item analysis it is important to check the reliability of the test.⁴

In this work, the results of the two parts of the written test i.e. the MCQ (50%) and the subjective part (50%) were plotted and the degree of agreement reflected by the correlation coefficient, R^2 , was taken as a measure reliability. The closer the value of R^2 to unity the more reliable the test will be. R^2 is a measure of how close values are from the trend line i.e the strength of the correlation, values of R^2 close to 1.0 indicates a strong correlation between variables.^[8,9]

Results:

Table 1 shows a list of the 32-sample set withdrawn from the sample randomly and systematically to reduce bias. It shows the scores of MSQ part marked electronically and the essay part marked by lecturer A (items 1-13 and lecturer B (items (14-25)

Table 1 Sample of 32 students with grades of MCQ and Essay type questions.

No. in the class list	No.	MCQ/50	Essay/50	Total/100
3	1	24	25	49
9	2	24	21	45
15	3	38	45	83
21	4	22	29	51
27	5	20	26	46
33	6	24	35	59
39	7	26	38	64
45	8	26	26	52
51	9	22	37	59
57	10	16	20	36
63	11	42	30	72
69	12	40	46	86
75	13	24	38	62
81	14	28	36	64
87	15	40	35	75
93	16	30	45	75
99	17	28	40	68
105	18	26	34	60
111	19	34	47	81
117	20	34	37	71
123	21	20	17	37
129	22	30	39	69
135	23	12	11	23
141	24	26	25	51
147	25	24	39	63
153	26	32	23	55
159	27	30	33	63
165	28	24	44	68
171	29	26	31	57
177	30	26	34	60
183	31	18	19	37
189	32	30	27	57

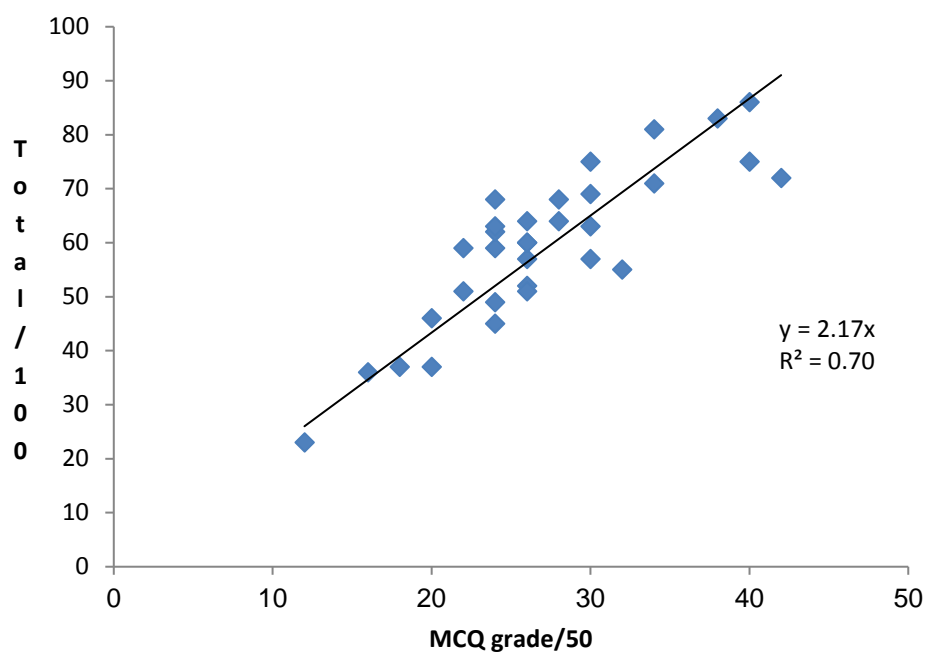


Figure 1: Reliability of MCQ grade of Pharmaceutical Chemistry

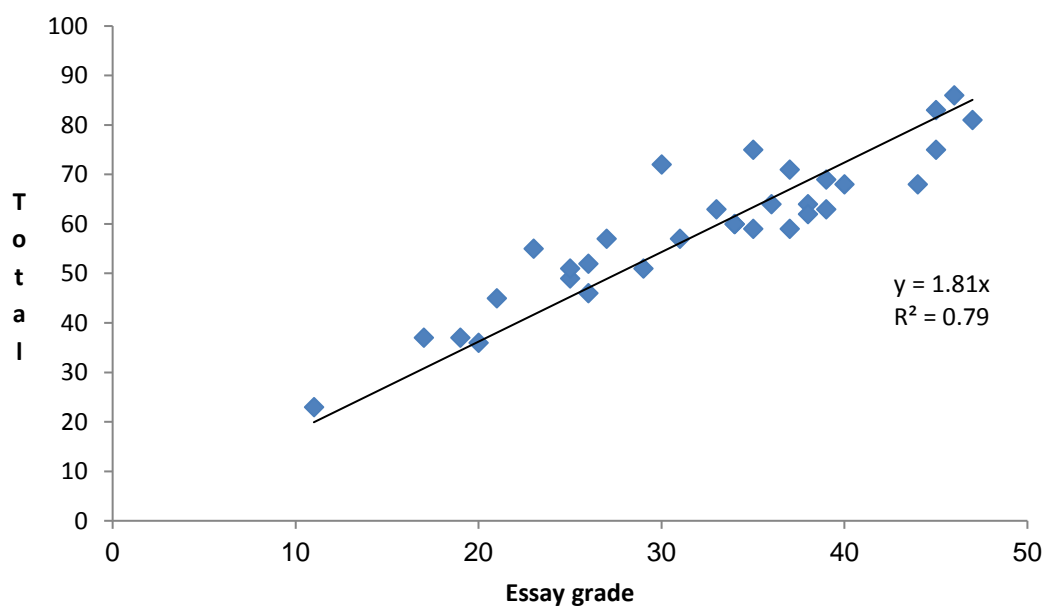


Figure 2: Reliability of Essay grades of Pharmaceutical Chemistry

Table 2: Facility Values and Discrimination Indices for the student Sample.

Q. No.	Correct Ans. /32	FV	Difficulty (P) Description	UQ, LQ	DI	Discrimination. Description
1	20	0.63	Mod. Diff.	6,4	0.13	Fairly Good
2	7	0.22	Mod. Diff.	5,1	0.25	Fairly Good
3	8	0.25	Mod. Diff.	4,1	0.19	Fairly Good
4	6	0.19	V. Diff.	4,0	0.25	Good Item
5	17	0.53	Mod. Diff.	8,1	0.44	V. Good
6	11	0.34	Mod. Diff.	6,1	0.31	Good Item
7	11	0.34	Mod. Diff.	5,1	0.25	Good Item
8	15	0.57	Mod. Diff.	7, 2	0.31	Good Item
9	14	0.44	Mod. Diff.	6, 2	0.25	Good Item
10	18	0.56	Mod. Diff.	6, 4	0.13	Fairly Good
11	9	0.28	Mod. Diff.	4,2	0.13	Fairly Good
12	12	0.38	Mod. Diff.	4,1	0.19	Fairly Good
13	9	0.28	Mod. Diff.	5,1	0.25	Fairly Good
		Mean=0.39			Mean=0.24	
		St Dev=0.15			St.Dev=0.087	

Key: $0 < P < 0.2$ Very Difficult, $0.2 < P < 0.8$ Moderately Difficult, $P > 0.8$ Very Easy
 $0 < D < 0.1$ Poor, $0.1 < D < 0.3$ Fairly Good, $0.3 < D < 0.4$ Good, $0.4 < D < 0.6$ Very Good,
D = Discrimination

Table 3: Students Questionnaire
Please fill-in the following questions as part of your duty to improve the educational process

No	Question	1	2	3	4	5	6	7	8	9	10
1	Do you have detailed information about the syllabus										
2	To what extent does the lecturer follow the syllabus										
3	Do you know about the following next lecture in advance										
4	Does the lecturer demonstrate knowledge in the subject material										
5	Does the lecturer taught you in away helped you to understand the subject?										
6	Does he/she motivated you to show interest in the subject?										
7	Does he/she use methods to help you understand the subject?										
8	What about the lecturer pace (neither fast nor slow)										
9	Was his/her language and writing clear and correct?										
10	Does his/her encourage you to participate in the lecture?										
11	Does his/her change method to help u understand when it needs?										
12	Does his/her create an atmosphere that helps in learning?										
13	Does the lesson go smoothly?										
14	Is there a waste of time as an attempt to control the lecture hall?										

[illegible]

Discussion:

Table-1 shows the results of the chosen sample ranked in descending order. The names of students were hidden for privacy. Random and systematic sampling methods were used to choose a sample free of bias. Figures 1 and 2 of the test reliabilities show consistent satisfactory results with R2 of 0.70 and 0.79 for MCQs and the essay parts of the score of the subject understudy respectively. This finding

indicates that MCQs can be used as a means of assessment in this subject. However, the essay type part is slightly more reliable than the MCQs. This might be attributed to the preference of the essay type questions by students leaving less time for the MCQs. ^[10]

Because two lecturers were involved (referred to as Lecturer A and lecturer B) were involved in teaching 3rd year with equal weighing, details of FA and ID analyses of the first 13 questions only is

shown in Table 2. For the first 13 questions (set by lecturer A) the mean and standard deviation of FA and DI were 0.39 ± 0.15 and 0.24 ± 0.087 respectively. However, only the final means and standard deviations FA and DI of the all the 25 items will be reported as mean and standard deviation and found to be 0.17 ± 0.12 and 0.54 ± 0.13 respectively.

The results are of expectable difficulty and the questions were of high standards and reflect the coverage of the curriculum and the efforts in setting of the exam and satisfactory coverage of the curriculum and its implementation.^[11,12]

Using the widely accepted^[2,3] range of acceptable item criteria of

$0.15 < FA < 0.85$ and $DI > 0.1$.

All the tested 13 items were acceptable. However, this statement does not necessarily apply for the rest of the test items 14-25.

For further improvement, questions $ID < 0.3$ should be the subject further improvement of stems and distracters.

All Items with $ID < 0.1$ should be eliminated and items with $ID < 0.2$ should be revised.

Items beyond the range $0.1 < FV < 0.9$ should be eliminated or revised because they are too easy or too difficult hence they are of poor quality.

For lecture evaluation, the IA results generally agreed with of student's evaluation of lecture A. When the upper and lower quarters students were asked to fill in the questionnaire shown in Table 3, it was found that upper quarter candidates have positive evaluation of the use of MCQs as a main part of the test have positively evaluated lectures with a score of 80% to lecturer A and 75% to lecturer B. On other hand, lower quarter candidates complained of shortage of time and bad time management for answering the essay type question leaving inadequate time for the MCQs. They gave lectures a score of 55% to lecturer A and 60% to lecturer B. The student Questionnaire is attached. (13)

Therefore, in this study, the student evaluation confirmed the finding of item analysis.

Some researcher uses DIF instead of FV, However. It helps in determining whether the students learned the concept being tested.^[14, 15]

A 100 MBBS students of medicine for 100 MCQs, mean DIF I of 48.90 ± 13.72 was reported with P value of 35 (22%) items was in the acceptable range (30–70%).^[16]

In another study on item analysis done by Patel and Mahajan reported 40 (80%) items were in acceptable range ($P = 30-70\%$).^[17]

Item analysis done by Mehta and Mokhasi and found (62%) items in the acceptable range ($P = 30-70\%$ Kolte reported (65%) items was in acceptable range (30–70%).¹⁸

It is recommended that Items with high FV I ($>70\%$) should be placed either at the beginning of the test as “warm-up” questions to boost the confidence of students, On the other hand items with low high I ($< 30\%$) should be either revised.

In our study only one item which was too difficult. It is recommended to include some difficult items for better normal distribution of students' abilities, with 25% easy, 50% medium and 25% difficult questions.

In conclusion; our results were of similar features as reported by literature. However, our results have higher acceptable rate than most researchers.¹⁹ This study has also proved that IA method can be successfully used as an additional quantitative indicator for evaluating the entire teaching process for organic pharmaceutical chemistry lecturer performance as it has been in agreement with responses shown by the higher quarter students who filled the students' questionnaire. Therefore, the high rate of acceptable MCQ questions reflecting the great efforts set in preparing high quality questions by lecturers of this institution, the College of Pharmacy, as lecturer A was chosen randomly for suitability (availability of data).

References

- 1- Popham, W. J. Classroom assessment: What teachers need to know. 5th edition. USA: Allyn & Bacon; 2008.
- 2- J.A. Talebi. Item Analysis as Effective Tool for Assessing Exam Quality. Res. Dev. Med. Edu. 2013; 2 (2): 20-23.
- 3- Quaigrain, K. and Arhin, A.K. Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. Cogent Education. 2017; 4(1).
- 4- Mozafer, R. H., & Farhan, J. Analysis of one-best MCQs: The Difficulty index, discrimination index and distractor efficiency. Journal of Pakistan Medical Association. 2012; 62, 142–147.
- 5- Downing, S. The effects of violating standard item writing principles on tests and students: Advances in Health Sciences Education Theory and Practice. 2003; 10, 133–143.
- 6- Dufresne, R. J., Leonard, W. J., and Gerace, W. J. Marking sense of students' answers to multiple-choice questions. The Physics Teacher. 2002; 40, 174–180
- 7- Tarrant, M., Ware, J., and Mohammed, A. M. An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. BMC Medical Education; 2001, 9, 861.
- 8- Carroll, R. G. Evaluation of vignette-type examination items for testing medical physiology. American Journal of Physiology. 1993; 264, 11–5.
- 9- Orak SN, Alpar EŞ. Validity and reliability of the Nurses' Professional Values Scale's MÜSBED. 2012; 2:22-31.
- 10- Siri A, Freddano M. The use of item analysis for the improvement of objective examinations. Procedia - Social and Behavioral Sciences. 2011; 29: 188-197.
- 11- Fidan NK. Prospective teachers' view of values education. Theoretical Educ. Sci. 2009; 2(2):1-18.
- 12- Ory J. Teaching Evaluation: Past, Present, and Future, New Directions Teaching and Learning. 2009; 83:13.
- 13- Gray M. Student Teaching Evaluations, Academic Search Premier. 2003; 89, 5, pp: 46.
- 14- Singh D., Gupta P., Singh T. Principles of Medical Education. 4th Ed. India: Jaypee Brothers Medical Publishers. 2013.
- 15- Gajjar, S., Sharma, R., Kumar, P. and Rana, M. 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. 2014; 39(1), 17-20.
- 16- Patel VC and Patel HV. Item Analysis of Medicine MCQs for Undergraduate 3rd year MBBS students. Res J Phar BiolChemSci 2015; 6, 1242-1251.
- 17- Patel KA and Mahajan NR. Itemized Analysis of questions of MCQs. Int J Sci Res 2013; 2:279-280.
- 18- Kolte V. Item Analysis of MCQS in Physiology examination. Indian J Basic Appl Med Res 2015; 4: 320 -326.
- 19- Mahajan R. Item Analysis of MCQS in Pharmacology examination. Indian J Basic Appl Med Res 2016; 6: 170 -173