

# Item Analysis of Multiple Choice Questions in Medical Licensing Examination

**Alimaa Gomboo<sup>1</sup>, Bayarbat Gombo<sup>1</sup>, Tumennast Munkhgerel<sup>1</sup>, Sumberzul Nyamjav<sup>1</sup>, Oyungoo Badamdorj<sup>2</sup>**

<sup>1</sup>Department of Medical Education Academic Affairs, Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia; <sup>2</sup>Division of Educational Policy and Management; Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia

**Submitted:** December 10, 2018

**Revised:** March 26, 2019

**Accepted:** June 10, 2019

## Corresponding Author

Oyungoo Badamdorj MD, PhD, Assoc Prof

Division of Educational Policy and Management, Mongolian National University of Medical Sciences, Ulaanbaatar 14210, Mongolia

**Tel:** +976-11-312059

**E-mail:** oyungoo@mnums.edu.mn

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/bync/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. Copyright© 2019 Mongolian National University of Medical Sciences

**Objectives:** To evaluate multiple-choice questions of the Medical Licensing Examination.

**Methods:** The study design was a cross sectional descriptive and item analysis method was used. This study was conducted in Health Development Center of the Ministry of Health. An item analysis was done on 1500 multiple-choice questions for the internal medicine licensing exam and of 2077 subjects from 2014 to 2016. The item analysis involved identifying reliability coefficient, difficulty index, and discrimination index for each question in relation to the doctors' exam performance. Microsoft Excel 2010, and SPSS were used for the statistical analyses.

**Results:** Reliability coefficient (KR20) of Mongolia's medical licensing examination in 2014-2016 ranged from 0.79 - 0.92. The frequency of multiple-choice questions with a negative or zero discrimination index was 11.6 percent and a high difficulty index was 21.9 percent. **Conclusion:** Even though reliability of Mongolia's medical licensing examinations were good, one-third of the multiple-choice questions used in the exams fail to evaluate examinee's knowledge.

**Keywords:** Reliability Coefficient, Discrimination Index, Difficulty Index, Medical Licensing Examination, Mongolia

## Introduction

In 1996, Mongolian government passed health legislation focusing on improving the quality of health services by requiring medical licenses to qualified medical professionals in their action plan. In 1997, health legislation required medical professionals to obtain an official license to practice nursing, pharmacy, midwifery, and rehabilitation. These new laws started licensing medical procedures in 1999<sup>1</sup>.

Every year roughly 2500 graduates matriculate from health sciences universities in Mongolia and 20-30% of the graduates obtain their Bachelor of Medicine degree to become general physicians<sup>2</sup>.

A medical license is an occupational license that permits a person to legally practice medicine according to professional and educational standards. Licensing is essential for improving the skills and performance of the graduates which directly influences the quality of their patient's care and safety<sup>3</sup>.

State and private medical schools have developed curricula based on requirements of Mongolia's Higher Educational Standard and the knowledge base to required to pass the medical licensure examination, with the health sector putting the same knowledge requirements in front of all graduates.

Assessment of health workers as students and professionals has a profound impact on their learning and is an essential safety valve before certification. It is used for their training, their placement, their certification, and their promotion. The multiple-choice question (MCQ) type of tests represents one of the most important examination tools that is commonly used in this assessment<sup>4</sup>.

Today the knowledge of the graduates is rated by their final exam, which is organized by the school's internal standards and also on the educational objectives of the Center for Health Development.

The medical schools select the tests from their own test pools, and the type of exams are similar. The experts monitor the tests developed by departments at the university level. Professional board members develop the pool of the MCQs for the licensing exam under the Ministry of Health and the members consist of various specialists from institutions, professional agencies, NGOs, physicians from health organizations, faculties of universities.

Methodology to assess the performance of such tests is available. An item analysis is typically used to examine the response of the students to individual test questions and to assess the quality of those questions and their contribution to the test as a whole. An item analysis was done on the Mongolian National University Medical School matriculation knowledge examination by Professor Sumberzul Nyamjav in 2003. Such reviews not only help improving questions which can be used again in later tests but also can be used to eliminate ambiguous or misleading items in a single test administration<sup>5</sup>. These analyses allow the examiners to categorize questions into categories that include valid and robust questions that discriminate between those applicants who have the desired knowledge base from those who do not, so that the items can be improved, or excluded. These analyses are essential in the development and maintenance of stringent standardized tests of professional licensure and are necessary to ensure that the tests are comparable year after year and produce results that are fair to the examinees. However, the methodology to examine test performance has not been widely used in Mongolia and to

our knowledge has not previously been applied to the medical licensure examination in Mongolia.

The purpose of this study is to examine the reliability, difficulty, and discrimination ability of the questions for Mongolian medical licensure examination based on exam taker's item results.

## Materials and Methods

### Data collection

A cross-sectional descriptive study was carried out of the 1500 MCQs used for the medical licensing exams in Mongolia between 2014-2016. Thirteen versions of the Mongolia medical licensing exam were used during the study period, and questions for these exams were taken from different but somewhat overlapping question pools each year. The answers for each item on each test were gathered from examinees who graduated from Mongolian National University of Medical Sciences (public), "Ach" Medical University (private), and "Etugen" University (private) medical schools. The correct answers were provided by the Health Development Center of the Ministry of Health, who approved this study and allowed access to the examination data.

## Methods

To quantitatively evaluate each exam question, we determined the following for each exam:

1. Reliability of the exam as a group of questions.
2. Difficulty index of each MCQ
3. Discrimination index of each MCQ.

### Reliability:

There are several statistical methods to quantify the reliability of a test. Some of the more common measures used in studies are Cronbach's Alpha, KR20 (Kuder-Richardson 20), and KR21 (Kuder-Richardson 21)<sup>6,7</sup>. We used the Kuder-Richardson 20 method for analyzing reliability. The Kuder-Richardson 20 is calculated as follows<sup>8</sup>:

$$KR20 = \frac{K}{K-1} \left[ 1 - \frac{\bar{x} \sum_{i=1}^K p_i q_i}{kS^2} \right]$$

Where:

KR20 is the reliability coefficient

K is the number of MCQs

$p_i$  is the proportion of correct responses to test item  $i$ .

$q_i$  is the proportion of incorrect responses to test item  $i$ .

$S^2$  is the variance of the total scores of all students that took the test.

KR20 values can range from zero to one. The closer the KR20 is to one, the more reliable an exam is considered because its questions consistently discriminate between higher and lower performing students.

The KR20 results are typically interpreted as follows:

.90 to 1 Excellent reliability; at the level of the best-standardized tests

.80 - .89 Very good for a classroom test

.70 - .79 Suitable for a classroom test; in the range of most tests.

There are probably a few items which could be improved.

.60 - .70 Somewhat low. This test needs to be supplemented by other measures (e.g., more tests or questions) to determine grades. There are probably some items which could be improved.

.50 - .60 Suggests a need for revision of the test unless it is quite short (ten or fewer items). The test needs to be supplemented by other measures (e.g., additional tests or questions) for grading.

0 - 0.50 Questionable reliability. This test should not contribute heavily to the course grade and needs revision<sup>9</sup>.

### Difficulty index

The difficulty index (P) is the parameter used to evaluate the difficulty of each question in a MCQ examination.

$$P = \frac{\text{Number of examinees with the correct answer}}{\text{Total number of examinees}} * 100$$

Since P is the percentage of examinees with the correct answer, it can range from 0 to 100. For MCQs, a test item is considered acceptable when P is between 60-65. When it is more than 90, an exam item should be regarded as too easy, and when it is less 30 it is too difficult<sup>10</sup>.

### Discrimination index

Item discrimination refers to the ability of an item to differentiate among students based on how well they know the material being tested. Item discrimination is calculated by ranking the

students according to total score and then selecting groups with the top quartile and the lowest quartile scores. For each item, the percentage of students in the top and bottom quartile answering the item correctly is calculated.

$$DI = \frac{P_u - P_L}{2\sqrt{pq}}$$

**DIDI** is the Discrimination index.  $P_u$  - Percentage of examinees in the upper quartile with the correct answer.

$P_L$  is the Percentage of examinees in the bottom quartile with the correct answer

$p$  is the Arithmetical average of  $P_u$  and  $P_L$

$q$  is the Arithmetical average of  $1 - P_u$  and  $1 - P_L$

DI can range from -1 to 1. A positive DI indicates that high scorers have a high probability of answering correctly, and low scorers have a low probability of answering correctly. A negative DI indicates that low scorers got the question correct or that examinees with high overall test scores did not get the item right. A result near or equal to zero means that the examinees that scored in the top and bottom quartiles answered the question correctly with the same frequency indicating the item was unable to distinguish between highest and lowest performers on the exam.

The proportion of students answering an item correctly indicates the level of difficulty of the item, and it is affected by the examinees' knowledge, as well as the clarity and relevance of the question. Practically, it is essential to use items with positive discrimination level when implementing MCQ in medical licensing examination<sup>10-16</sup>.

### Statistical analysis

Data were gathered by entering the answer for each item for the 2077 medical doctors who took a medical licensing exam over the 2014 to 2016 study period into MS Excel 2010. The reliability index (KR20), difficulty index, and discrimination index were calculated. Data analyzed with simple proportions, mean  $\pm$  standard deviations, 95% CI, and the chi-square test was used to compare differences in proportions. The linear relationship between DIF and DI was assessed using Pearson's correlation test. The statistical analyses were carried out using MS Excel 2010 and SPSS version 21, and the significance level was  $p \leq .05$ .

**Ethical statements**

Ethical approval and clearance was obtained from the Ethical Review Board of the Mongolian National University of Medical Sciences on April 16, 2013.

**Results**

We assessed MCQs of 2077 graduates from the Mongolian National University of Medical Sciences, "Ach" Medical University, "Etugen" University in 2014-2016 who took a medical licensing examination using item analysis.

The mean medical licensing examination score of all graduates from the three medical schools' in 2014 was 55.5 ± 8.9, 79.5 ± 13.6 in 2015, and 68.61 ± 12.5 CI in 2016. The variation in the scores is primarily due to different MCQs, the number of MCQ items used, and the number of versions of the medical licensing examination.

The reliability coefficient of Medical licensing examination in 2014 was .79, which is good, .84 in 2015 and .92 in 2016, which is considered very good (Table 2).

Using the discrimination index we identified that approximately 3.4% (N = 17) MCQ had a negative DI, 8.2% (N = 41) MCQ had a DI near to or equal to 0, and 88.4% (N = 442) MCQ had a positive DI, and 11.6% of the used MCQ were not able to discriminate high test performers from low test performers. Furthermore, the DI was different every year (p=.034) (Table 3).

The graduates from the Mongolian National University of Medical Sciences, "Ach" Medical University, "Etugen" University found that over the study period 88 (21.9%) of the medical licensing items were difficult, 684 (42.6%) were acceptable, and 198 (43.3%) were easy (Table 4). However, the proportion of difficult, fair, and easy questions varied significantly depending on which year the exam was taken (p=.0001).

Table 5 compares of DIF and DI of the MCQ items from 2014 to 2016. The overall difficulty index of the items ranged from 55.51 ± 8.8 in 2015 to 80.36 ± 13.6 the following year. The discrimination index of the items was more homogeneous, ranging from 0.24 ± 0.2 in 2014 and 2015 to 0.28 ± 0.21 in 2016.

**Table 1.** Descriptive results of Medical Licensing Examination Score by year and version

Years	Version*	Number of MCQ	Number of examinees	Mean ± S.D	95 % CI		Min score	Max score
					Lower	Upper		
2014	A	100	91	52.53 ± 8.2	50.9	54.1	38.0	75.0
	B	100	79	56.82 ± 10.6	54.7	58.9	27.0	84.0
	C	100	89	54.11 ± 9.1	52.3	55.9	37.0	100.0
	D	100	92	54.15 ± 8.7	52.4	55.9	29.0	72.0
	E	100	89	59.12 ± 8.2	57.5	60.7	38.0	88.0
	F	100	100	56.47 ± 8.4	54.8	58.1	38.0	84.0
	<b>Total</b>	<b>600</b>	<b>540</b>	<b>55.5 ± 8.9</b>	<b>53.8</b>	<b>57.3</b>	<b>34.5</b>	<b>83.8</b>
2015	A	150	185	86.48 ± 13.7	84.3	88.7	41.0	134.0
	B	150	158	84.59 ± 14.7	82.2	87.0	42.0	132.0
	C	150	122	68.04 ± 12.7	66.0	70.1	42.0	118.0
	D	150	186	78.76 ± 13.4	76.6	80.9	39.0	111.0
	<b>Total</b>	<b>600</b>	<b>651</b>	<b>79.5 ± 13.6</b>	<b>77.3</b>	<b>81.6</b>	<b>41.0</b>	<b>123.8</b>
2016		100	297	69.82±12.3	67.4	72.2	22.0	100.0
		100	272	69.41±12.1	67.0	71.8	28.0	90.0
		100	287	66.59±13.0	64.0	69.1	27.0	100.0
	<b>Total</b>	<b>300</b>	<b>856</b>	<b>68.61±12.5</b>	<b>66.2</b>	<b>71.1</b>	<b>25.7</b>	<b>96.7</b>

\*New versions of the exam were used each year, with a total 13 exams from 2014 to 2016.

CI - Confidence Interval; MCQ - Multiple Choice Questions

**Table 2.** Reliability Coefficient of MCQ items

Years	Number of MCQs items	Response of MCQs items	Reliability Coefficient
2014	600	5	.79
2015	600	5	.84
2016	300	5	.92
Mean of Reliability coefficients			.85

MCQ: Multiple Choice Questions; Kuder-Richardson 20 method was used to calculate reliability coefficient

**Table 3.** Discrimination Index (percentage) of MCQ items

Years	DI (percentage)			p-value*
	Negative	Zero	Positive	
2014	30 (5.0)	48 (8.0)	522 (87.0)	.034
2015	18 (3.0)	48 (8.0)	534 (89.0)	
2016	3 (1.0)	27 (9.0)	270 (90.0)	
Average	17 (3.4)	41 (8.2)	442 (88.4)	

DI – Discrimination Index; \*Chi-Square Test

**Table 4.** Difficulty Index of MCQ items.

Years	DIF (proportion)			p-value*
	Difficult	Acceptable	Easy	
2014	120 (20.0)	288 (47.6)	198 (32.7)	.0001
2015	120 (20.0)	312 (52.05)	204 (34.0)	
2016	24 (8.05)	84 (28.0)	192 (64.0)	
Average	88 (21.9)	684 (42.6)	198 (43.3)	

DIF- Difficulty Index Factor; \*Chi-Square Test

**Table 5.** Comparison of Difficulty Index and Discrimination Index of MCQ items.

Item analysis parameters	Years	Difficulty index	Discrimination index
Mean ± SD	2014	55.51 ± 8.8	0.24 ± 0.18
	2015	80.36 ± 13.6	0.24 ± 0.17
	2016	68.61 ± 12.5	0.28 ± 0.21
Range	2014	27 - 100	-0.24 – 0.73
	2015	39 - 134	-0.29 – 0.63
	2016	22 - 100	-0.27 – 0.70

SD - Standard Deviation

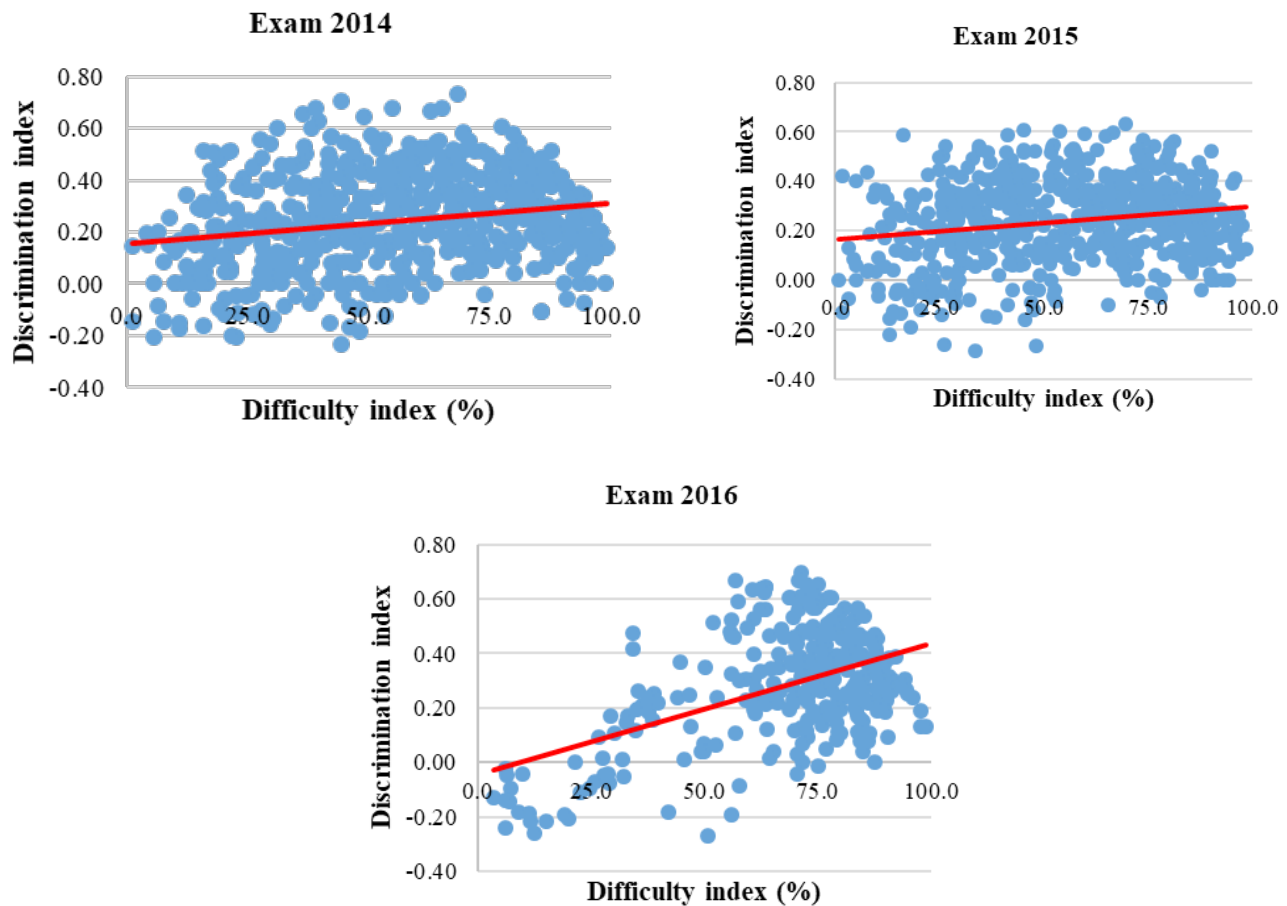


Figure 1. Pearson Correlation of Difficulty and Discrimination index.

Figure 1 shows the relationship between the difficulty index and the discrimination index for each of the MCQs used during the 3-year study period. There was a weak correlation between the discrimination index and the difficulty index in 2014 ( $r = .2189, p < .0001$ ), and in 2015 ( $r = .1965, p < .0001$ ). However, there was a moderate correlation between the two in 2016 ( $r = .50, p < .0001$ ) indicating that fewer easy questions were used that year.

## Discussion

Our results show that the overall validity of the medical school licensure exam varied in recent years, from 0.79 (satisfactory) in 2014, 0.84 (good) in 2015, to 0.92 (excellent) in 2016. Approximately 11.6 percent of the test questions were had negative or near to zero validity scores indicating they were not valid questions, while 21.9 percent were extremely difficult items that could not assess fairly students' knowledge. The Division

for Educational Policy and Management of Mongolian National University of Medical Sciences has performed internal monitoring of its final exam every year to improve the quality of the exam. As part of this improvement effort, recommendations to upgrade or to exclude some questions were given to the medical licensing examination organizers. The average of the reliability is now excellent and is now above 0.90.

The final exam result of the "Ach" medical school during the period of time 2015-2016 had a reliability of 0.94-0.96, and it was identified that 3% of the test questions were too difficult, and 4% were too easy. It brought to our attention that their final exam reliability was excellent, and the percentage of questions that were too difficult and too easy test were minimal.

The reliability results of the medical licensing exams of 5 medical schools in Switzerland was 0.91<sup>17</sup>. This result was higher than Mongolia's exams' result.

Since 2012, The Korean government started to share professional licensing examination items, including the results

of exam takers between 2009 and 2014 (n=20,455) from 41 medical schools that covered 5548 items. Passing success rates were 93.60, 97.02, 94.44, 96.91, 96.22, and 96.65 in 2009 through 2014, respectively due to the similarity of discrimination indices of the items<sup>18</sup>.

From 2014-2016, the reliability of the exam increased (0.79-0.92), the proportion of DI of negative and near zero test items decreased (13-10 percent), and the number of difficult questions (20.0-8.05) decreased. This may be due to the corrections and improvements in their medical licensing examinations.

The discrimination index of medical licensing examination in Mongolia between 2014-2016 correlated positively with the difficulty index ( $p < .0001$ ), which was statistically significant. In a similar type of study reported by Anjali et al., the discrimination index did not correlate with the difficulty index ( $r = .31$ ,  $p = .053$ )<sup>15</sup>. The difficulty and discrimination indices are often reciprocally related. However, this may not always be true. Questions having a high P (easier questions) discriminate poorly; conversely, questions with a lower P (harder questions) are considered to be good discriminators<sup>19</sup>.

Based on our item analysis, we need to improve the MCQ that have difficulty indexes that are too high or too low and questions with negative or near zero discriminatory indices that are failing to evaluate examinee's knowledge.

The medical licensing examination was revamped in 2014 with WHO's financial support and used for new graduates, and it affected the examination's assessment.

Throughout the world, competency-based educational assessments are currently the most acknowledged methodology; yet still, many challenges exist for both highly developed and developing countries<sup>20</sup>. Therefore, traditional MCQ based knowledge examinations remain the most widely used valid student and graduate assessment tool in the world<sup>21</sup>.

Furthermore, current review articles demonstrate the development directions in global medical licensing examination practices; the outcome-based workforce planning is becoming the primary arms of this professional assessment<sup>22</sup>. This study has laid the foundation for medical licensing examination validation in Mongolia. However, more modern approaches that implement sophisticated data mining and artificial intelligence technologies in item analysis that are time-saving and more efficient are being used in the developed world, particularly in the United States<sup>23</sup>. Using in-depth, content analysis of professional criteria and the

item analysis; officials in the United States have analyzed over 550000 exam materials to align exam questions with workforce tasks<sup>24</sup>.

Globally, skills and performance assessments are an inseparable part of medical licensing examinations. In Germany, Schickler et al. have demonstrated how practical examinations results are correlated with traditional knowledge assessments as well as oral examinations<sup>25</sup>. However, in Mongolia, skill and performance exam are not included in the medical licensing examination and remain a challenge in medical education of Mongolia.

Our most significant limitation was the absence of comparative analytic methods to assess test performance, and we plan to include a comparative analysis of the medical school examination passing rates with the medical school exit knowledge examination to identify associated factors.

## Conclusion

The reliability of medical licensing examinations was quite high, but almost one-third of MCQs used in the exam were unable to evaluate the examinees' knowledge. In other words, the frequency of MCQs with negative or near zero discrimination index and the number of difficult questions were high.

## Conflict of Interest

The authors state no conflict of interest.

## Acknowledgement

The lead author expresses gratitude to the Center of Health Development and Mongolian National University of Medical Sciences for the golden opportunity to do this research study.

## References

1. Ministry of health. Department of Accreditation and Medical license 1999-2002 III. National center for health development. Ulaanbaatar, Mongolia: Munkhiin Useg; 2002. p 6-7.
2. Center for Health Development. Annual report of National Center for Health Development 2014-2016. Ulaanbaatar,



- Mongolia: Munkhiin Useg; 2016. p 45.
3. Archer J, Lynn N, Roberts M, Coombes L, Gale T, Bere SR. A Systematic review on the impact of licensing examinations for doctors in countries comparable to the UK: Final Report. General Medical Council 2015.
  4. Abdel-Hameed AA, Al-Faris EA, Alorainy IA, Al-Rukban MO. The criteria and analysis of good multiple choice questions in a health professional setting. *Saudi Med J* 2005; 26: 1505-10.
  5. Sahoo DP, Singh R. Item and distracter analysis of multiple choice questions (MCQs) from a preliminary examination of undergraduate medical students. *IJRMS* 2017; 5: 5351.
  6. The technical whitepaper series. KR20 & Coefficient Alpha Their equivalence for binary scored items 2007 [accessed on 15 March 2015]. Available at: <http://www.pbarrett.net/techpapers/kr20.pdf.p2-4>.
  7. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika* 1951; 16: 297-334. doi: 10.1007/BF02310555.
  8. Real statistics. Kuder–Richardson Formula 20 [accessed on 15 September 2018]. Available at: [https://en.wikipedi.org/wiki/Kuder%E2%80%93Richardson\\_Formula\\_20](https://en.wikipedi.org/wiki/Kuder%E2%80%93Richardson_Formula_20).
  9. Understanding Item Analyses.2019; Available at: <http://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis/>.
  10. Sumberzul N. Oyunbileg SH. Medical education. Ulaanbaatar, Mongolia: Admon; 2001. p 100-104.
  11. Oyungoo B. Some aspects of clinical competence evaluation development of medical students [dissertation]. Ulaanbaatar, Mongolia: Mongolian National University of Medical Sciences; 2010.
  12. Sumberzul N. Development of knowledge evaluation of medical students [dissertation]. Ulaanbaatar, Mongolia: Mongolian National University of Medical Sciences; 2003.
  13. Shete A, Kausar A, Lakhkar K, Khan S. Item analysis: An evaluation of multiple choice questions in Physiology examination. *JCME* 2015; 3: 106-9. doi: 10.5455/jcme.20151011041414.
  14. Rao C. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. *IJEPR* 2016; 4: 201-4.
  15. The University of Texas at Austin. MC-item-analysis-handout 2003 [accessed on 15 March 2015]. Available at: [www.utexas.edu/academic/diia](http://www.utexas.edu/academic/diia).
  16. Angale AS, Giri PA, Doibale MK. Study on item and test analysis of multiple choice questions amongst undergraduate medical students. *IJCMPh* 2017; 4: 1562.
  17. Guttormsen S, Beyeler C, Bonvin R, Feller S, Schirlo C, Schnabel K, et al. The new licencing examination for human medicine: from concept to implementation. *Swiss Med Wkly* 2013; 143:w13897. doi: <https://doi.org/10.4414/smw.2013.13897>.
  18. Yang EB, Lee MA, Park YS. Effects of test item disclosure on medical licensing examination. *Adv Health Sci Educ Theory Practice* 2018; 23: 265-74. doi: 10.1007/s10459-017-9788-8.
  19. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *J Pak Med Assoc* 2012; 62: 142-7.
  20. Huber-lang M, Palmer A, Grab C, Boeckers A, Boeckers TM, Oechsner W. Visions and reality: the idea of competence-oriented assessment for German medical students is not yet realised in licensing examinations. *GMS J Med Educ* 2017; 34: Doc25.
  21. Swanson DB, Roberts TE. Trends in national licensing examinations in medicine. *Med Educ* 2016; 50: 101-14.
  22. Price T, Lynn N, Coombes L, Roberts M, Gale T, de Bere SR, et al. The International Landscape of Medical Licensing Examinations: A Typology Derived From a Systematic Review. *Int J Health Policy Manag* 2018; 7: 782-90.
  23. Robson B, Boray S. Data-mining to build a knowledge representation store for clinical decision support. Studies on curation and validation based on machine performance in multiple choice medical licensing examinations. *Comput Biol Med* 2016; 73: 71-93.
  24. Haist SA, Butler AP, Paniagua MA. Testing and evaluation: the present and future of the assessment of medical professionals. *Adv Physiol Educ* 2017; 41: 149-53.
  25. Schickler A, Brüstle P, Biller S. The Final Oral/Practical State Examination at Freiburg Medical Faculty in 2012--Analysis of grading to test quality assurance. *GMS Z Med Ausbild* 2015; 32: Doc39.