

# Item's Analysis in Health Education at the High School of Health Science of Tunis

Kaouther MEJRI<sup>1</sup>, Sonia MAHJOUB<sup>2</sup>, Fatma AYDI<sup>3</sup>, Issam SALOUAGE<sup>4</sup>

<sup>1,2,3,4</sup> *The high school of health science and technologies of Tunis. El Manar University, Tunisia*

<sup>2</sup> *Faculty of Medicine of Tunis, El Manar University, Tunisia*

**Abstract:** Assessment is a crucial step in learning evaluation since it reflects the improvement of the quality of education. Thus, educators are always challenged while analyzing the performance of their students. The study aims to evaluate the quality of written exams set for under graduated students (2017-2019) in the biology department of the high school of health science and technologies. We used the item analysis tools: difficulty index, discrimination index and for the reliability the Cronbach alpha assessing available data during the study period (2017-2019). The study involved 2960 copies related to 104 tests including 1367 questions. We found that 5.3% of the questions were difficult, 44.18% were easy 49.52% with moderate difficulties, 19.38% with excellent discrimination, 10.53% with good discrimination. The test reliability was good in only 10.5% of the tests. Our research objectively analyzed the quality of questions, yet it revealed the presence of some deficiencies suggesting the improvement of the quality of assessment.

**Key words:** assessment, difficulty index, discrimination index, reliability.

**Abbreviations:** SAQ: short assay questions, LEQ: long assay questions, MCQ: multiple choice questions, IWF: items writing flaws, FD: functional distractors, NFD: nonfunctional distractor, DE: distractor efficiency, SD: standard deviation

## I. INTRODUCTION

Assessment is a crucial step in learning evaluation. In fact, the role of student's appraisal in learning and their feed-back is to improve teaching-learning experience [1]. Thus, educators are always challenged while analyzing the performance of their students. Since they will be safe practitioners in future, paramedics and health students as well as their medical colleagues need to be efficiently assessed for their knowledge, skills and competency [2]. Health science students are evaluated by various methods: multiple choice questions (MCQ), short answer questions (SAQ), long essay questions (LEQ), case studies, etc [3]-[6]. Because of their objectivity, elimination of the assessor's favoritism, extensive coverage of the subject in short period, MCQ are frequently used compared to the other types. Actually, marking the later ones is time consuming and may involve assessor's biases.[7][8]

Item analysis is the tool used to evaluate the quality of tests. It analyzes the performance of the individual and the overall test using different parameters such as difficulty index, discrimination index and test reliability. [9][10]

The aim of the study was to evaluate the quality of written exams set for under graduated students (2017-2019) in the biology department.

## II. METHODS

### 1. Study Context and Ethical Approval

Ethical approval for the project was granted by the High school of health science and technologies ethics committee Ref: ESSTST-EC 19/07.

Our research is a retrospective descriptive study. It was conducted in July 2019 on assessment examination comprising MCQ forms, SAQ, LEQ and clinical case studies. The project included examinations of the main sessions of the academic years 2016-2017, 2017-2018 and 2018-2019. The assessment involved all levels of biological studies 1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> year.

A total of 1360 items were analyzed. The quality of the questions was evaluated in terms of Difficulty index (P), Discriminating index (DI), non-functioning distractors (NFDs), test reliability (Cronbach alpha), and student's performance (mean score and overall passing rate).

### 2. Difficulty Index (p)

It describes the percentage of students who answered correctly to given questions. It was calculated for each item using the formula:  $p = \text{mean rate} / \text{marking scheme}$  [11]. The difficulty index varied from 0 to 1 with cut-off values of  $\leq 0,3$  (difficult), from 0,3 to 0,7 (moderate) and  $\geq 0,7$  (easy). [12]

The MCQ did have neither the same number of options nor the same type or scoring, thus the corrected DI was calculated according to the formula  $p' = p - \frac{1-p}{N-1}$  with p the uncorrected difficulty index and N the number of options. [11]

### 3. Discrimination Index (DI)

Discriminating index (DI) is the ability of a test item to discriminate between high and low scoring examinees. It ranges between 0 and 1. Higher discriminating indices of a test indicate better and greater distinguishing competence of the test. It was calculated for each question after selection of the upper and lower 27% which has been demonstrated to be the most efficient fraction. It is calculated using the formula  $DI = (\rho+) - (\rho-)$  with (p+) the difficulty index for the high performers group and (p) the difficulty index of the low performers group [14].

In some circumstances the DI can be negative which means that the students with lower ability answer more correctly than those with higher ability. Such situation could be because of the complexity of item making and the process of guessing. The cut-off values of DI were interpreted

according to Ebel in 1965: [12] thus  $DI \geq 0,4$  excellent, 0,2-0,39 good, 0,1-0,19 fair and  $< 0,1$  poor discrimination. Discrimination index of 0.2 or higher was acceptable. Questions are considered ideal if their difficulty index was acceptable and their discrimination index was good. [15] The relation between difficulty index and discrimination index for each exam was determined by Pearson correlation.

#### 4. Distractor Efficiency

Functional distractors were considered as the ones selected by at least 5% of the students. Nonfunctioning distractors (NFD) are options of a question other than correct answer and selected by less than 5% of the examinees. Since MCQ did not have the same number of options, distractor efficiency (DE) was calculated for each question depending on the presence of NFD. For MCQ with 5 options, if the question contains respectively 0 NFD, 1 NFD, 2 NFD and 3 NFD thus DE corresponds to 100%, 66%, 33% and 0%. [16]

#### 5. Test Reliability

Exam's consistency was evaluated using Cronbach alpha. It was calculated for each test using the formula:  $\alpha = \frac{k}{k-1} \left[ 1 - \frac{\sum \sigma_j^2}{\sigma_T^2} \right]$  with  $k =$  the number of questions,  $\sigma_j^2 =$  rate variance for question  $j$  and  $\sigma_T^2 =$  rate variance of the test or the exam.

Consistency was interpreted according to George and Mallery [13]:  $\alpha \geq 0,90$  (Excellent), 0,80 - 0,90 (good), 0,70 - 0,79 (acceptable) and  $< 0,7$  (non-acceptable)

#### 6. Statistical Analysis:

The data were entered in the Microsoft Office Excel file and analyzed using SPSS software (version 21.0). Pearson's chi-square test was used to evaluate the association. The statistical significance level was maintained as  $p$ -value  $< 0.05$  in the statistical analysis.

### III. RESULTS

We analyzed tests with available scheming marks for all the students (1<sup>st</sup>, 2<sup>nd</sup> and 3<sup>rd</sup> level) in the biology department during 3 years (2017-2019). However, while collecting data, 20% of the evaluations were excluded because of ambiguities in scoring. This led to the analysis of 104 tests related to various courses such as anatomy, physiology, microbiology, biochemistry, hematology, immunology, genetics, cell biology ...

The study involved 2960 copies associated to 104 tests including 1367 questions. Those exams were addressed to 155 students. SAQ were noticed in 664 copies (48,57 %) while MCQ and LEQ covered respectively 495 (36,21 %) and 181 (13,24%) copies. Twenty-seven clinical-case-studies covered only 1,97%. The detailed distribution of these questions is shown in table I.

Table I: question distribution

Parameters	Academic years			p value
	2016-2017	2017-2018	2018-2019	
Students n*	90	94	101	-
Total n	402	410	555	-
SAQ n (%)	192 (47,76)	201 (49,0)	271(48,82)	NS
LEQ n(%)	58(14,42)	78 (19,02)	45 (8,10)	<0.001
Clinical case n(%)	15 (3,8)	4 (0,97)	8(1,44)	0.009
MCQ n(%)	137 (34,1)	127(30,97)	231(41,62)	0.001

NS: no significant; SAQ: short assay questions; LEQ: long assay questions; MCQ: multiple choice questions

\*(the total seems to be more than 155 because some students were counted at the first level in 2017 and at the 2<sup>nd</sup> level in 2018 as well as at the 3<sup>rd</sup> level in 2019)

The most common questions were SAQ (around 48%). They were similarly noticed during the academic years. A tendency toward decreasing the use of LEQ (till 8.1% in 2019) was observed with a significant difference through time ( $p < 0.001$ ). Likewise, the number of clinical case studies significantly diminished ( $p = 0.009$ ). However, a significant increase in the frequency of MCQ was noticed ( $p = 0.001$ ).

The assessment of the difficulty of the questions revealed that the mean value was within the moderate range during the three years. Yet a significant difference was observed in 2018 (55%) compared to 2017 and 2019. As for the easy ones, although they covered more than the third of questions, the lowest frequency was detected in 2018 (37%) with significant difference compared to those respectively in 2017 ( $p = 0.03$ ) and in 2019 ( $p = 0.007$ ).

We also noticed that the frequency of the difficult ones slightly diminished from 7,46% in 2017 to 5,22% in 2019 yet with no significant difference (table II). Detailed information of items distribution is represented in table III.

Table II: item distribution

Parameters	Academic years			p
	2016-2017	2017-2018	2018-2019	
Difficulty index (mean (SD))	0,64 (0,09)	0,61 (0,13)	0,64 (0,1)	NS
Difficult n (%)	30 (7,46)	27 (6,58)	29 (5,22)	NS
Moderate n(%)	191(47,51)	229 (55,85)	257(46,30)	0.008
Easy n (%)	181(45,03)	154 (37,57)	269(48,46)	0.003
Discrimination Index (mean (SD))	0,32 (0,08)	0,38 (0,15)	0,34 (0,08)	NS
Ideal questions n (%)	112(27,86)	145(35,36)	173(31,45)	NS
Test reliability (mean (SD))	0,3 (1,08)	0,43 (0,44)	0,34 (0,99)	NS
Passing rate %	82,43	78,25	82,45	NS

SD: standard deviation; NS: no significant

As for the discrimination index, its mean value was around 0,3 which means a good discrimination between students. In fact, 35.8% of the questions ( $n = 490$ ) had excellent discrimination and 16.9% had good discrimination ( $n = 232$ ).

The overall rate of ideal questions (with moderate difficulty and good discrimination) was 32%; it varied non significantly from 27% to 35%. The analysis of the correlation of difficulty and discrimination indices revealed a highly correlation in 13 tests, a moderate correlation in 12 tests and no correlation in the remaining 102 tests thus in 80%.

Studying tests consistency revealed a mean value of 0,37. Although they had no acceptable homogeneity, tests (n=11) with good reliability reached 10,5%.

When we focused on MCQ, (table IV) we noticed the presence of multiple response answers in the options. Thus, we assessed item writing flaw (IWF) in the stem (negative clause) and in the choices (the presence of more than one correct answer). We remarked that they varied from 80 to 85% in MCQ.

Table IV: MCQ item distribution

Parameters	Academic years			p
	2016-2017	2017-2018	2018-2019	
MCQ n	137	127	230	-
Options n	663	574	1114	
1 Correct answer per MCQ n (%)	24 (17,51)	19 (14,96)	50 (21,74)	NS
IWF n (%)	113 (82,49)	108 (85,04)	180 (78,26)	
Total distracters n (%)	323 (48,71)	282 (49,12)	312 (28)	-
FD n (%)	214 (66,25)	211 (74,82)	150 (48,07)	<0.001
NFD n (%)	109 (33,74)	71 (25,18)	162 (51,93)	
Distracters' Efficiency (mean (SD))	67 (35,2)	68,9 (39,06)	62,3 (35,97)	NS
DE 100% n (%)	61 (18,88)	62 (21,98)	68 (21,79)	0.001
DE 0% n (%)	17 (5,26)	20 (7,09)	27 (8,65)	0.006
Difficulty index (mean (SD))	0,62 (0,21)	0,58 (0,18)	0,63 (0,2)	NS
Difficult n (%)	12 (8,75)	9 (7,08)	10 (4,34)	NS
Moderate n (%)	63 (45,98)	84 (66,14)	120 (52,17)	0.003
Easy n (%)	62 (45,25)	34 (26,77)	101(43,91)	0.002
Discrimination Index (mean (SD))	0,25 (0,21)	0,27 (0,23)	0,42 (0,68)	0.001
Ideal questions n (%)	33 (24,08)	35 (27,55)	63 (27,39)	NS

NS: no significant; IWF: items writing flaws; FD: functional distractors;

NFD: nonfunctional distractor; DE: distractor efficiency; SD: standard deviation

As for the distractors they reached 917 (38%) of the options, 342 among them (37%) were nonfunctional. Mean values of distractors' efficiency varied non significantly from 62,3 to 68,9% per year. The highest rate of FD was observed in 2018 with a significant difference ( $p < 0.001$ ) compared to those in 2017 and in 2019. MCQ during 2017 had the lowest rate of DE of respectively 100% and 0% compared to the following years (with p value respectively of 0.001 and 0.006). Similarly, to the overall questions, no

significant difference was observed when considering the distribution of the mean difficulty index. Moreover, we noticed that MCQ with moderate difficulty had the highest rate in 2018 with significant difference compared to 2017 and 2019. As for the discrimination index, the highest level was detected in 2019 ( $p = 0.001$ ), however the comparison of ideal question distribution revealed no significant difference through academic years of the study.

We analyzed students' performance by assessing the rate of null items or "scoring-zero-item", and the percentage of correct items within a test "fully noted items". The mean frequency of null items (either false or unanswered statements) was 18,71 +/- 21,09 (range from 0 to 100% per test). Questions with more than 30% null items were observed in 32 tests (25,19%) of the cases. The investigation of fully noted items revealed that they varied from 4% to 50% per test with no significant difference among levels and through time (2017-2019).

The assessment of success rate among tests varied from 33% per test to 100% with a mean value of 80%. In fact, 16 tests (15.38%) were fully succeeded.

#### IV. DISCUSSION

Students' assessment is a necessary step in improving education, yet few (n=4) are the published data that analyzed the quality of exams in the Tunisian high educational system [14],[17]-[19]. Besides, in order to develop competent future health professionals (in our case laboratory technicians) a good assessment is needed. Thus, we elaborated this study analyzing the quality of assessment in the biology department of the high school of health science and technologies of Tunis. This work involved only exams data with available scheming marks (80% of the questions).

We noticed that SAQ are frequently used, they cover almost half the questions per year. MCQ are present in more than 30% each year, while the remaining part included LEQ and clinical case studies.

A high proportion of flawed item was noticed in MCQ, ranging from 80 to 85% per year. This frequency is similar to those reported by Pais et al while analyzing IWF in 8 anatomy exams. In fact, they reported a frequency varying from 40% to 70% per test [20]. Our result may reflect either the lack of preparation and time invested by educators when constructing MCQ, or lack of familiarity with MCQ items writing guidelines.

This lack of knowledge was also observed in a cross-sectional survey among dental faculty members of the gulf cooperation council countries which revealed that more than 48% of the 216 participants were unable to identify the flawed MCQ [21]. Our study reveals the need to introduce training sessions in order to improve MCQ writing qualities. In fact, a recent study reported that a longitudinal development program on MCQ writing skills significantly reduces the presence of IWF from 8,5 to 3% [22] and improves the quality of MCQ [23][24]. Similarly, a short training session led to the reduction of IWF, however, the high proportions of flawed MCQ revealed the need to continuous faculty development programs [25][26]

Assessing students' performance revealed slight variations in test passing rates through the three years yet with no significant difference. Similar results were reported in the literature [22]. When we focused on fully noted items, we noticed that 45% of the tests contained 4 to 50% of items with full score per test. A recent study analyzing full score items within hematology exams during 6 years, described a smaller range from 9 to 33% with a mean value of 18,7% [14]. As for null items, the current work showed that the quarter of the tests included more than 30% per test. This frequency is lower than that reported by Ben Salah et al (45%). Such results could be related to the lack of preparations or ambiguities related to IWF.

Evaluating the difficulty index, we noticed that its mean value was in the moderate range. This result is similar to recent Tunisian studies elaborated within the faculty of medicine assessing hematology and microbiology exams[14][19].

In the MCQ group, the mean difficulty index was within the acceptable range 0,49+/- 0,29 leading to the presence of 49% of the questions with moderate difficulty, 28% easy ones and 23% difficult. The analysis of distractors within this type of questions showed that 30% needs to be replaced and removed from the items bank as they were nonfunctioning. It was reported that low distractor efficiency provides valuable information about IWF [27]. The proportion of questions containing all functional distractors (DE=100%) varied from 17 to 21%. This rate is higher than that reported by Tarrant et al (13,8%) [28] and lower to that described by Patil et al (56.7%) [16]. Reevaluation the efficiency of these plausible distractors in different groups of students may be useful in comparative studies. The proportion of FD and 3NFD were during the academic year 2016-2017, 18% and 5%, during the academic year 2017-2018 were 21% and 7% and during the academic year 2018-2019 were 17% and 5%. In the other hand, a recent study reported that longitudinal training of educators greatly influenced this proportion by increasing FD while decreasing NFD [22]. A study by Testa et al, reported that DE within MCQ at the application levels was greater than those at "knowledge" and "comprehension" levels. [29]

As for the discrimination index, the mean value was in the acceptable range (0,3). Discrimination of items was excellent in 37% of the cases and good in 17% of the questions. Although similar indices were reported in medical exams, item distribution revealed some discrepancies [14][18][19]. In fact, items with excellent and good discrimination were respectively noticed in 19% and 29% [19] and in 20% and 17% of the cases [14]. Recently, it was reported that question items with high-level taxonomy performed better in discrimination indices. [30]

Ideal questions reached only 32% yet higher levels were reported in literature 47% [31].

Assessing the reliability revealed that 72% of the tests had unacceptable consistency. Since low Cronbach alpha values reveal that items are unrelated, thus the high rate noticed in our study may be due to the use of multiple formats questions within the same test. Moreover, these results show

that further analysis after developing training sessions on items writing guidelines are needed in order to assess assessment evolution.

*Limitation:* Our study lacks the analysis of cognitive levels of the items and their relation with questions difficulty and student's performance.

## V. CONCLUSION

Our study objectively assessed the quality of questions in the biology department of the high school of health science and technologies of Tunis, yet it revealed the presence of some deficiencies. Introducing training course about items writing guidelines and analyzing their influence on the quality of assessments will be helpful in evaluating the educational system. Thus, the present results with future inter collaborative work between professors and students' perceptions will be helpful in enhancing the quality of assessment.

*Funding:* This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## REFERENCES

- [1] Fuentealba C. (2011) The role of assessment in the student learning process. *J Vet Med Educ*: 38(2):157-62. doi: 10.3138/jvme.38.2.157.
- [2] Wass V, Van der Vleuten C, Shatzer J, Jones R.(2001) Assessment of clinical competence. *Lancet*: 357(9260):945-9. doi: 10.1016/S0140-6736(00)04221-5.
- [3] Palmer E, Devitt P. (2007) Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? *BMC Med Educ*:7:49. doi: 10.1186/1472-6920-7-49.
- [4] Mujeeb AM, Pardeshi ML, Ghongane BB. (2010) Comparative assessment of multiple choice questions versus short essay questions in pharmacology examinations. *Indian J Med Sci*: 64(3):118-24. doi: 10.4103/0019-5359.95934.
- [5] Baig M, Ali SK, Ali S, Huda N. (2014) Evaluation of Multiple Choice and Short Essay Question items in Basic Medical Sciences. *Pak J Med Sci*: 30(1):3-6. doi: 10.12669/pjms.301.4458.
- [6] Khan MU, Aljarallah BM.(2011) Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students. *Int J Health Sci (Qassim)*:5(1):39-43.
- [7] Holzinger A, Lettner S, Steiner-Hofbauer V, Capan Melser M. (2020) How to assess? Perceptions and preferences of undergraduate medical students concerning traditional assessment methods. *BMC Med Educ* :20(1):312. doi: 10.1186/s12909-020-02239-6.
- [8] Tariq, S., Tariq, S., Maqsood, S., Jawed, S., & Baig, M (2017). Evaluation of Cognitive Levels and Item writing flaws in Medical Pharmacology Internal Assessment Examinations. *Pakistan Journal of Medical Sciences*;33(4), 866–870. <https://doi.org/10.12669/pjms.334.12887>
- [9] Kaur M, Singla S, Mahajan R. (2016) Item analysis of in use multiple choice questions in pharmacology. *Int J Appl Basic Med Res*: 6(3):170-3. doi: 10.4103/2229-516X.186965
- [10] Abdel-Hameed AA, Al-Faris EA, Alorainy IA, Al-Rukban MO (2005). The criteria and analysis of good multiple choice questions in a health professional setting. *Saudi Med J*: 26(10):1505-10.
- [11] Laveault and Grégoire (2017). Introduction to the Theories of Tests in Psychology and Sciences of Education. (3rd Edition). Brussels: De Boeck.
- [12] Puthiaparampil T., Rahman M., Fong Lim I (2017). From Item Analysis to Assessment Analysis: Introducing New Formulae. *MedEdPublish*. 6, [1], 7, <https://doi.org/10.15694/mep.2017.000007>

- [13] George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: Allyn & Bacon.
- [14] Ben Salah, N., Salouage, I., Hatira, Z. Z., Goucha, R., Meherzi, A., & Kallel, K (2019). Multidisciplinary exams in medical studies: Interest of docimologic analysis. *La Tunisie Medicale*; 97 (1), 93–99.
- [15] Hingorjo, M. R., & Jaleel, F (2012). Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *JPMA. The Journal of the Pakistan Medical Association*;62 (2), 142–147.
- [16] Patil, R., Palve, S. B., Vell, K., & Boratne, A. V (2016). Evaluation of multiple-choice questions by item analysis in a medical college at Pondicherry, India. *International Journal Of Community Medicine And Public Health*; 3,(6). <http://dx.doi.org/10.18203/2394-6040.ijcmph20161638>
- [17] Sakly, N., Massaoudi, S., Mastouri, M., Azaiez, R (2019). Analyse docimologique des épreuves du concours de résidanat en pharmacie en Tunisie. Available at : <http://www.fphm.rnu.tn/sites/default/files/Annexe%2074%20Analyse%20docimologique%20concours%20r%C3%A9sidanat.pdf>
- [18] Hermi, A., & Achour, W (2015). Difficulty, discrimination and cognitive level of Microbiology exam questions of the Faculty of Medicine of Tunisia. *La Tunisie Medicale*; 93(8–9), 487–490.
- [19] Hermi, A., & Achour, W (2016). Item analysis of examinations in the Faculty of Medicine of Tunis. *La Tunisie Medicale*; 94 (4), 247–252.
- [20] Pais, J., Silva, A., Guimarães, B., Povo, A., Coelho, E., Silva-Pereira, F., Lourinho, I., Ferreira, M. A., & Severo, M (2016). Do item-writing flaws reduce examinations psychometric quality? *BMC Research Notes*; 9(1), 399. <https://doi.org/10.1186/s13104-016-2202-4>
- [21] Kowash M, Alhobeira H, Hussein I, Al Halabi M, Khan S (2020). Knowledge of dental faculty in gulf cooperation council states of multiple-choice questions' item writing flaws. *Med Educ Online*;25(1):1812224. doi:10.1080/10872981.2020.1812224.
- [22] Abdulghani, H. M., Irshad, M., Haque, S., Ahmad, T., Sattar, K., & Khalil, M. S (2017). Effectiveness of longitudinal faculty development programs on MCQs items writing skills: A follow-up study. *PloS One*; 12(10), e0185895. <https://doi.org/10.1371/journal.pone.0185895>
- [23] Shaikh S, Kannan SK, Naqvi ZA, Pasha Z, Ahamad M (2020). The Role of Faculty Development in Improving the Quality of Multiple-Choice Questions in Dental Education. *J Dent Educ*; 84(3):316-322. doi: 10.21815/JDE.019.189.
- [24] Abdulghani HM, Ahmad F, Irshad M, Khalil MS, Al-Shaikh GK (2015). Faculty development programs improve the quality of multiple-choice questions items' writing: 5: 9556. DOI: 10.1038/srep09556
- [25] Naeem N, van der Vleuten C, Alfaris EA (2012). Faculty development on item writing substantially improves item quality. *Adv Health Sci Educ Theory Pract*: 17:369–376. DOI: 10.1007/s10459-011-9315-2
- [26] Gupta P, Meena P, Khan AM, Malhotra RK, Singh T. Effect of Faculty Training on Quality of Multiple-Choice Questions (2020). *Int J Appl Basic Med Res*: 10(3):210-214. doi: 10.4103/ijabmr.IJABMR\_30\_20.
- [27] Sajjad M, Iltaf S, Khan RA. Nonfunctional distractor analysis: An indicator for quality of Multiple-choice questions (2020). *Pak J Med Sci*: 36(5):982-986. doi: 10.12669/pjms.36.5.2439.
- [28] 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, 40. <https://doi.org/10.1186/1472-6920-9-40>
- [29] Testa S, Toscano A, Rosato R (2018). Distractor efficiency in an item pool for a statistics classroom exam: Assessing its relation with item cognitive level classified according to Bloom's taxonomy. *Front Psychol*;9:1-12. doi:10.3389/fpsyg.2018.01585
- [30] Hamamoto Filho PT, Silva E, Ribeiro ZMT, Hafner MLMB, Cecilio-Fernandes D, Bicudo AM (2020). Relationships between Bloom's taxonomy, judges' estimation of item difficulty and psychometric properties of items from a progress test: a prospective observational study. *São Paulo Med J*;138(1):33–39.
- [31] Shad, M. N., Fatima, A., Fatima, S., & Chiragh, S (2018). Item analysis of MCQs of a pharmacology term exam in a private medical college of Pakistan. *Pakistan Journal of Medical and Health Sciences*; 12, 700–703.

Table III: item distribution according to levels

Parameters	Academic years											
	2016-2017				2017-2018				2018-2019			
Levels		1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>		1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>		1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>
Students *	90	41	21	28	94	34	39	21	101	31	30	40
<b>Total questions</b>	402	144	87	143	410	140	121	149	555	237	135	183
SAQ n (%)	192 (47,76)	55	39	98	201 (49,0)	67	46	88	267 (48.1)	74	70	123
LEQ n (%)	58 (14,42)	17	25	16	78 (19,02)	31	27	20	45 (8.1)	25	15	5
Clinical case studies n (%)	15 (3,8)	8	0	7	4 (0,97)	0	1	3	8 (1.44)	2	3	3
MCQ n (%)	137 (34,1)	64	23	50	127 (30,97)	42	47	38	231 (41.62)	136	47	48
<b>Difficulty index (mean (SD))</b>	0,64 (0,09)	0,58(0,22)	0,67(0,21)	0,66 (0,06)	0,61 (0,13)	0,56(0,16)	0,63(0,1)	0,65(0,13)	0,64 (0,1)	0,64(0,11)	0,61(0,10)	0,67(0,08)
Difficult n (%)	30 (7,46)	18	5	7	27 (6,58)	13	4	10	29 (5,22)	14	9	6
Moderate n (%)	191 (47,51)	70	35	86	229 (55,85)	90	72	67	257 (46.3)	118	66	73
Easy n (%)	181 (45,03)	58	47	76	154 (37,57)	37	45	72	269 (48.46)	105	60	104
Discrimination Index (mean (SD))	0,32 (0,08)	0,34(0,21)	0,33(0,10)	0,27(0,06)	0,38 (0,15)	0,34(0,09)	0,36(0,2)	0,44(0,12)	0,36 (0,24)	0,34(0,09)	0,34(0,09)	0,35(0,06)
Test reliability (mean (SD))	0,47 (0,34)	0,66(0,17)	0,37(0,51)	0,39 (0,2)	0,43 (0,44)	0,27 (0,6)	0,45 (0,31)	0,56 (0,34)	0,34 (0,9)	0,53(0,26)	0,61(0,21)	0,4(0,7)
Passing rate %	82,43	67,43	91,13	88,39	78,25	72,16	85,27	76,74	81,55	81,68	78,77	86,90

\*(the total seems to be more than 155 because some students were counted at the first level in 2017 and at the 2<sup>nd</sup> level in 2018 as well as at the 3<sup>rd</sup> level in 2019)