

A Simple Statistical Estimation of One's Performance in an MCQ Examination, Based upon Mock Test Results, Using Binomial Distribution of Probability

Sudipto Roy¹, Priyadarshi Majumdar^{2*}

¹Department of Physics, St. Xavier's College, Kolkata, India

²Jyotinagar Bidyasree Niketan Higher Secondary School, Kolkata, India
Email: roy.sudipto1@gmail.com, *majumdar_priyadarshi@yahoo.com

Received June 4, 2012; revised July 8, 2012; accepted July 20, 2012

ABSTRACT

A simple statistical model is proposed regarding the estimation of one's overall performance in an MCQ examination along with the calculation of probability of obtaining a certain percentage of marks in the same. Using the data obtained from the results of a sufficiently large number of mock examinations, conducted prior to the main examination, certain parameters quantifying one's knowledge or preparation for the examination has been calculated. Based on those parameters, the probability of obtaining a certain percentage of marks has been computed using the theory of binomial probability distribution. The dependence of this probability function on various parameters has been depicted graphically. A parameter, called the performance index, has been defined in terms of the expectation value and standard deviation of marks computed from probability distribution. Using this parameter, a new parameter called the relative performance index has been defined. This index estimates one's performance with respect to the best possible performance. The variation of relative performance index with respect to the preparation index has been shown graphically for different parameter values quantifying various aspects regarding the examination and the examinee.

Keywords: MCQ; Preparation Index; Knowledge Index; Application Index; Performance Index; Relative Performance; Binomial Probability Distribution; Estimation of Examination Result

1. Introduction

Although it is widely acknowledged that the scores or grades one makes in an examination can never be the ultimate judge of a student's talent or academic capability. But for all practical purposes the importance of an examination is universally accepted. It is a possible method to assess students because it judges them by a common technique (having the same set of questions) applied to all the examinees under the same circumstances (at least tried to be maintained). The general examination process, with long or descriptive answer type questions, has a major disadvantage that there exists no unique answer to a particular question, for any discipline whatsoever. In other words, it is impossible to prepare a unique model answer for a descriptive type question paper. Therefore, we have an obvious lack of uniformity in the ways of manual evaluation of the answer scripts by different examiners. It is not possible to use computers to evaluate such answer scripts because of the absence of a unique set of model answers. In-depth knowledge of sub-

ject and critical analysis are required to examine such scripts. An examiner, well versed with the subject, can only do this properly. Other shortcomings of the old examination process are that its implementation has a lot of difficulties and it is very much time consuming because of its manual nature of evaluation. Despite all these disadvantages the old examination process is still of great importance worldwide for University degrees that we are very much concerned about.

A new examination method, quite different from the old one, has been developed all over the world in various branches starting from medical and nursing sciences [1,2] to engineering and statistics [3]. It becomes increasingly popular in comparison to the traditional examination process with descriptive answer type questions. Here, one faces an examination with multiple choice type questions (MCQ). The answer to any question in such a paper is very brief and is presented in a single word or sentence in the question paper itself along with some incorrect answers (apparently close to the right answer) as options at the end of the question. The examinee has to choose the correct option(s) (sometimes it has more than one

*Corresponding author.

correct option). The degree of difficulty or stiffness of the question paper becomes higher as the given options become closer to one another in meaning. A great advantage with this MCQ type examination is that one can maintain uniformity of questions and evaluation process because of the availability of unique model answer. Apart from this, computers can be programmed to evaluate such answer scripts, reducing the cost and time of the process. The shortcomings of MCQ are that an examinee can make high scores in this kind of tests without having in-depth knowledge of the subject. Taking recourse to wild guess one can score moderately high marks without having any workable knowledge in the subject.

Srinivasa and Adkoll [1] studied the application of MCQ in medical education and suggested the necessity of developing a MCQ bank for future purpose. In another particular study [2] the three option and four option MCQ tests in nursing education are compared with each other and it was concluded that the three option tests performs equally well in compared to the other. The additional option(s) most often does not improve the test reliability and validity. Costa, Olivera and Ferrao [3] studied the psychometric properties of the three multiple choice tests used to measure skills in Statistics in the scope of the engineering and management courses of the University of Minho. Steif and Dantzer [4] quantified the learning of concepts in Statics using MCQ. Steif and Handen [5] analysed the results of MCQ tests that has been uses to measure learning in statics. In a more recent work Ventouas *et al.* [6] on the other hand discussed about the relative advantage and disadvantages of MCQ and CRQ (constructed response) based papers. In fact they admitted the advantages of MCQ with positive and negative marking concerning the objectivity in the grading process and speed of production of results.

MCQ s has a long and widespread history of use in support of teaching and assessment across a range of disciplines as mentioned. The increased use in the recent years relates to the governance and resource of higher education and changing student characteristics [7]. There is a substantial amount of assessment literature associated with MCQ s and it supports claims of benefits arising from effective practice and also identifies significant limitations when application is inappropriate or the need for assessment efficiencies takes precedence over pedagogical concerns and considerations [7-11]. MCQs have proved effective ways of assessing learning at the lower levels of Bloom's Taxonomy [12], identification and understanding. However, MCQ s can be constructed to be more challenging than was once believed [13] and so application at higher levels of Bloom's Taxonomy has also been demonstrated [14].

In contrary one of the most frequent criticisms of MCQ s is that they trivialize the learning process through

an over emphasis lower levels of cognitive demand [15] and that this often fails to reflect the true intentions of course learning objectives. According to Paxton [16] the learners may be quite skilled on taking MCQ tests but their ability to solve problems, exercise personal judgments and to communicate their understandings may remain underdeveloped. Other research [17,18] has found that students perceive MCQ s as assessing at a surface level and this consequently misdirects and otherwise influences the way in which they prepare for examinations.

In order to overcome such difficulties, the questions are chosen uniformly from all parts of the syllabus. These are of different patterns, such as knowledge based, understanding based, application based and skill based in appropriate proportions. The answers given at the end of each question must appear to be very close to one another to curb the tendency of answering without knowledge of the subject as mentioned earlier. We generally have negative markings for wrong attempts in MCQ examinations to discourage the random guessing process to a large extent.

In a recent work Ding and Beichner [19] discussed about some commonly used approaches of MCQ data analysis in physics research namely, classical test theory, factor analysis, cluster analysis, item response theory and model analysis.

2. Modelling

In the present article a statistical model has been developed regarding the estimation of probable performance of a student taking a MCQ examination. Some parameters, quantifying ones knowledge and understanding of the subject, have been defined in this context. Using this model, it would be possible to make a prediction of one's performance in the final examination on the basis of set of data from a sufficiently large number of mock tests preceding the final examination. It is possible to make an estimate of one's preparation by subjecting oneself to mock tests with question papers identical in pattern to that of the actual one. The parameters reflecting one's knowledge and perception, based on mock test data, are defined in the following way.

$$q = \frac{\text{number of questions attempted}}{\text{total number of questions in the papers}} \quad (1)$$

$$= \frac{\sum_{i=1}^{\Omega} A_i}{\sum_{i=1}^{\Omega} N_i},$$

$$r = \frac{\text{number of questions answered correctly}}{\text{total number of questions in the papers}} \quad (2)$$

$$= \frac{\sum_{i=1}^{\Omega} C_i}{\sum_{i=1}^{\Omega} N_i},$$

$$p = r/q = \frac{\text{number of questions answered correctly}}{\text{number of questions attempted}} \tag{3}$$

$$= \frac{\sum_{i=1}^{\Omega} C_i}{\sum_{i=1}^{\Omega} A_i}.$$

In the above expressions the total number of mock tests taken is denoted by Ω . For the i^{th} mock test, C_i , A_i and N_i denote the number of correct answers, questions attempted by the candidate and the total number of questions provided in the question paper respectively.

Equation (1) defines the parameter q as a rough indication of the familiarity of the examinee with the syllabus, assuming small chance of random guessing by him. We further assume that the question paper of each mock test is based upon the entire syllabus of the subject and the questions are chosen from all topics belonging to all parts of the syllabus uniformly. Hence for a sufficiently large Ω , the parameter q may also be regarded as a measure of one's knowledge according to one's own perception.

On the other hand Equation (2) defines the parameter r as a quantitative measure of preparation of the candidate for the examination. Let us call it the *preparation index*. One may conclude that the candidate has acquired $(100 \times r)\%$ knowledge of the syllabus. For a sufficiently large Ω , this parameter reflects one's true understanding of the subject (we are assuming that the candidate

read all the questions and had sufficient time to answer them). Regarding the ranges of q and r one can say that $0 \leq q \leq 1, 0 \leq r \leq 1$.

The parameter p as defined by Equation (3) can be looked upon as the probability of answering a question correctly. For a sufficiently large Ω , this ratio ($p \equiv r/q$), obtained from mock test data, can be considered as the chance of success in every attempt to be made in the forthcoming final examination. Let us call it the *answering efficiency*. Larger value of p indicates smaller difference between one's actual knowledge and one's own perception of his knowledge.

In reality, two different classes of questions, namely knowledge based and application based, should be present in an ideal question paper. Application based questions are mainly intended for the assessment of one's logical reasoning ability and also the ability of solving numerical problems. Thus, instead of having a single *preparation index* (r), we find it reasonable to define two separate indices for these two different types of question sets. Let us define *knowledge index* (k) as a measure of one's ability to answer knowledge or information based questions. While for the application-based questions one should define a new parameter related to intuition, power of understanding and application. Let us designate it as *application index* (a). These parameters can be mathematically expressed as

$$k = \frac{\text{number of knowledge based questions answered correctly}}{\text{total number of knowledge based questions}} = \frac{\sum_{i=1}^{\Omega} C_i^k}{\sum_{i=1}^{\Omega} N_i^k}, \tag{4}$$

$$a = \frac{\text{number of application based questions answered correctly}}{\text{total number of application based questions}} = \frac{\sum_{i=1}^{\Omega} C_i^a}{\sum_{i=1}^{\Omega} N_i^a}. \tag{5}$$

Here N_i^k and N_i^a are respectively the numbers of knowledge-based and application-based questions in the i^{th} test. Two other quantities C_i^k and C_i^a are respectively the numbers of correct answers to knowledge based and application based questions in the i^{th} test. Combining Equations (2), (4) and (5) we may write

$$r = \frac{\sum_{i=1}^{\Omega} C_i^k + \sum_{i=1}^{\Omega} C_i^a}{\sum_{i=1}^{\Omega} N_i} = k \left(\frac{\sum_{i=1}^{\Omega} N_i^k}{\sum_{i=1}^{\Omega} N_i} \right) + a \left(\frac{\sum_{i=1}^{\Omega} N_i^a}{\sum_{i=1}^{\Omega} N_i} \right), \tag{6}$$

where

$$C_i^k + C_i^a = C_i. \tag{7}$$

The bracketed parts of the right hand side of Equation (6) are simply the numbers of questions of both types, expressed as fractions of the total number of questions of all types in all tests. This equation expresses the relation among the parameters k , a and r .

To formulate a mathematical model that explores the relationship between the preparation and performance in any MCQ examination process, let us think of a question paper in the final examination consisting of full marks F , with N number of questions with equal weight. Hence the marks allotted for each question is F/N . The questions are on a subject where one needs to memorise and apply objective information only. Based on the information obtained from a sufficiently large number of mock tests taken before the main examination, one can reasonably assume that the parameter called *answering efficiency* ($p \equiv r/q$) is the probability of answering a question correctly. An answer to any of such multiple-choice type questions (MCQ) can be only RIGHT or WRONG. To keep matters simple we have assumed further that the process of attempting a question and its result is independent of attempting any other question.

Let us define a parameter Q as the fraction of questions attempted by a candidate in the final or main ex-

amination. Hence, QN is the number of questions attempted by the candidate. Using the binomial distribution, the probability of having y answers RIGHT in QN attempts is given by [20,21],

$$P_y = {}^{QN}C_y (p)^y (1-p)^{QN-y} \tag{8}$$

Now in writing the above equation we have assumed that the difficulty level of the questions (item difficulty) is same for all the questions under consideration and this assumption is most crucial because our paper tries to quantify those concepts that are difficult to quantify in educational research.

Following (8) now the expectation or mean value of y

$$y_{\text{exp}} = pQN \tag{9}$$

While the corresponding standard deviation

$$y_{SD} = [QNp(1-p)]^{1/2} \tag{10}$$

As the marks awarded for each correct answer is F/N the negative marks allotted for each wrong answer is $[-n(F/N)]$ where n is a positive fraction, which may be called the *negative marking factor*.

Hence, for y correct answers, the marks obtained by the candidate is given by

$$M = y(F/N) - (QN - y)n(F/N) \tag{11}$$

The percentage of marks obtained by the candidate is given by

$$m = 100(M/F) = 100[(y/N)(1+n) - nQ] \tag{12}$$

So

$$y = f(m) = (0.01m + nQ)[N/(1+n)] \tag{13}$$

An expression for the probability (P_m) of securing $m\%$ marks can be obtained from Equations (8) and (13) as

$$P_m = {}^{QN}C_{f(m)} (p)^{f(m)} (1-p)^{QN-f(m)} \tag{14}$$

Equation (14) determines the probability distribution of m . Depending upon the values of p ($\equiv r/q$) and Q , m should vary over a range determined by this probability distribution. Using Equations (9), (10) and (13) the expectation value and standard deviation of m are given by

$$\begin{aligned} m_{\text{exp}} &= 100[(y_{\text{exp}}/N)(1+n) - nQ] \\ &= 100Q[p(1+n) - n] \end{aligned} \tag{15}$$

$$\begin{aligned} m_{SD} &= 100(y_{SD}/N)(1+n) \\ &= 100(1+n)N^{-1/2}[Qp(1-p)]^{1/2} \end{aligned} \tag{16}$$

m_{SD} is a measure of dispersion in the probable values of m . Consistency in the probable values of marks percentage (m) is determined generally by the smallness of the ratio (m_{SD}/m_{exp}) [20,22,23]. Using the value of p

obtainable from the mock tests, one can predict the performance to be made in the main examination up to a certain extent. Generally, the performance is considered to be reasonably good when m_{exp} is sufficiently high and (m_{SD}/m_{exp}) is sufficiently low. So, as a quantitative measure of the probable performance, we can define a term called *performance index* (P) in the following manner.

$$\begin{aligned} P &= \text{Expectation value} \times \text{consistency} \\ &= m_{\text{exp}} \frac{m_{\text{exp}}}{m_{SD}} = \frac{m_{\text{exp}}^2}{m_{SD}} \end{aligned} \tag{17}$$

It is evident from Equation (16) that, $m_{SD} = 0$ when $p = 0$ or, $p = 1$. For these two extreme cases, the function P becomes infinity (using Equation (17)). To avoid this difficulty we propose to define P as

$$\begin{aligned} P &= \left[\frac{1}{m_{\text{exp}}} + \frac{1}{m_{\text{exp}}/m_{SD}} \right]^{-1} = \frac{m_{\text{exp}}}{1+m_{SD}} \\ &= \frac{100Q[p(1+n) - n]}{1+100(1+n)N^{-1/2}[Qp(1-p)]^{1/2}} \\ &= \frac{100Q[r(1+n) - nq]}{q+100(1+n)N^{-1/2}[Qr(q-r)]^{1/2}} \end{aligned} \tag{18}$$

The parameter r , which is calculated from the data from a large number of mock tests, reflects one's true knowledge. If N_1 is the total number of questions in all mock tests, rN_1 is the number of questions answered correctly. Since qN_1 is the total number of questions attempted in all mock tests, we can write

$$qN_1 = rN_1 + \alpha(N_1 - rN_1) \text{ where } 0 \leq \alpha \leq 1, \tag{19}$$

Hence

$$q = r + \alpha(1-r) \tag{20}$$

Here, $\alpha = 0$ and $\alpha = 1$ corresponds to the situation where one has not attempted any question in the mock tests beyond one's knowledge and the situation where one has attempted all the questions respectively. For $r = 1$, indicating complete knowledge of the subject, q is independent of α . Clearly, for $r < 1$, the parameter α is a measure of one's tendency for attempting questions completely unknown to him.

Similarly, the parameter Q (fraction of questions attempted in the final examination) should also depend on one's preparation or knowledge of the syllabus. Since r is a measure of one's knowledge, as obtained from mock tests, we can write

$$Q = r + \beta(1-r) \text{ where } 0 \leq \beta \leq 1. \tag{21}$$

Using Equations (20) and (21) in Equation (18) we have

$$P = \frac{100[r(1-\beta) + \beta][r(1+n\alpha) - n\alpha]}{r(1-\alpha) + \alpha + 100(1+n)N^{-1/2}[r\alpha(1-r)(r + \beta - \beta r)]^{1/2}} \tag{22}$$

From the above equation we can calculate the maximum and minimum values of P as $P_{\max} = 100$ (for $r = 1$) and $P_{\min} = -100n\beta$ (for $r = 0, \beta = 1$).

For convenience in analysis, we may define a parameter, called the *relative performance index* (P_r), which judges one's relative performance with respect to the best possible performance as defined below.

$$P_r = \frac{P}{P_{\max}} = \frac{P}{100} \tag{23}$$

Using Equations (22) and (23) the first four figures (Figures 1 to 4) of this article have been drawn. These figures depict how one's performance index (P_r) depends on one's preparation index (r).

Figure 3 clearly indicates that, negative marking in an examination, has very little effect on the probable performance of an examinee having an extremely good preparation.

We now recast Equation (14) as

$$P_m = {}^Q C_{f(m)} \left(\frac{r}{q}\right)^{f(m)} \left(1 - \frac{r}{q}\right)^{Q - f(m)} \tag{24}$$

where q and Q are already defined (Equations (20) and (21))

$$f(m) = [0.01m + n(r + \beta(1-r))][N/(1+n)] \tag{25}$$

Using the above two equations we have drawn Figures 5 to 9. These curves depict the nature of variation of probability (P_m) as a function of m . The peak of any of

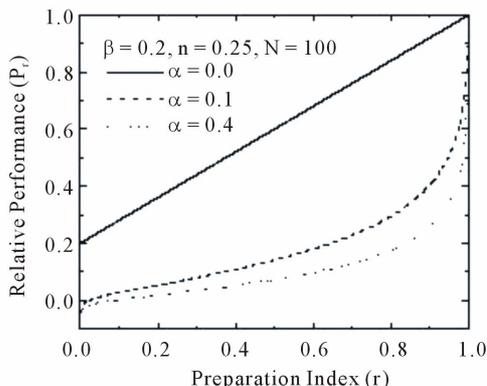


Figure 1. The variation of the relative performance (P_r) as a function of preparation index (r) for different values of α . As r increases, P_r increases. For $\alpha = 0$ the variation is linear (as evident from Equation (22)). For other values of α , the slope increases with r . At any fixed value of r , P_r decreases as α increases.

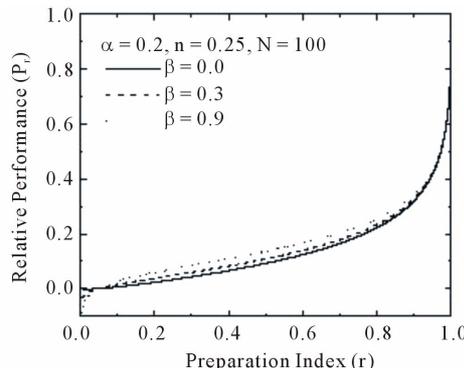


Figure 2. The variation of the relative performance (P_r) as a function of r for different values of β . As r increases, P_r increases. At any value of r , P_r is larger for higher values of β . These curves become closer to one another at higher values of r .

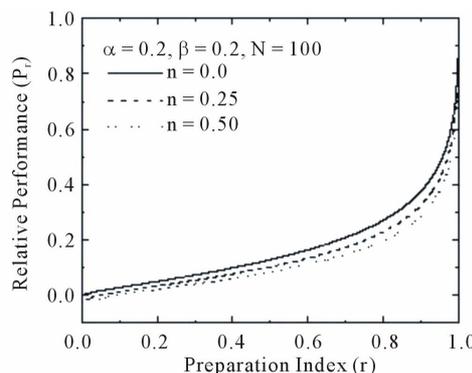


Figure 3. Relative performance (P_r) vs r plot for different values of n . As r increases, P_r increases. At any particular level of preparation r , P_r becomes smaller for higher values of n . These curves become closer to one another at higher values of r . These curves converge in the direction of increasing r . At values of r close to 1 P_r becomes independent of n .

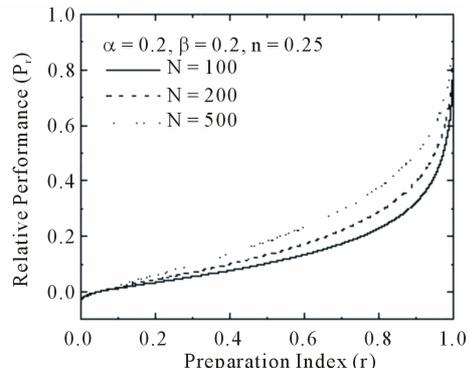


Figure 4. This figure shows the variation of relative performance (P_r) as a function of preparation index (r) for different values of N . At any fixed value of r , P_r increases as N increases. As r increases, the slope of the curve increases. Initially at very small values of r these curves are very close to each other. As r increases, curves diverge from one another and they again become closer at values of r close to 1.

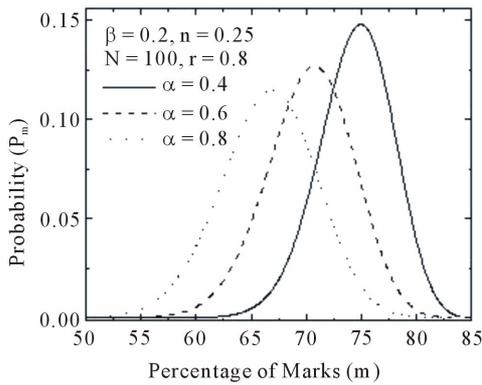


Figure 5. This figure shows the variation of P_m as a function of m for different values of α . As α decreases, the curve becomes narrower, indicating smaller dispersion of the probable values of m . The peak of the curve, indicating the most probable value of m , becomes higher for smaller values of α . As α decreases the peak shifts towards higher values of m .

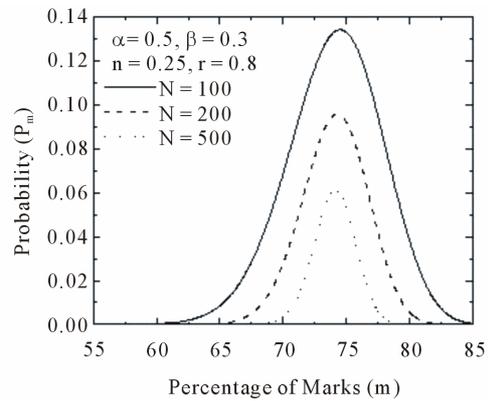


Figure 8. Variation of P_m as a function of m for different values of N . The peaks of these curves appear at the same value of m , indicating the constancy of the most probable value of m . As N increases, the height of the curve decreases. For greater values of N , the probability (P_m) is smaller at all values of m .

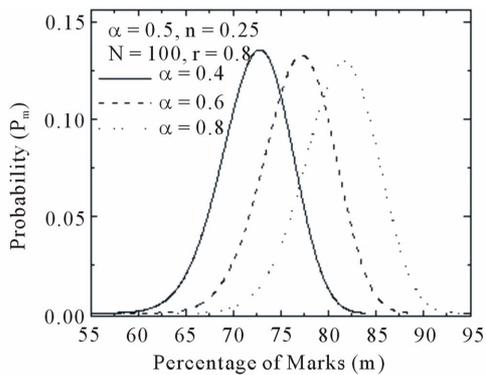


Figure 6. This figure shows the variation of P_m with m for different β . As β increases, the curves become broader, indicating greater dispersion of the probable values of m . The peak of the curves, indicating the most probable value of m , becomes higher for smaller β . With increase in β , the peak shifts towards higher m .

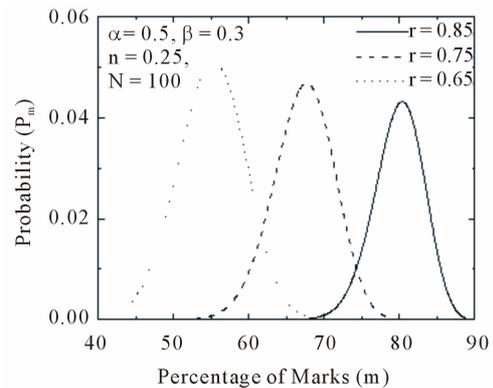


Figure 9. The variation of P_m with m for different values of r . As r increases the peak shifts towards higher values of m . The most probable value of m increases as r increases. The curve becomes narrower and the peak height decreases with rise in r -value indicating ones true measure of preparation for the examination.

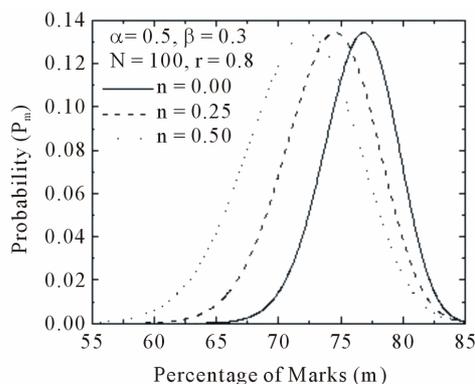


Figure 7. Figure shows P_m - m variation for different n . As n decreases, the curve becomes narrower, indicating smaller dispersion of the probable values of m . The peaks of these curves have the same height, although their positions are different. As n increase, the peaks shift towards smaller values of m , indicating smaller expectation value of m .

these curves corresponds to the most probable value of m .

In **Figure 9** the narrowing of curves with rise in r indicates smaller dispersion of the probable values of m , implying smaller chances of deviation from the most probable value. A candidate, having a high score of r in the mock tests, is likely to acquire high marks in the final examination, with smaller deviation of probable values from the most probable m .

3. Limitations

There may be limitations and implementation difficulty of this statistical model, caused by some assumptions, made for the sake of simplicity of mathematical analysis. The success of this model depends upon the largeness of the parameter Ω (in a statistical sense) α has to be de-

terminated from the mock test data and the estimation of performance has to be made for different values of β . The larger the value of Ω , greater will be the validity and applicability of this approach where these parameters can be used to predict one's performance. This model shows that one is likely to make better performance in examinations having larger number of questions. This dependence on N is a mathematical consequence that cannot generally be guessed from common sense. It has also been assumed that the process of attempting a question and its result is independent of attempting any other question. This assumption does not hold for linked comprehension questions where, the process of attempting a question and its result depends on attempting other linked questions. In this regard a modification of our simple theory using the conditional probability [21] is required. Let the events of attempting successive questions in a linked comprehension be A, B, C , etc. Then, according to the conditional probability [21] we have

$$P(B/A) = P(B \cap A)/P(A), \quad (26)$$

$$P(C/B) = P(C \cap B)/P(B), \quad (27)$$

and so on.

These ideas can be incorporated for theoretical interests. Calculations, based on such ideas, are likely to make this model so complicated that it would not be very useful to examinees preparing for competitive examinations. The mathematical simplicity in its present form is important in the sense that one can use this model successfully with considerable ease, for an estimation of performance, without making too much effort to grasp the underlying concept. The present analysis reveals important and useful features, which one can't discover just by intuition. It enables one to make an effective self-assessment, and thereby modify one's plans, while preparing for an important examination.

REFERENCES

- [1] D. K. Srinivasa and B. V. Adkoll, "Multiple Choice Questions: How to Construct and How to Evaluate?" *Indian Journal of Pediatrics*, Vol. 56, No. 1, 1989, pp. 69-74.
- [2] M. Tarrant and J. Ware, "Impact of Item-Writing Flaws in Multiple-Choice Questions on Student Achievement in High-Stakes Nursing Assessments," *Medical Education*, Vol. 42, No. 2, 2008, pp. 198-206. [doi:10.1111/j.1365-2923.2007.02957.x](https://doi.org/10.1111/j.1365-2923.2007.02957.x)
- [3] P. Costa, P. Olivera and M. E. Ferrao, "Equalização de Escalas com o Modelo de Resposta as Item de Dois Parâmetros," In: M. Hill, *et al.*, Eds., *Estatística-da Teoria à Prática, Actas do XV Congresso Annual da Sociedade Portuguesa de Estatística*, Edições SPE, 2008, pp. 155-166.
- [4] P. Steif and J. Dantzler, "Astatics Concept Inventory: Development and Psychometric Analysis," *Journal of Engineering Education*, Vol. 33, 2005, pp. 363-371.
- [5] P. Steif and M. A. Handsen, "Comparisons between Performances in a Statics Concept Inventory and Course Examinations," *International Journal of Engineering Education*, Vol. 22, No. 3, 2006, pp. 1070-1076.
- [6] E. Ventouas, D. Triantis, P. Tsiakas and C. Stergiopoulos, "Comparison of Examination Methods Based on Multiple Choice Questions," *Computers & Education*, Vol. 54, No. 2, 2010, pp. 455-461. [doi:10.1016/j.compedu.2009.08.028](https://doi.org/10.1016/j.compedu.2009.08.028)
- [7] D. Nicol, "E-Assessment by Design: Using Multiple-Choice Tests to Good Effect," *Journal of Further & Higher Education*, Vol. 31, No. 1, 2007, pp. 53-64. [doi:10.1080/03098770601167922](https://doi.org/10.1080/03098770601167922)
- [8] L. Thompson, "The Uses and Abuses of Multiple Choice Testing in a University Setting," Annotated Bibliography Prepared for the University Centre for Teaching and Learning, University of Canterbury, Canterbury, 2005.
- [9] P. Nightingale, *et al.*, "Assessing Learning in Universities," Professional Development Centre, University of New South Wales, 1996, pp. 151-157.
- [10] J. Heywood, "Assessment in Higher Education: Student Learning, Teaching Programmes and Institutions," Jessica Kingsley Publishers, London, 2000.
- [11] N. Falchikov, "Improving Assessment through Student Involvement: Practical Solutions for Aiding Learning in Higher and Further Education," Routledge Falmer, London, 2005.
- [12] D. Krathwohl, "A Revision of Bloom's Taxonomy: An Overview," *Theory into Practice*, Vol. 41, No. 4, 2002, pp. 212-218. [doi:10.1207/s15430421tip4104_2](https://doi.org/10.1207/s15430421tip4104_2)
- [13] S. Brown, "Institutional Strategies for Assessment," In: S. Brown and A. Glasner, Eds., *Assessment Matters in Higher Education*, SRHE and Open University Press, Buckingham, 1999, pp. 3-13.
- [14] M. Culwick, "Designing and Managing MCQs," University of Leicester, The Castle Toolkit, 2002.
- [15] S. Kvale, "Contradictions of Assessment for Learning in Institutions of Higher Education," In: D. Boud and N. Falchikov, Eds., *Rethinking Assessment in Higher Education: Learning for the Longer Term*, Routledge, London, 2007, pp. 57-71.
- [16] M. Paxton, "A Linguistic Perspective on Multiple Choice Questioning," *Assessment and Evaluation in Higher Education*, Vol. 25, No. 2, 2000, pp. 109-119. [doi:10.1080/713611429](https://doi.org/10.1080/713611429)
- [17] G. Gibbs and C. Simpson, "Conditions under Which Assessment Supports Students' Learning," *Learning and Teaching in Higher Education*, Vol. 1, No. 1, 2004, pp. 3-29.
- [18] K. Scouller, "The Influence of Assessment Method on Students' Learning Approaches: Multiple-Choice Question Examination versus Assignment Essay," *Higher Education*, Vol. 35, No. 4, 1998, pp. 453-472. [doi:10.1023/A:1003196224280](https://doi.org/10.1023/A:1003196224280)
- [19] L. Ding and R. Beichner, "Approaches to Data Analysis of Multiple Choice Questions," *Physical Review Special*

- Topics—Physics Education Research*, Vol. 5, 2009, Article ID: 020103.
- [20] N. G. Das, “Statistical Methods,” Tata McGraw-Hill Publishing Company Ltd., New Delhi, 2008.
- [21] A. M. Goon, M. K. Gupta and B. Das Gupta, “Fundamentals of Statistics,” The World Press Pvt. Ltd., Kolkata, 1971.
- [22] M. R. Spiegel, *et al.*, “Schaum’s Outlines of Statistics,” 3rd Edition, McGraw Hill, New York, 1999.
- [23] M. R. Spiegel, *et al.*, “Schaum’s Outlines of Probability and Statistics,” 3rd Edition, McGraw Hill, New York, 2009.