ORIGINAL ARTICLE

Item Analysis of Single Best Response Type Multiple Choice Questions for Formative Assessment in Obstetrics and Gynaecology

Shabdika Kulshreshtha[1] · Ganesh Gupta[2] · Gourav Goyal[3] · Kalika Gupta[4] · Kush Davda[5]

Abstract

Background MCQs (multiple choice questions) can be used to assess higher-order skills and be utilised for creating a question bank.

Purpose of Study (Aim) To perform post-validation item analysis of MCQs constructed by medical faculty for formative assessment of final-year medical students and to explore the association between difficulty index (p value) and discrimination indices (DI) with distractor efficiency (DE).

Methods An observational study was conducted involving 50 final-year MBBS students and 10 faculty members for a period of one year (October 2020 to September 2021). After training the faculty in item analysis, a MCQ test was prepared after collecting the peer-reviewed 25 MCQs (five each from various subtopics of the subject). The result of this test was used to calculate the FV (facility value), DI (discrimination index), and DE (distractor efficiency) of each item. Student and faculty feedback was also obtained on a five-point Likert scale.

Results The mean FV was 46.3 ± 19.3 and 64% of questions were difficult; the mean DI was 0.3 ± 0.1 and 92% of questions could differentiate between HAG (high achiever's group) and LAG (low achiever's group); the mean DE was $82\% \pm 19.8$ and 48% of items had no NFDs (non-functional distractors).

Conclusion MCQs can be used to assess all the levels of Bloom's taxonomy. Item analysis yielded 23 out of 25 MCQs that were suitable for the question bank.

Keywords Difficulty index · Distractor efficiency · Item analysis · MCQs (multiple choice questions) · Non-functional distractor

Introduction

The millennial years have shown that multiple choice questions are one of the best ways of assessing higher levels of cognitive domain. For most comprehensive, licencing, and screening tests, MCQs are used as a versatile tool to gauge the competencies of a medical student. If constructed appropriately, they can assess higher cognitive processing of Bloom's taxonomy like interpretation, synthesis, and knowledge application rather than merely testing recall of facts [1]. In the dynamic field of medical education, it is indeed cumbersome and time-consuming to frame MCQs as compared to descriptive questions [2].

Shabdika Kulshreshtha is Professor, Ganesh Gupta is an Associate Professor, Gourav Goyal is an Associate Professor, Kalika Gupta is an Associate Professor, and Kush Davda is a Demonstrator.

✉ Kalika Gupta
  kalikagupta007@gmail.com

1 Department of Obstetrics and Gynaecology, Geetanjali Medical College and Hospital, Udaipur, Rajasthan, India

2 Department of Anaesthesiology, Pacific Medical College and Hospital, Bhilo Ka Bedla, Udaipur, Rajasthan, India

3 Department of Paediatrics, Ananta Institute of Medical Sciences and Research Centre, Rajasthan University of Health Sciences, Rajsamand, Rajasthan, India

4 Department of Community Medicine, Ananta Institute of Medical Sciences and Research Centre, Rajasthan University of Health Sciences, B5, Doctors Quarters, Ananta Institute of Medical Sciences and Research Centre, Rajsamand, Rajasthan 313202, India

5 Department of Forensic Medicine, Ananta Institute of Medical Sciences and Research Centre, Rajasthan University of Health Sciences, Rajsamand, Rajasthan, India

The most commonly used type of MCQ is the single best response type, i.e. type A MCQ with four options [3]. The major problems with MCQs are difficulty in constructing plausible distractors, especially in assessing higher cognitive skills; ambiguity with more than one correct answer; score getting influenced by the reading ability of the student; probability of guessing the correct answer; and the inability of a MCQ to differentiate between high and low performers. Difficulty comprehending the reason for opting for an incorrect answer by the student and overinterpreting an MCQ (item) are the other problems encountered by the students. These problems are well noted in the MCQs framed in most medical undergraduate subjects, including Obstetrics and Gynaecology.

Although it is a fact that MCQs are not a preferred tool (for psychomotor and affective domains), if properly constructed, MCQs can easily overcome these above-mentioned flaws [4]. How each MCQ (item) functions as a level of difficulty and in identifying the spread of high and low performers is decided by item analysis [4]. This will help in meeting all learning outcomes, providing highly structured, well-designed tasks and meeting the uniform standards. Also validity, reliability, and educational impact are taken care of.

The MCQ item analysis consists of the difficulty index (DIF I) (percentage of students that correctly answered the item), discrimination index (DI) (distinguishes between high achievers and nonachievers), distractor effectiveness (DE) (whether the items are well constructed), and internal consistency reliability (how well the items are correlated to one another). By items, one means questions, statements, or scenarios that are used as an assessment instrument. Each item is evaluated for these indices because if an item is flawed, it can become a distractor and the assessment can fail [5]. Item analysis is a relatively simple and valuable procedure that provides a method for analysing observation, the interpretation of the knowledge achieved by the students and information regarding the quality of test items. In this study, we have performed item analysis of single best response type MCQs, as they are seen as an efficient tool for assessing the student's level of academic learning.

The index study would provide a platform to change the way MCQs are selected in the formative assessment strategy as a part of undergraduate curriculum implementation. It would also help in the preparation of a standard question bank in Obstetrics and Gynaecology. The objectives of our study are to perform post-validation item analysis of MCQs constructed by medical faculty for formative assessment of final-year medical students and to explore the association between difficulty index (p value) and discrimination indices (DI) with distractor efficiency (DE). We also assessed their feedback on a 5-point Likert scale.

Methodology

This was a Prospective and Analytical study carried out at the Department of Obstetrics and Gynaecology in the year 2021, from January to December. The study was carried out involving the first 50 students of the final year of M.B.B.S. who gave their consent to take the MCQ test at the semester's end. It also involved all the faculty members of the Department of Obstetrics and Gynaecology, Ananta Medical College & Research Centre, Rajsamand, Rajasthan.

Item Analysis: [6]

The result of the student's performance in the formative assessment was used to find out FV (facility value), DI(discrimination index), and DE(distractor efficiency).

Facility Value (FV) or Difficulty Index or Facility Index:

$$\text{FV} = \text{HAG (High achievers)} + \text{LAG (Low achievers)}/$$
$$\text{N (Total students in two groups) represented in \%}$$

HAG = First 30% scorers;

LAG = Last 30% scorers (they were classified as per their performance in the MCQ test).

FV ranges from 0 to 100; FV > / =85% means easy question; 51−84% moderate; / =50 hard.

Discrimination Index (DI):

$$\text{DI} = 2 * (\text{HAG} - \text{LAG})/N$$

Its maximum value is 1.0. Its value > / = 0.35 is considered good, while < 0.2 is unacceptable, and in between is intermediate, depending on the type and intention of the test.

Distractor Efficiency

Those incorrect options in the MCQ are distractors. A poor distractor (NFD or non-functional distractor) is the one that is not even picked by 5% of the students, while a good one (Functional Distractor, FD) is selected by 5% or more students. This suggests the distractors are plausible and not dummies [7].

Validation

After the item analysis presentation, feedback was taken from the involved faculties. All the MCQs submitted by faculties

were peer reviewed before the test was given to the undergraduate students. All the students were given a five-point Likert questionnaire as feedback on the formative assessment. The feedback questionnaire consisted of a 5-point Likert scale.

The following phases were observed while conducting the study:

Phase 1: After obtaining due permission from the Principal and Controller and ethical approval from the IEC, the MEU department was informed of the seminar followed by a group activity. A written circular was dispatched through HOD (Department of Obstetrics and Gynaecology) to all faculties in the department of Obstetrics and Gynaecology.

A sensitization seminar was conducted in two sessions: the first covered the theory aspect of designing an MCQ along with item analysis, and the second half consisted of a group activity for setting an MCQ and applying item analysis with examples.

Phase 2: A total of 25 type A MCQs from topics already covered in previous classes (taking into account 5 topics with 5 questions from each topic) were collected. The questions were prepared after a pooling of peer-reviewed (by Professors and Associate professors) 25 MCQs from junior faculty members (Senior residents and Assistant professors). All MCQs were a combination of recall, image-based, and case-based questions. Every type A MCQ consisted of a stem and four options. All 50 students in the final year of M.B.B.S. had to select the best answer out of four choices. Each correct answer was given 1 mark, and there was no negative marking in this test. The duration of this assessment was 60 min. The result of the student's performance was used to determine the level of difficulty and power of discrimination using Microsoft Office Excel. Based on the marks obtained, students were divided into 3 groups: high achievers (top 33%), mid achievers (middle 33%), and low achievers (bottom 33%). After the assessment, feedback was obtained from students and faculty in the form of a five-point Likert questionnaire.

Phase 3: Applying item analysis for post-validation of the MCQ question paper.

Phase 4: To prepare a question bank and resource material after applying item analysis.

Statistical Analysis

The indices were calculated using the formulae referred to in Methods. All the values have been expressed as the mean ± SD of the total number of items. The correlation at the 0.01 level was considered significant. Analysis was done using IBM SPSS 23.0 software.

Results

The OBGYN MCQ test, consisting of 25 single best response MCQs, was taken by 50 final-year M.B.B.S. students. Their mean score was 11.58 ± 4.18 (maximum marks: 25). The mean score in the two groups, i.e. LAG and HAG, was 8.44 ± 2.22 and 14.72 ± 3.21, and the difference was statistically significant ($p = 0.001$) (Table 1). The highest score was 22, and the lowest was 6. The respective values of FV, DI, and DE for all 25 MCQs are given in (Fig. 1). The mean value of FV is $46.3\% \pm 19.4\%$, which indicates that the test paper was moderately difficult. There were no questions that were easy (FV > 85%), while 36% of questions were moderately (FV between 51 and 84%) difficult. Sixty-four per cent of the questions were hard (FV < 50%) for the students. (Fig. 2).

The mean DI of the test was 0.3 ± 0.1, which is in the acceptable range of the discrimination index. Out of 25 MCQs (items), two items (8%) had DI < 0.2, which is unacceptable. The remaining 23 items (92%) could discriminate between HAG and LAG. Nine (36%) MCQs had excellent DI, as shown in (Fig. 3). Though items 4, 10, 13, 17, and 18 were too easy, they had acceptable DI to differentiate between HAG and LAG. (Fig. 1).

Out of 25 items, there were 75 distractions. The mean DE was $82 \pm 19.8\%$, which is quite good. Forty-eight per cent of the items had functional distractors (Fig. 4). Only 20% of items had two NFDs (DE 50%).

Feedback was taken from students and faculty on item analysis and the quality of MCQ questions. Questions included whether they were confident item analysis would be useful, how interesting the whole exercise was, how much effort was involved in making MCQs, if doing item
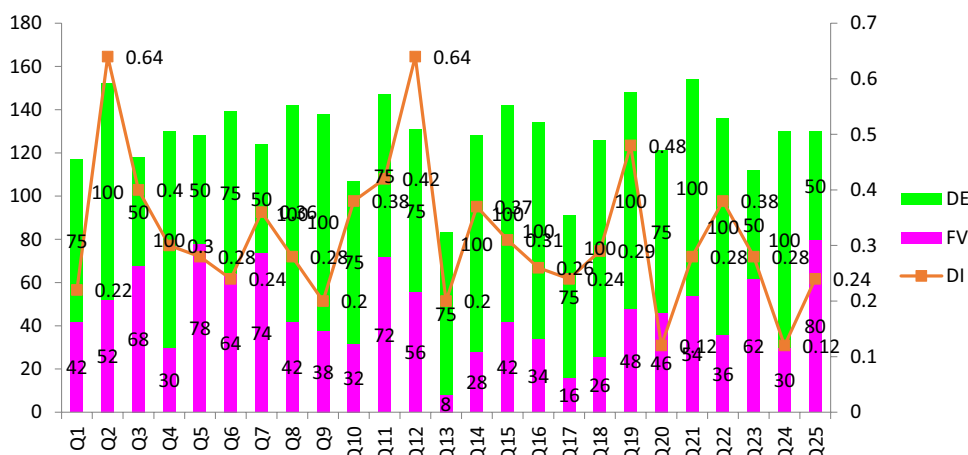
Table 1 Distribution of mean score of HAG, LAG, FV, DI & DE

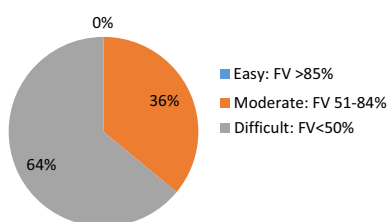| Parameters*	Mean ± SD
HAG	14.72 ± 3.21
LAG	8.44 ± 2.22
FV	46.3 ± 19.4
DI	0.3 ± 0.1
DE	82 ± 19.8

Correlation coefficient between DI and DE was found to be 0.05 ($p > 0.01$). Negative strong correlation was observed between FV and DE − 0.61 ($p < 0.01$). FV correlated positively with DI, $r = 0.23$ ($p > 0.01$)

*HAG, High Achieving group; LAG, low Achieving Group; FV, Facility Value; DI, Discrimination Index; DE, Distractor Efficiency

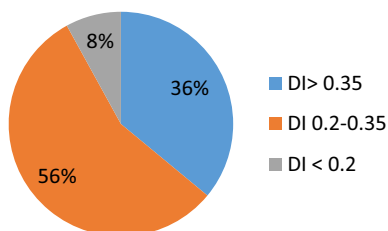**Fig. 1** Distribution of scores of all the students according to the MCQs

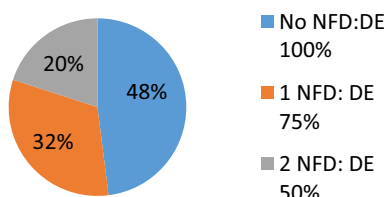**Fig. 2** Facility value and distribution

Facility value [FV] and distribution
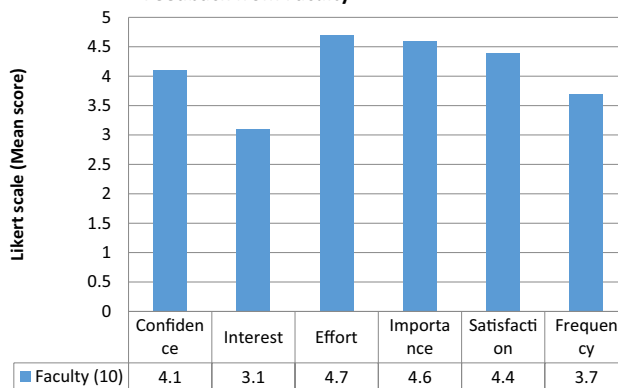
Fig. 3 Distribution of discrimination index [DI] in the study

Distribution of DI in the study

Fig. 4 Distribution of NFDs [non-functional distractors] with the distractor efficiency [DE]

Distribution of NFDs in the study MCQs

Feedback from Faculty

Fig. 5 Feedback on 5-point Likert scale

analysis is important, if this whole exercise was satisfying, and how frequently such an activity should be done by faculty. According to the faculty, 4.7/5 efforts were required for constructing valid MCQs, and they found this whole exercise less interesting (3.1/5) (Fig. 5).

Discussion

Framing MCQs has always been challenging, as multiple parameters are supposed to be kept in mind. The majority of high-stakes assessments follow this pattern. Many studies have been conducted on item analysis of MCQs so that they become more valid, reliable, and have a measurable educational impact.

Our question paper had a mean FV score of 46, which means it was relatively difficult. In our study, 23 (92%) MCQs had acceptable to excellent Discriminating power, and 2 items had unacceptable DI. This means those 2 items need to be revised; the rest were good at discriminating between HAG and LAG students. The DI values of the

present study are comparable with studies on item analysis by Date et al. [8] and others [9], as similar findings with 78% of items having acceptable to excellent discriminating power (DI > 0.20) and 24% having poor discriminating power (DI ≤ 0.20) were reported. Sometimes DI can be negative, i.e. low achievers answer a particular item more correctly than high achievers, as reported in some studies [10, 11]. The reasons for negative DI can be the wrong key, ambiguous framing of questions, or generalised poor preparation of students. Items with negative DI decrease the validity of the test and should be removed from the collection of questions. Three items that are either too easy or too difficult have poor DI. In item analysis, FV and DI should be interpreted together.

Distractor analysis tells whether the distractors used in items are functional or non-functional. We have a mean DE of 82%, which indicates that around half of the items had functional distractors. More NFDs make an item easier and decrease DE. In a similar study by Garg et al. [12] among medical students in Delhi, the mean DI was 0.3 ± 0.17 and the mean DE was 63.4 ± 33.3. The DI here is similar to that of the index study, but the DE is much lower, indicating the presence of better distractors in the items of our study.

Pande et al. [13], Shete et al. [14], and Karelia et al. [15] showed the difficulty index correlated positively with the discrimination index, which was not significant statistically. These studies replicated the findings of our study. Sim and Rasiah et al. [16] and Mitra et al. [17] studies showed a poor correlation between difficulty index and discrimination index. Similar to ours, Khilnani et al. [18] also observed a positive correlation between the DI-DE pair and the FV-DI pair. These authors also observed a negative correlation between FV and DE, like us.

Questions that are too easy or too difficult are less discriminating. Hence, these questions need to be reconstructed to a moderate level of difficulty, either by changing the stem or by supplying better plausible distractors that will not test the interpretative or language skills of students as also inferred by Rao et al. [19] in their study [19]. A properly developed MCQ to suit a particular group of students will have moderate difficulty and high discrimination as also concluded by Izah SC et al. [20] in their article [20]. Thus, the difficulty and discrimination index serve as an indicator of the functional quality of each item [21].

Carneson et al. [1] found in their study that if appropriately constructed, MCQs can assess higher functions of Bloom's taxonomy like interpretation, synthesis, and application of knowledge [1]. Case et al. expressed the item index in percentages ranging from 0 to 100. Accordingly, the higher the percentage, the easier the item, with the recommended range being 30–70% [2]. Another famous text book by Singh et al. in their chapter 'Item analysis' define it as a

process of assembling, summarising and using information from students' responses to assess the quality of the given test [6]. Kaur et al. [22] in their findings also support the use of this technique to modify/remove a faulty item from subsequent tests. Thus a valid, reliable question bank of any speciality can be prepared [22].

Strengths of the study: It is the first study of its kind on item analysis in a major clinical subject, Obstetrics and Gynaecology. Most of the items were of acceptable difficulty and good to excellent discrimination. The majority of the items had acceptable to excellent discrimination scores. Around 50% of the items had functional distractions. We took feedback from students and faculty on a 5-point Likert scale.

Limitations: Sixty-four per cent of the items framed were hard as per facility value. A well-constructed MCQ questionnaire should have the maximum number of questions in the moderate difficulty range. The index study involved a smaller number of items and only a few students. It included only one subject out of various medical subjects. Increasing the number of items can improve the reliability of the study design. PBS (point bi-serial correlation) identifies those items that are odd ones out, i.e. they do not test the same construct as the remaining test. This parameter could also be included in further studies as it increases both the reliability and validity of the test. Similarly, reliability coefficients and standard errors of measurement can be used to further make the items more reliable. The refined items can thus be part of a unique question bank consisting of type A MCQs for both formative and summative assessment purposes.

Conclusion

There is a dire need to train medical faculty in constructing MCQs for formative and summative assessment, as MCQs constitute the majority of the professional exam pattern. Very easy and very difficult MCQs have low discrimination scores.

Item analysis should be an integral and regular activity for medical faculty in order to build subject-specific question banks. Items having average difficulty and high discrimination with functioning distractors should be incorporated into tests to improve the validity of the tests as well as the effectiveness of the questions.

Appendix

Single Best Response Type MCQs in OB GYN

Time duration : 40 min

Instructions

+ 4 marks for each correct answer

− 1 mark for each incorrect answer

Read the questions carefully

Select the single best option as the answer

Those who submit after the given time will be disqualified

A. Section: AUB (Abnormal Uterine Bleeding)

1. Meera, a 28-year-old nulliparous woman, complains of heavy menstrual bleeding and lower abdominal pain since six months. On examination, there is a 14 weeks size uterus with intramural fibroid. The treatment of choice is

   a. Wait & watch
   b. Myomectomy
   c. GnRH analogues
   d. Hysterectomy

2. A 45-year-old female comes to your clinic with complaints of heavy menstrual bleeding and excessive pain during periods. Endometrial biopsy is normal and USG shows diffusely enlarged uterus with no adnexal mass. What is your diagnosis?

   a. Fibroid uterus
   b. Endometritis
   c. Endometriosis
   d. Adenomyosis

3. Which of the following is not used for medical management of AUB?

   a. NSAIDs
   b. Tranexamic acid
   c. Combined oral contraceptive pills
   d. Letrozole

4. Transcervical endometrial resection (TCRE) is used in:

   a. Carcinoma endometrium
   b. Endometriosis
   c. Benign endometrial hyperplasia
   d. Submucous fibroid

5. The most common symptom of endometrial hyperplasia is:

   a. Vaginal discharge
   b. Vaginal bleeding
   c. Pelvic pain
   d. Amenorrhoea

B. Section: General Obstetrics

6. All of the following are prerequisites for medical management of ectopic pregnancy EXCEPT

   a. Patient is hemodynamically stable and desires future fertility
   b. Gestational sac size is less than 4 cm
   c. Lack of operative facilities and critical care unit
   d. Methotrexate is the most widely used drug

7. Which of the following is not included in Modified Bishop's score?

   a. Cervical dilatation
   b. Position of cervix
   c. Cervical effacement
   d. Cervical consistency

8. Recently the MTP Act,1971 has been amended. All of the following are true regarding the amendments EXCEPT

   a. MTP amendment Act, 2021 came into force on 25.3.21
   b. It permits MTP until 24 weeks of gestation in selected cases
   c. The medical board consists of a gynaecologist and a paediatrician
   d. Advice of only one specialist is sufficient for MTP up to 20 weeks

9. A term primi presented to labour room with per abdominal finding of soft mass being felt at lower uterine pole. The FHS is localized above umbilicus. All of the following are true regarding its management EXCEPT

   a. ECV can be tried in this case
   b. In case of an unsuccessful ECV, patient should be counselled on the risks and benefits of planned vaginal birth vs planned LSCS
   c. The perinatal mortality is about 0.5/1000 with LSCS and 2/1000 with planned vaginal birth

d. Planned vaginal births have high APGAR scores

10. A, $G_2 P_1 A_0 L_1$ patient presented with labour pains and no discharge per vaginum. She had delivered a male child by C-section 1.5 years back for foetal distress. Her PR -126/min; BP—90/60; PA findings—uterus term size, suprapubic tenderness + + +, foetal cardiac activity is absent. Which of the following is true regarding its management EXCEPT

a. Induction can be done
b. Blood and blood products should be reserved
c. Abnormal CTG is the most reliable finding
d. Large bore cannula is inserted

C. Section: Menopause

11. All are the causes of premature menopause EXCEPT

a. Sheehan's syndrome
b. Prolonged high doses of alkylating agents
c. Prolonged GnRH therapy
d. Prolonged high doses of antibiotics

12. What is the AMH value at menopause?

a. < 0.2 ng/ml
b. 0.2–1 ng/ml
c. 1–7 ng/ml
d. > 7 ng/ml

13. Which of these is NOT an effective treatment for menopause related vaginal dryness?

a. Vitamin E
b. Laser therapy
c. Masturbation
d. Petroleum jelly

14. A 60-year-old post-menopausal woman presents with vaginal bleeding. Which one of the followinginvestigation is not required?

Mark only one oval

a. Endometrial biopsy
b. Diagnostic laparoscopy
c. Pap smear
d. Hysteroscopy

15. HRT is a contraindication in all of the following conditions EXCEPT

a. Breast Cancer
b. Low grade endometrial sarcoma
c. Hepatocellular carcinoma
d. Age related macular degeneration

D. Section: Infertility

16. Cytokines involved in OHSS (ovarian hyperstimulation syndrome)

a. IL-2
b. VEGF
c. THF alpha
d. Endothelin

17. Semen analysis of a male partner (of infertile couple) shows azoospermia but presence of fructose. The most probable diagnosis is:

a. Prostatic infection
b. Mumps orchitis
c. Block in efferent duct system
d. All of the above

18. Pooja, a 26-year-old female, married since 2 years with regular unprotected coitus visits infertility OPD. Her hormonal profile is normal, Her USG also shows no abnormality. Her husband's semen analysis is as follows: Sperm count 39 million; Sperm concentration 15 million/ ml; 58% live sperms; 40% are progressively motile; semen volume is 5 ml. What will be your plan of management?

a. Ovulation induction by clomiphene citrate
b. Ovulation induction by letrozole
c. Reassurance with advice of timed coitus
d. Ovulation induction with intrauterine insemination

19. Select the most accurate instrument used to perform this procedure shown in the image given below:

a. MR SYRINGE CANNULA
b. RUBIN'S CANNULA
c. IUI CANNULA
d. MVA CANNULA

20. True regarding the given image of cervical mucous is as follows:

a. Five days after LH surge
b. Serum oestrogen level is higher than progesterone level
c. Serum progesterone level is higher than oestrogen level
d. Corpus luteum is present in the ovary

E. Section: High Risk Pregnancy

21. In a developing country like India, nutritional anaemia is commonly seen in *majority of pregnant females. Following are true for its management EXCEPT

    a. The earliest parameter to increase after oral iron therapy is reticulocyte count
    b. Parenteral iron is indicated in antenatal patients presenting with anaemia between 30 and 36 weeks
    c. Iron dextran is safer than iron sucrose
    d. 1 unit of blood transfusion raises the Hb levels within 24 h

22. Couvelaire uterus is seen in APH.AII of the following are true regarding this condition EXCEPT

    a. It is seen in concealed variety of APH
    b. Uterus appears patchy or of diffuse port wine colour
    c. It is an indication for doing hysterectomy
    d. It is also known as uterine apoplexy

23. All of the following are signs/ symptoms of impending eclampsia EXCEPT

    a. Epigastric pain
    b. Headache/dizziness
    c. Diplopia/ scotoma
    d. Polyuria

24. While managing a case of COVID positive pregnant patient in labour, following points are to be kept in mind EXCEPT

    a. LSCS is indicated to reduce mother to child viral transmission
    b. Foetal scalp blood sampling can be done
    c. Epidural analgesia is safe
    d. Continuous intrapartum foetal monitoring is required

25. Which of the following manoeuvres is used in shoulder dystocia?

    a. Pinard's manoeuvre
    b. Mauriceau–Smellie–Veit manoeuvre
    c. McRobert's manoeuvre
    d. Lovset's manoeuvre

Declarations

References

1. Carneson J, Delpierre G, Masters K, Designing and Managing MCQs: and Bloom's taxonomy; 2011. Retrieved from http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html.
2. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. 3rd ed. National Board of Medical Examiners; 2010. Retrieved from http://www.nbme.org/publications/item-writing-manual.html.
3. Skakun EN, Nanson EM, Taylor WC, et al. An investigation of three types of multiple choice questions. Ann Conf Res Med Educ. 1977;16:111–6.
4. Zubairi AM, Kassim NL. Classical and Rasch analysis of dichotomously scored reading comprehension test items. Malays J ELT Res. 2006;2:1–20.
5. Kumar D, Jaipurkar R, Shekhar A, et al. Item analysis of multiple choice questions: a quality assurance test for an assessment tool. Med J Armd Forc India. 2021;77:S85–9.
6. Ciraj AM. MCQ, Item analysis and question banking. In: Anshu ST, editors. Principles of assessment in medical education. NewDelhi: Jaypee Publishers; 2012. p. 88–106,116–127.
7. Ananthkrishnan N. The item analysis. In: Medical education principles and practice. 2nd ed. Pondicherry: JIPMER; 2000. p. 131–7.
8. Date AP, Borkar AS, Badwaik RT. Item analysis as tool to validate multiple choice question bank in pharmacology. Int J Basic Clin Pharmacol. 2019;8:1999–2003.
9. Patil VC, Patil HV. Item analysis of medicine multiple choice questions (MCQs) for undergraduate (3rd year MBBS) students. Res J Pharmaceut Biol Chem Sci. 2015;6:1242–51.
10. Gajjar S, Sharma R, Kumar P, et al. Item and test analysis to identify quality multiple choice questions (MCQS) from an assessment of medical students of Ahmedabad, Gujarat, Indian. J Commun Med. 2014;39:17–20.
11. Mukherjee P, Lahiri SK. Analysis of multiple choice questions (MCQs): item and test statistics from an assessment in a medical college of Kolkata, West Bengal. IOSR J Dent Med Sci. 2015;14:47–52.
12. Garg R, Kumar V, Maria J. Analysis of multiple choice questions from a formative assessment of medical students of a medical college in Delhi, India. Int J Res Med Sci. 2019;7:174–7.
13. Pande SS, Pande SR, Parate VR, et al. Correlation between difficulty and discrimination indices of MCQs in formative exam in physiology. South East Asian J Med Educ. 2013;7:45–50.
14. Shete AN, Kausar A, Lakhkar K. Item analysis: an evaluation of Mcqs in physiology examination. J Contemp Med Educ. 2015;3:106–9.
15. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and relationship between them in four response type multiple choice questions of pharmacology summative tests of year II MBBS students. Ie JSME. 2013;6:41–6.
16. Sim SM, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. Ann Acad Med Singapore. 2006;35:67–71.
17. Mitra NK, Nagaraja HS, Ponnudurai G. The levels of difficulty and discrimination indices in type A multiple choice questions of preclinical semester 1 multidisciplinary summative tests. Jpn Soc Mech Eng. 2009;3:2–7.
18. Khilnani AK, Thaddanee R, Khilnani G. Development of multiple choice question bank in otorhinolaryngology by item analysis: a cross-sectional study. Int J Otorhinolaryngol Head Neck Surg. 2019;5:449–53.
19. Rao C, Kishan Prasad H, Sajitha K, et al. Item analysis of multiple choice questions: assessing an assessment tool in medical students. Int J Educ Psychol Res. 2016;2(4):201.
20. Item analysis of Multiple Choice Questions (MCQs) from a formative assessment of first year microbiology major students [Internet]. www.oatext.com. [cited 2023 Sep 13]. Available from: https://www.oatext.com/item-analysis-of-multiple-choice-questions-mcqs-from-a-formative-assessment-of-first-year-microbiology-major-students.php#Article.
21. Lowe D. Set a multiple choice question (MCQ) examination. BMJ. 1991;302:780–2.
22. Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. Int J Appl Basic Med Res. 2016;6(3):170.