SKEWNESS

Jurnal Statistika, Aktuaria dan Sains Data

Volume 2, No. 1, April 2025

The Modeling of Multiple-Choice Questions Using Binomial Distribution

Agung Prabowo [1*], Hasriati [2]

[1]*Statistics Study Program, Faculty of Mathematics and Natural Sciences,*

Universitas Jendral Soedirman, Purwokerto, Indonesia

[2]*Mathematics Department, Faculty of Mathematics and Natural Sciences,*

Universitas Riau, Riau, Indonesia

Corresponding e-mail: agung.prabowo@unsoed.ac.id

Abstract. One weakness of multiple-choice tests is that it gives a chance for test takers to correctly answer the question without knowing for sure that the selected response is the right one. A good admission test is a multiple-choice test combination that minimizes the probability of answering the question correctly by randomly guessing it. In this way, the multiple-choice test will be able to serve its function as a performance differential or for each test taker. In this article, a mathematical model that can be used to see the risk of admitting low-performing students for passing the admission selection using multiple-choice tests will be built. The model is made based on a binomial distribution combination. The model can be designed to choose the sets of multiple-choice tests that minimize the probability of admitting low-performing participants. The research methodology is to review the literature. In this article, we analyze two examples of applications. The illustration in an example of application 1 gives a result that setting the passing grade at 50 allows only 1 low-performing participant out of 1,000,000 participants to pass the test for answering the questions correctly merely by randomly guessing it. Test II in an example of application 2 is more homogeneous and sharper in measuring the participants' performance.

Keywords: *binomial distribution, performance, minimize the probability, randomly guessing, survival function.*

1 Introduction

Multiple choice tests have been a popular method for mass selection when test participants are abundant. The multiple choice test questions are selected based on such criteria as difficulty level, discriminating power and distractor efficiency. Prior to its use, they need to be tested for their validity, reliability and discrimination ability. A good test needs to take the easy, moderate, and difficult question composition into consideration [1].

Multiple choice tests ensure that the grading time is brief and its ability to select accurately can always be improved based on the combination of the test types used. Multiple choice tests are used, for example, in university admission test. In this test, two or more combination tests are given and all of them are multiple choice. For example, Test I is to test Mathematical ability and Test II for English proficiency.

The record of applied research on multiple choice test combination has been started since 2013. [2], [3], [4], [5], [6], [7], [8], [9], [10], and [11] studied the relationship between university admission selection test and students' score acquisition in Mathematics and Natural Sciences Faculty at Charles University. Similar research at a different faculty and university has been conducted by Zvára and Anděl [12]. A more recent similar article was published by Wahyuni et al. [13] which provides variations in the types of multiple-choice tests.

One weakness of multiple choice test is that it gives a chance for test takers to correctly answer the question without knowing for sure that the selected response is the right one. This is a case of answering correctly by randomly guessing it [14]. A good admission test is a multiple choice test combination which minimizes the probability of answering question correctly by randomly guessing it. In this way, the multiple choice test will be able to serve its function as a performance differentiator of each test taker.

[15], [16], and [17] has produced research on minimizing the probability of answering multiple choice test correctly by randomly guessing it. The solution they offered was to convert the obtained test score (raw data) into a test score in a standard percentage.

In [18] concluded that multiple choice tests manage to distinguish the high-performing participants from their low-performing counterparts. The risk of admitting low-performing students can be supressed up to 1 out of 1 million participants [18]. In his further research, [19] once again suggests that the use of multiple choice tests manages to minimize the risk of admitting students with low ability. However, for a small number of test participants, [19] suggests to using a standardized test (without multiple choice) rather than a multiple choice test.

The minimum number of types/fields of multiple choice tests during an admission test is two. Each test type should be independent to each other. This means that the acquired score in one test does not affect the other test. [20] research concluded that there is an inter-independent relationship between multiple choice tests for all the fields being tested. As a result, the attained test score for a tested ability has no influence on other abilities. The results of [20], [21] research are crucial since his inter-independence nature makes the mathematical calculation easier.

The elements influencing the choice of a test are (1) the number of subjects to be tested, (2) the number of questions for each subject, (3) the number of response choices for each question in every tested subject, and (4) the score for each question [22].The rules used are that the maximum score of every test type is 100 and the score will not be deduced for every incorrect response [19].

The multiple choice tests are used using two or more test combinations. In Table 1, the test consists of two types/fields/abilities, making a total of two test subjects. For example, [19] uses two test subjects, namely Mathematics and English. For Mathematics test, five options are available for each question. Ten questions are given with the score for each question being 5 and 5 other questions are given with their score being 10 each. The total score for Mathematics test is 100. For the English multiple choice test, 40 questions are given with their respective score being 2.5 and four response options are provided for each questions 4 (Table 1).

Table 1. Multiple Choice Tests with Two Test Subjects

Test Elemenets	**Multiple Choice Test Type**		
Subjects	Test I Mathematics		Test II English
Group number	1	2	1
Number of questions	10	5	40
Score for each question	5	10	2.5
Number of alternatives	5		4
Total scores	100		100

Source: [19].

In another model, [23] uses a multiple choice test as a combination of two subjects, i.e., Mathematics and English, yet the total number of questions for English test is 50 (Table 2).

Table 2. Multiple Choice Tests with Two Test Subjects

Test Elemenets	**Multiple Choice Test Type**		
Subjects	Test I Mathematics		Test II English
Group number	1	2	1
Number of questions	10	5	50
Score for each question	5	10	2
Number of alternatives	5		4
Total scores	100		100

Source: [23].

The tests in Tables 1 and 2 have similar combinations, with each of them consisting of two tests. Test I consists of two question groups and Test II consists of one question group. The difference lies in the number of questions and score for each question in Test II. In section 3.2 Example of Application 1, an example of how to use of binomial distribution with a test combination based on Table 1 is provided. Meanwhile, in section 3.3 Example of Application 2, an example of how to develop Tables 1 and 2 is given. In Example of Application 3, two test types are used with Test I consisting of three test groups and Test II consisting of two test groups.

The accuracy of multiple choice test will be better when the number of ability fields being tested is increased. In other words, the number of tests is increased. In [24] suggests to add another test for university admission selection, making a total of three tests, for example Mathematics, English and General Knowledge (Table 3). Mathematics ability is tested in 50 multiple choice questions with 4 response options and scored 2 for every correct answer. The English test consists of 40 questions with 4 response options and 2.5 score for each question. Meanwhile, for the General Knowledge test, it is suggested to increase the number of response options to 5 with 40 questions and 2.5 score for each question. Every test type in Table 3 consists only of one test group.

Table 3. Multiple Choice Tests with Three Test Subjects

Test Elemenets	**Multiple Choice Test Type**		
Subjects	Test I Mathematics	Test II English	Tes III General Knowledge
Number of groups	1	1	1
Number of questions	50	40	40
Score for each question	2	2.5	2.5
Number of alternatives	4	4	5
Total scores	100	100	100

Source: [24].

Analysis of Table 3 provides us with information that Mathematics and English tests are designed in such a way that their mean is the same, yet the standard deviation of Mathematics test is lower than English test. This means that the passing participants' Mathematics ability will be more homogeneous than their English proficiency. Furthermore, the probability to score 40 or more in English test is greater than in Mathematics test. General knowledge test is added to make the overall test ability more accurate and selective in filtering the passing participants. Adding General Knowledge test generates lower probability of scoring 40 or more in General Knowledge test than in English Test. Meanwhile, the average score of General Knowledge test is lower than Mathematics and English tests.

Based on the explanation above, in this article a mathematical model which can be used to see the risk of admitting low-performing students for passing the admission selection using multiple choice tests will be built. The admission test consists of k multiple choice test types, each of which is characterized by the number of questions, score for each question and number of response options. Every test type will be assigned a maximum score of 100 and no penalty is given for every incorrect response.

The model is made based on binomial distribution combination. Every test type brings with them the appropriate binomial distribution. Using binomial distribution requires an assumption that the response options are provided randomly (with the same probability). Using binomial distribution, the probability for the correct response to exceed a certain amount can be calculated and the number of correct responses can be expected. The generated model can be designed to choose the sets of multiple choice tests which minimize the probability of admitting low-performing participants. The criterion for a participant to be classified as low-performing is those passing the test only due to luck of randomly guessing yet giving the correct responses which qualify the minimum passing grade.

2 Materials and Methods

Suppose a random variable is $X \sim bin(n, p)$ then probability mass function for this random variable is (Walpole and Myers, 1986):

$$Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad ; x = 0,1,2, \ldots, n \; , 0 < p < 1 \tag{1}$$

where

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)\,!}$$

Expectation and variance of random variable X is

$$\begin{aligned} E[X] &= np \\ Var(X) &= np(1-p) \end{aligned} \tag{2}$$

Based on the explanation above, in this article a mathematical model which can be used to see the risk of admitting low-performing students for passing the admission selection using multiple choice tests will be built. The admission test consists of k multiple choice test types, each of which is characterized by the number of questions, score for each question and number of response options. Every test type will be assigned a maximum score of 100 and no penalty is given for every incorrect response.

The cumulative distribution function of random variable $X \sim bin(n, p)$ is

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k} \quad (3)$$

Binomial distribution is a discrete distribution. The cumulative distribution function for discrete random variable $X \sim bin(n, p)$ which has continuous values for random variable X is:

$$F(x) = Pr(X \leq x) = \begin{cases} 0 & ; x < 0 \\ \sum_{k=0}^{[x]} \binom{n}{k} p^k (1-p)^{n-k} & ; x \geq 0 \end{cases} \quad (4)$$

where $[x]$ is integer part of x.

The survival function is defined as

$$\boldsymbol{S(x) = 1 - F(x)} \quad (5)$$

It is important to have the survival function since for the participants to pass the test they are required to be able to correctly answer x questions with x = 0, 1, 2, ….

3 Results and Discussion

3.1 Binomial Distribution Modeling

3.1.1 The Binomial Model Construction for Test I

Suppose $X_1, X_2, \ldots\ldots, X_k$ is k random variables each of which has an inter-independent binomial distribution, respectively with parameters n_1, p_1, n_2, p_2, …., n_k, p_k. In short, suppose $X_i \sim bin(n_i, p_i)$; $i = 1,2, \ldots, k$. As an alternative, it is possible $p_i = p$ for every i.

The expectation and variance for random variable X_i ; $i = 1,2, \ldots, k$ is obtained from (2), i.e.,

$$E[X_i] = n_i \cdot p_i \qquad ; i = 1,2, \ldots, k \quad (6)$$
$$Var(X_i) = n_i \cdot p_i \cdot (1 - p_i) \qquad ; i = 1,2, \ldots, k$$

Suppose for Test I, there are k test groups all of which are multiple choice tests with each of them having the number of questions$n_1, n_2, \ldots., n_k$ and for each test group, the score of every question is $a_1, a_2, \ldots., a_k$, with $\sum_{i=1}^{k} a_i n_i = 100$. Suppose the variable Y indicates the total score obtained using:

$$Y = a_1 X_1 + a_2 X_2 + \ldots\ldots + a_k X_k \tag{7}$$

where,

$a_1, a_2, \ldots., a_k > 0$

$x_1 = 0,1,2, \ldots., n_1,\ x_2 = 0,1,2, \ldots., n_2, \ldots\ldots.., x_k = 0,1,2, \ldots., n_k$

The equation (7) can be stated using

$$Y = \sum_{i=1}^{k} a_i X_i \quad ; a_i > 0\ ; x_i = 0,1,2, \ldots., n_i \tag{8}$$

The random variable $X_i\ ; i = 1,2, \ldots, k$ in equations (7) and (8) has the probability mass function as per equation (1), i.e.

$$Pr(X_i = x_i) = \binom{n_i}{x_i} p_i^{\,x_i} (1 - p_i)^{n_i - x_i} \quad ; x_i = 0,1,2, \ldots., n_i\ , 0 < p_i < 1 \tag{9}$$

where p_i is the probability of guessing/answering correctly whose score is one divided by the number of response options.

Assuming that each random variable is inter-independent, then

$$E[Y] = E\left[\sum_{i=1}^{k} a_i X_i\right] = \sum_{i=1}^{k} a_i E[X_i] = \sum_{i=1}^{k} a_i n_i p_i \tag{10}$$

$$Var(Y) = Var\left(\sum_{i=1}^{k} a_i X_i\right) = \sum_{i=1}^{k} a_i^2 Var(X_i) = \sum_{i=1}^{k} a_i^2 n_i p_i\,(1 - p_i) \tag{11}$$

Suppose the admission selection for Test I is determined by two test groups. The first one is Test I Group 1 and the second one is Test I Group 2. The test score for Group 1 is stated with random variable X_1 and the test score for Group 2 is stated with random variable X_2. Assuming that the random variables X_1 and X_2 have binomial distribution respectively and are inter-independent to each other, then the probability for participants to score y or more can be calculated using the survival function in (5), namely

$$P(Y > y) = P(X_1 > y) + P(X_2 > y) \tag{12}$$

where $0 \leq y \leq 100$.

If Test I uses k test groups and every test group is stated as a random variable which has binomial distribution and is inter-independent to another, then the probability for participants to score y or more is

$$P(Y > y) = \sum_{i=1}^{k} P(X_i > y) \tag{13}$$

3.1.2 Binomial Model Construction for Test II

Suppose $W_j \sim bin(m_j, p_j)$; $j = 1,2,\dots,l$. As an alternative, it is possible $p_j = p$ for every j. In this case, W_j ; $j = 1,2,\dots,l$ has the probability mass function which follows (1), i.e.,

$$Pr(W_j = w_j) = \binom{m_j}{w_j} p_j^{w_j} (1 - p_j)^{m_j - w_j} \quad ; w_j = 0,1,2,\dots,m_j \ , 0 < p_j < 1 \tag{14}$$

where p_j is the probability of correctly answering the score of which is one divided by the number of response options. The expectation and variance for random variable W_j ; $j = 1,2,\dots,l$ is formed based on (2), i.e.,

$$\begin{aligned} E[W_j] &= m_j \cdot p_j & ; j = 1,2,\dots,l \\ Var(W_j) &= m_j \cdot p_j \cdot (1 - p_j) & ; j = 1,2,\dots,l \end{aligned} \tag{15}$$

Suppose for Test II there are l test groups, all of which are multiple choice tests with their number of questions being $m_1, m_2, \dots, m_l$ and for every test group, the score for each question is $b_1, b_2, \dots, b_l$, with $\sum_{j=1}^{l} b_j m_j = 100$. Suppose the random variable Z signifies the total score obtained, where

$$Z = \sum_{j=1}^{l} b_j W_j \quad ; b_j > 0 \ ; w_j = 0,1,2,\dots,m_j \tag{16}$$

Assuming that every random variable is inter-independent, then

$$E[Z] = E\left[\sum_{j=1}^{l} b_j W_j\right] = \sum_{j=1}^{l} b_j E[W_j] = \sum_{j=1}^{l} b_j m_j p_j \tag{17}$$

$$Var(Z) = Var\left(\sum_{j=1}^{l} b_j W_j\right) = \sum_{j=1}^{l} b_j^2 Var(W_j) = \sum_{j=1}^{l} b_j^2 m_j p_j \left(1 - p_j\right) \tag{18}$$

Suppose for Test II, there are two test groups, namely Test II Group 1 and Test II Group 2, each of which is stated as a random variable with binomial distributions W_1 and

W_2. Assuming the random variables W_1 and W_2 are inter-independent, then the probability for participants to score z or more can be calculated using the survival function in (5), i.e.,

$$P(Z > z) = P(W_1 > z) + P(W_2 > z) \quad (19)$$

where $0 \leq z \leq 100$.

If Test II uses l test groups and each test have a binomial distribution and is inter-independent, then the probability for participants to score z or more is

$$P(Z > z) = \sum_{j=1}^{l} P\left(W_j > z\right) \quad (20)$$

Basically, Test II is the replication of Test I, thus the way to construct Test II is exactly the same as constructing Test I. If the admission test requires the use of two or more test types, then Tests III and so forth are constructed using the same way as in constructing Tests I and II.

3.2 Score v and Choosing Set of Tests

Suppose the admission selection is determined by two tests, i.e., Tests I and II. For example, Test I is Mathematics test and Test II is English test. The mathematics test score is stated as random variable $Y$and the English test score is stated as random variable Z, each has binomial distribution. A participant is declared passing the test if they score greater than v with $0 \leq v \leq 100$. for Test I (Mathematics) and Test II (English). Assuming the random variables Y and Z, are inter-independent, then the probability of this student to pass the test is

$$P(Y > v\,, Z > v) = P(Y > v) \cdot P(Z > v) \quad (21)$$

If the admission selection uses t test types and each test has binomial distribution which is inter-independent to each other, then the probability of this student to pass the test is:

$$P(T_1 > v\,, T_2 > v\,, \ldots\ldots, T_t > v) = P(T_1 > v) \cdot P(T_2 > v) \cdot \ldots \cdot P(T_t > v) \quad (22)$$

In this case, the score v can be chosen to allow the used set of tests to minimize the probability of admitting low-performing participants. For the same set of tests, increasing the score v will lessen the probability of admitting low-performing participants. For the same score v, the most accurate set of tests can be determined to filter the admitted participants. The analysis to select the set of tests is provided in section 3.4.

3.3 Example of Application 1

In an admission test, two test types are given, namely Tests I and II. Test I has two question groups. Every group has multiple choice questions with 5 response options, hence the probability to answer correctly for every question is $p = \frac{1}{5}$. Suppose for Test I, Group 1 consists of 10 questions with the score for each question being 5. Group 2 consists of 5 questions with the score for each question being 10. Thus, $a_1 = 5$ and $a_2 = 10$. Suppose X_1 is the random variable of the number of correct responses in Group 1, and X_2 is the random variable of the number of correct responses in Group 2. Assuming X_1 and X_2 have binomial distribution respectively with $n_1 = 10, n_2 = 5$ and $p_1 = p_2 = p = \frac{1}{5}$ and X_1 and X_2 are inter-independent. The probability mass function for the random variables X_1 and X_2 refers to (9), and provided as equations (23) and (24).

$$P(X_1 = x_1) = \binom{10}{x_1}\left(\frac{1}{5}\right)^{x_1}\left(\frac{4}{5}\right)^{10-x_1} \quad ; x_1 = 0,1,2,\ldots,10 \tag{23}$$

$$P(X_2 = x_2) = \binom{5}{x_2}\left(\frac{1}{5}\right)^{x_2}\left(\frac{4}{5}\right)^{5-x_2} \quad ; x_2 = 0,1,2,\ldots,5 \tag{24}$$

Suppose Y is the random variable in (7) which indicates the score obtained by participants in Test I, where

$$Y = a_1X_1 + a_2X_2 = 5x_1 + 10x_2 \tag{25}$$

where $0 \leq y \leq 100$. The score $Y = y$ is obtained as the sum of scores obtained from Groups 1 and 2. For Group 1, the scores possibly obtained are $5x_1$; $x_1 = 0,1,2,\ldots,10$ For Group 2, the score possibly obtained are $10x_2$; $x_2 = 0,1,2,\ldots,5$.

For example, based on (25) the score $y = 15$ is obtained as the sum of score 5 in Group 1 and score 10 in Group 2 or score 15 in Group 1 and score 0 in Group 2. The probability of obtaining score $y = 15$ is calculated using (23) and (24).

$$P(Y = 15) = P(X_1 = 1)P(X_2 = 1) + P(X_1 = 3)P(X_2 = 0)$$

$$= \binom{10}{1}\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^9\binom{5}{1}\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^4 + \binom{10}{3}\left(\frac{1}{5}\right)^3\left(\frac{4}{5}\right)^7\binom{5}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^5$$

$$= 0{,}175922$$

Based on equation (25) where $y = 5x_1 + 10x_2; x_1 = 0,1,2,\ldots,10, x_2 = 0,1,2,\ldots,5$ the scores possibly obtained for the random variable Y is Y = 0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, and 100. Table 4 presents the score of probability mass function as calculated using equations (23) and (24), the cumulative distribution function as calculated using (3) and the survival function as calculated using (5).

$$P(Y \geq 30) = 1 - P(Y < 30) = 1 - [P(Y = 0 \cup Y = 5 \cup \dots . \cup Y = 25)]$$
$$= 1 - [P(Y = 0) + P(Y = 5) + \cdots .. + P(Y = 25)]$$
$$= 1 - 0{,}762649$$
$$= 0{,}237351$$

$$P(Y \geq 40) = 1 - P(Y < 40) = 1 - 0{,}933933 = 0{,}066067$$

$$P(Y \geq 50) = 1 - P(Y < 50) = 1 - 0{,}988161 = 0{,}011839$$

Table 4. Frequency Distribution for Test I Score

$\boldsymbol{y}$	$\boldsymbol{P(Y = y)}$	$\boldsymbol{P(Y \leq y)}$	$\boldsymbol{P(Y \geq y)}$
0	0.035184	0.035184	1
5	0.087961	0.123145	0.964816
10	0.142937	0.266082	0.876855
15	0.175922	0.442004	0.733918
20	0.174547	0.616551	0.557996
25	0.146098	0.762649	0.383449
30	0.105227	0.867876	0.237351
35	0.066057	0.933933	0.132124
40	0.036467	0.970400	0.066067
45	0.017761	0.988161	0.029600
50	0.007634	0.995795	0.011839
55	0.002890	0.998685	0.004205
60	0.000957	0.999642	0.001315
65	0.000275	0.999917	0.000358
70	0.000067	0.999984	8.3×10^{-5}
75	0.000014	0.999998	1.6×10^{-5}
80	0.000002	1	2×10^{-6}
85	3×10^{-7}	1	0
90	4×10^{-8}	1	0
95	1×10^{-9}	1	0
100	3×10^{-11}	1	0

Based on Table 4, the following results are obtained. These indicate that the greater the score$y$taken, the more accurate the Test I in selecting the admitted participants. In score$y = 50$, 11,839 out of 1,000,000 participants or 12 out of 1000 participants passed the test by luckily guessing the answers randomly. Statistically speaking, admitting 12 low-performing participants (for passing because of luckily guessing the answers randomly) out of 1000 participants is still a fair result. The admitted participants' performance can be improved by increasing the passing grade. Setting the score $y = 80$ only generates 2 participants who pass the test by guessing the answers randomly out of 1,000,000 participants.

In regard to this explanation, if the admission selection only has one test type, then the passing grade should be made higher, for example by a minimum score of 80. When two or more sets of tests are used, the passing grade can be maintained at 50. Further analysis is provided in Table 6.

Based on formulation (6) for expectation and variance,

$$E[X_1] = n_1 \times p_1 = 10 \times \frac{1}{5} = 2$$

$$Var(X_1) = n_1 \times p_1 \times (1 - p_1) = 10 \times \frac{1}{5} \times \frac{4}{5} = \frac{8}{5}$$

$$E[X_2] = n_2 \times p_2 = 5 \times \frac{1}{5} = 1$$

$$Var(X_2) = n_2 \times p_2 \times (1 - p_2) = 5 \times \frac{1}{5} \times \frac{4}{5} = \frac{4}{5}$$

Note equation (25), i.e., $Y = 5X_1 + 10X_2$. Using equations (10) or (17) and (11) or (18), the expectation and variance scores for random variable Y are:

$$E[Y] = E[5X_1 + 10X_2] = E[5X_1] + E[10X_2] = 5E[X_1] + 10E[X_2] = 10 + 10 = 20$$

$$E[Y] = \sum_{i=1}^{k} a_i n_i p_i = \left(5 \times 10 \times \frac{1}{5}\right) + \left(10 \times 5 \times \frac{1}{5}\right) = 20$$

$$Var(Y) = Var(5X_1 + 10X_2) = 25Var(X_1) + 100Var(X_2) = 120$$

$$Var(Y) = \sum_{i=1}^{k} a_i^2 n_i p_i (1 - p_i) = \left(25 \times 10 \times \frac{1}{5} \times \frac{4}{5}\right) + \left(100 \times 5 \times \frac{1}{5} \times \frac{4}{5}\right) = 120$$

The mean is equal 20 means that every student will score 20. From the variance 120, a standard deviation = 10.95 is obtained. If the 3 sigma criterion is used, a mean plus 3 sigma = 20 + 3(10.95) = 52.85 is obtained. This result can be used as a benchmark that this test can select fairly if the lowest score is 52.85. Based on the previous analysis, the use of a minimum score of 50 only gives a probability of 12 low-performing participants out of 1000 participants to pass the test. This is supported by the analysis result with a mean and standard deviation which suggests the use of a minimum score of 52.85 as the passing grade.

Test II only consists of one question group with 50 questions and the score for each question is 2. Each of these multiple choice questions have four response options, thus the probability of answering them correctly is $p = \frac{1}{4}$. Suppose W is the random variable which suggests the number of correct responses for 50 questions in Test II. The random variable W has a binomial distribution and is stated as $W \sim bin\left(m = 50, p = \frac{1}{4}\right)$, thus it has a probability mass function

$$P(W = w) = \binom{50}{w}\left(\frac{1}{4}\right)^{w}\left(\frac{3}{4}\right)^{50-w} \quad ; w = 0,1,2,\ldots,50 \tag{26}$$

Suppose the random variable Z suggests the score that participants will obtain in Test II where $Z = 2w$. The score $Z = 20$ is obtained for $W = 10$, thus

$$P(Z = 20) = P(W = 10) = \binom{50}{10}\left(\frac{1}{4}\right)^{10}\left(\frac{3}{4}\right)^{40} = 0{,}098518 \tag{27}$$

Table 5. Frequency Distribution for Test II Score

z	$P(Z = z)$	$P(Z \le z)$	$P(Z \ge z)$	z	$P(Z = z)$	$P(Z \le z)$	$P(Z \ge z)$
0	0.000001	0.000001	1	24	0.129368	0.510986	0.618382
2	0.000009	0.000010	0.999999	26	0.126050	0.637036	0.489014
4	0.000077	0.000087	0.999990	28	0.111044	0.748080	0.362964
6	0.000411	0.000498	0.999913	30	0.088836	0.836916	0.251920
8	0.001610	0.002108	0.999502	32	0.064776	0.901692	0.163084
10	0.004938	0.007046	0.997892	34	0.043184	0.944876	0.098308
12	0.012345	0.019391	0.992954	36	0.026390	0.971266	0.055124
14	0.025865	0.045256	0.980609	38	0.014816	0.986082	0.028734
16	0.046341	0.091597	0.954744	40	0.007655	0.993737	0.013918
18	0.072087	0.163684	0.908403	42	0.003645	0.997382	0.006263
20	0.098518	0.262202	0.836316	44	0.001602	0.998984	0.002618
22	0.119416	0.381618	0.737798	46	0.000650	0.999634	0.001016

Table 5. (Continue) Frequency Distribution for Test II Score

z	$P(Z = z)$	$P(Z \leq z)$	$P(Z \geq z)$	z	$P(Z = z)$	$P(Z \leq z)$	$P(Z \geq z)$
48	0.000244	0.999878	0.000366	62	3 x 10^{-8}	1	0
50	0.000084	0.999962	0.000122	64	6 x 10^{-9}	1	0
52	0.000027	0.999989	3.8 x 10^{-5}	66	1 x 10^{-9}	1	0
54	0.000008	0.999997	1.1 x 10^{-5}	68	0	1	0
56	0.000002	0.999999	3 x 10^{-6}	.	0	1	0
58	0.000001	1	1 x 10^{-6}	.	0	1	0
60	1 x 10^{-7}	1	0	100	0	1	0

Based on Table 5 is obtained:

$$P(Z \geq 30) = 1 - P(Z < 30) = 1 - [P(Y = 0 \cup Y = 2 \cup \ldots . \cup Y = 28)]$$
$$= 1 - [P(Y = 0) + P(Y = 2) + \cdots .. + P(Y = 28)]$$
$$= 1 - 0{,}748080 = 0{,}251920$$
$$P(Z \geq 40) = 1 - P(Z < 40) = 1 - 0{,}986082 = 0{,}013918$$
$$P(Z \geq 50) = 1 - P(Z < 50) = 1 - 0{,}999878 = 0{,}000122$$

Based on Table 5, in score$y = 50$, 122 participants pass the test by randomly guessing out of 1,000,000 participants. This means the set of Test II is more accurate in selecting the passing participants than the set of Test I. Even at score$y = 60$, the probability for participants to pass the test by randomly guessing is zero.

The expectation and variance scores of random variable W are:

$$E[W] = m \times p = 50 \times \frac{1}{4} = 12{,}5$$
$$Var(W) = m \times p \times (1 - p) = 50 \times \frac{1}{4} \times \frac{3}{4} = \frac{150}{16}$$

The expectation and variance scores of random variable Z are:

$$E[Z] = E[2W] = 2E[W] = 2 \times 12{,}5 = 25$$
$$Var(Z) = Var(2W) = 4Var(W) = 4 \times \frac{150}{16} = \frac{150}{4}$$

The Test II mean is greater than Test I. The standard deviation for Test II is 6.12. If the 3 sigma criterion is used, then the mean plus 3 sigma = 25 + 3(6.12) = 43.36 is obtained. This result can be used as a benchmark that this test can select fairly if the lowest score is 43.36. Based on the previous analysis, the use of a minimum score of 40 only gives a probability for 13,918 low-performing participants out of 1,000,000 participants to pass the test. At score 50, 122 low-performing participants pass the test out of 1,000,000 participants. Referring to Table 5, the closes score to 43.36 is 44 and only 2,618 low-performing participants pass the test out of 1,000,000 participants. Hence, the analysis with the survival function and mean supports each other.

Referring to (21), choosing$v = 50$**,** the probability to score 50 or more in Tests I and II is

$$P(Y \geq 50) \times P(Z \geq 50) \ = 0{,}011839 \times 0{,}000122 = 0{,}000001$$

If the passing grade is 50 in Test I and 50 in Test II, then the probability for participants to pass the test merely by relying on lucky guess is 0.000001. This means only one out of 1,000,000 participants passes the test merely by guessing the answers. This result provides adequate confidence that the combination of multiple choice tests can select participants well and fairly. Only participants with extraordinary ability can pass this test. The use of other and different pairs of scores v is provided in Table 6. In Table 6, the passing probability for scores 60, 70, and 80 in Test II is not included and the value is 0.

Table 6. The Probabilty to Pass on Test I dan II with Minimum Score v

Test I Score v	**Probability Test I**	**Test II Score v**	**Probability Test II**	**Probabilty to Pass**
50	0.011839	50	0.000122	0.000001
		60	0	0
		70	0	0
		80	0	0
60	0.001315	50	0.000122	1.6×10^{-7}
70	8.3×10^{-5}	50	0.000122	1.0×10^{-8}
80	2×10^{-6}	50	0.000122	2.4×10^{-10}

3.4 Example of Application 2

Suppose in Test I the questions are grouped into three by their low, medium, and high levels of difficulty. The higher the level of difficulty, the lesser the number of questions, yet the greater the score for each item will be. For example:

K1: Questions with low level of difficulty consists of 12 items with the score for each item being 3.

K2: Questions with medium level of difficulty consists of 10 items with the score for each item being 4.

K3: Questions with high level of difficulty consists of 3 items with the score for each item being 8.

The total number of questions is 25 multiple choice questions with 5 response options and a total score of 100. The total scores for questions with low, medium, and high levels of difficulty are 36, 40 and 24 respectively.

Suppose

Y : the score obtained by participants in Test I

X_1 : the number of correct responses for 12 items worth 3 score points in Group 1

X_2: the number of correct responses for 10 items worth 4 score points in Group 2

X_3: the number of correct responses for 3 items worth 8 score points in Group 3

The random variablesX_1, X_2, X_3 have binomial distributions each with $p = \frac{1}{5}$.

The probability density function for each dsitributions is

$$P(X_1 = x_1) = \binom{12}{x_1}\left(\frac{1}{5}\right)^{x_1}\left(\frac{4}{5}\right)^{12-x_1} \quad ; x_1 = 0,1,2,\ldots,12$$

$$P(X_2 = x_2) = \binom{10}{x_2}\left(\frac{1}{5}\right)^{x_2}\left(\frac{4}{5}\right)^{10-x_2} \quad ; x_2 = 0,1,2,\ldots,10$$

$$P(X_3 = x_3) = \binom{3}{x_3}\left(\frac{1}{5}\right)^{x_3}\left(\frac{4}{5}\right)^{3-x_3} \quad ; x_3 = 0,1,2,3.$$

The score points for each item in the three groups are known to be 3, 4 and 8, thus the score acquisitions for Mathematics test is

$$Y = 3X_1 + 4X_2 + 8X_3 \;\; ; x_1 = 0,1,2,\ldots,12 \;\; ; \; x_2 = 0,1,2,\ldots,10 \;\; ; \; x_3 = 0,1,2,3.$$

From $y = 3x_1 + 4x_2 + 8x_3$ the values for the random variable Y are y = 0, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, …., 99, 100.

Suppose the questions in Test II are grouped into two by their low and high levels of difficulty. For example:

K1 : Questions with low level of difficulty consist of 30 items worth 2 score points

K2 : Questions with high level of difficulty consist of 10 items worth 4 score points.

The total number of questions is 40 multiple choice questions with 4 response options and a total score of 100. The total score points for items with low and high levels of difficulty are 60 and 40 respectively.

Suppose that,

Z : the score obtained by participants in English test

W_1 : the number of correct responses for 30 items worth 2 score points in Group 1

W_2 : the number of correct responses for 10 items worth 4 score points in Group

The random variables W_1, W_2 have binomial distribution each with $p = \frac{1}{4}$. The probability density function for each dsitributions is

$$P(W_1 = w_1) = \binom{30}{w_1}\left(\frac{1}{4}\right)^{w_1}\left(\frac{3}{4}\right)^{30-w_1} \quad ; w_1 = 0,1,2,\ldots,30$$

$$P(W_2 = w_2) = \binom{10}{w_2}\left(\frac{1}{4}\right)^{w_2}\left(\frac{3}{4}\right)^{10-w2} \quad ; w_2 = 0,1,2,\ldots,10$$

The score points for each item in the two groups are known to be 2 and 4, thus the score acquisition for Mathematics test is

$$Z = 2W_1 + 4W_2 \quad ; w_1 = 0,1,2,\ldots.,30 \quad ; \quad w_2 = 0,1,2,\ldots.,10.$$

From $z = 2w_1 + 4w_2$ the values of the random variable Z is z = 0, 2, 4,…., 98, 100.

Referring to (21), participants are declared passing the test if they score a minimum of $v = 60$ in both tests. The probability to score 50 or more in Tests I and II is

$$P(Y \geq 60) \times P(Z \geq 60) = [1 - P(Y < 60)] \times [1 - P(Z < 60)] \tag{28}$$

For the first term on the right side from (28):

$$P(Y < 60) = P(Y = 0) + P(Y = 3) + P(Y = 4) + P(Y = 6) + P(Y = 7) + \cdots. + P(Y = 59)$$

$$P(Y = 0) = P(X_1 = 0, X_2 = 0, X_3 = 0) = P(X_1 = 0) \times P(X_2 = 0) \times P(X_3 = 0)$$

$$= \binom{12}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^{12}\binom{10}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^{10}\binom{3}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^3$$

$$= \binom{12}{0}\binom{10}{0}\binom{3}{0}\left(\frac{4}{5}\right)^{25}$$

$$P(Y = 3) = P(X_1 = 1) \times P(X_2 = 0) \times P(X_3 = 0)$$

$$= \binom{12}{1}\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^{11}\binom{10}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^{10}\binom{3}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^3$$

$$= \binom{12}{1}\binom{10}{0}\binom{3}{0}\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^{24}$$

$$P(Y = 4) = P(X_1 = 0) \times P(X_2 = 1) \times P(X_3 = 0)$$

$$= \binom{12}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^{12}\binom{10}{1}\left(\frac{1}{5}\right)^1\left(\frac{4}{5}\right)^9\binom{3}{0}\left(\frac{1}{5}\right)^0\left(\frac{4}{5}\right)^3$$

$$= \binom{12}{0}\binom{10}{1}\binom{3}{0}\left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^{24}$$

The continuous calculation is$P(Y = 6)$, $P(Y = 7)$,….., $P(Y = 58)$, and $P(Y =$

59). For the second term on the right side from (28):

$$P(Z < 60) = P(Z = 0) + P(Z = 2) + P(Z = 4) + \cdots . + P(Y = 58)$$

$$P(Z = 0) = P(W_1 = 0, W_2 = 0) = P(W_1 = 0) \times P(W_2 = 0)$$

$$= \binom{30}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^{30}\binom{10}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^{10} = \binom{30}{0}\binom{10}{0}\left(\frac{3}{4}\right)^{40}$$

$$P(Z = 2) = P(W_1 = 1, W_2 = 0) = P(W_1 = 1) \times P(W_2 = 0)$$

$$= \binom{30}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^{29}\binom{10}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^{10} = \binom{30}{1}\binom{10}{0}\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^{39}$$

$$P(Z = 4) = P(W_1 = 0, W_2 = 1) + P(W_1 = 2, W_2 = 0)$$

$$= P(W_1 = 0) \times P(W_2 = 1) + P(W_1 = 2) \times P(W_2 = 0)$$

$$= \binom{30}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^{30}\binom{10}{1}\left(\frac{1}{4}\right)^1\left(\frac{3}{4}\right)^{9} + \binom{30}{2}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^{28}\binom{10}{0}\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^{10}$$

$$= \binom{30}{0}\binom{10}{1}\left(\frac{1}{4}\right)\left(\frac{3}{4}\right)^{39} + \binom{30}{2}\binom{10}{0}\left(\frac{1}{4}\right)^2\left(\frac{3}{4}\right)^{38}$$

The continuous calculation is $P(Z = 6)$,$P(Z = 8)$,….., $P(Z = 56)$, and $P(Z = 58)$. Then, used the equation (28) for finalized calculation process. The result for this calculation is not presenting in this paper.

3.5 Performance of Set of Tests

The performance of set of tests is measured per example of application. In Example of Application 1, the passing is determined by two tests. Based on Table 7, the number of questions in both tests is different and the response options provided is also different. The mean of Test II is greater than Test I. However, the standard deviation of Test II is smaller than Test I. This means Test II is more homogeneous than Test I. It is also characterized by the lower result of mean calculation plus 3 times standard deviation for Test II. Based on the probability to score more than a certain point, Test II gives far lesser probability. The whole analysis leads to a conclusion that Test II is stronger in selecting the passing participants.

Table 7. Statistic for Multiple Choice Test for Example of Application 1

Pass Examination	Test I	Test II
Number of Multiple Choice Test	15	40
Number of Alternative Answers	5	4
Mean	20	25
Deviation Standard	10.95	6.12

Table 7. (Continue) Statistic for Multiple Choice Test for Example of Application 1

Pass Examination	Test I	Test II
Mean plus 3 times Deviation St.	52.85	43.36
Probability with score greater than 10	87.69%	99.79%
Probability with score greater than 20	55.80%	83.63%
Probability with score greater than 30	23.74%	25.19%
Probability with score greater than 40	6.61%	1.39%
Probability with score greater than 50	1.18%	0.0122%
Probability with score greater than 60	0.13%	0%
Probability with score greater than 70	0.000083%	0%
Probability with score greater than 80	0.000002%	0%
Probability with score greater than 90	0%	0%

The performance of set of tests is measured per example of application. In Example of Application 1, the passing is determined by two tests. Based on Table 6, Test II is more dominant in its ability to select the passing participants. In Test I, the probability of passing for passing grades 50, 60, 70, and 80 still has positive points. Meanwhile, the probability for Test II is 0, except for the passing grade 50 the probability of which still has positive point. The 0 probability in Test II leads the passing probability to have been 0 at passing grade 60.

Table 8. Statistic for Multiple Vhoice Test for Example of Application 2

Pass Examination	Test I	Test II
Number of Multiple Choice Test	25	40
Number of Alternative Answers	5	4
Mean	28	25
Standard Deviation	11.17	7.25
Mean plus 3 times Standard Deviation	61.51	46.75
Probability with score greater than 10	-	-
Probability with score greater than 20	-	-
Probability with score greater than 30	-	-
Probability with score greater than 40	-	-
Probability with score greater than 50	-	-

Table 8. (Continue) Statistic for Multiple Vhoice Test for Example of Application 2

Pass Examination	Test I	Test II
Probability with score greater than 60	-	-
Probability with score greater than 70	-	-
Probability with score greater than 80	-	-
Probability with score greater than 90	-	-

In Table 8, the amount of probability to score more than a certain point is not included. The calculation to obtain it is lengthy. The analysis is based only on the mean, standard deviation and data distribution values. Test I gives greater mean and standard deviation, thus the data distribution which is measured plus 3 times the standard deviation is also greater. This indicates that Test II is more homogeneous and sharper in measuring the participants' performance. From this result, it is assumed that Test II is stronger or more dominant in selecting the admitted participants than Test I.

4 Conclusion

In general, the illustration in Example of Application 1 gives a result that setting the passing grade at 50 allows only 1 low-performing participant out of 1,000,000 participants to pass the test for answering the questions correctly merely by randomly guessing it. The admission test in Example of Application 1 consists of two test types. Partially, Test II is more selective in filtering the passing participants. This is characterized by the fact that 11,839 out of 1,000,000 participants pass the test merely by guessing it, and only 122 in for test II at passing grade 50. If the passing grade is increased to 60, 1,315 out of 1,000,000 participants still pass the test merely by guessing it for Test I, and none for Test II. This means, at passing grade 60, when only Test II is used, all the admitted participants are truly those with the expected ability.

References

[1] Jannah, R., Hidayat, D.N., Husna, N., & Khasbani, I. (2021). An item analysis on multiple-choice questions: a case of a junior high school English try out test in Indonesia. *Leksika*, 15, 1, 9-17.

[2] Poláčková, J., & Svatošová, L. (2013). Relationship of success in university study and admission exam results. *Journal on Efficiency and Responsibility in Education and Science*, 6, 4, 281-293.

[3] Bartoška, J., Brožová, H., Šubrt, T., & Rydval, J. (2013). Incorporating practitioners' expectations to project management teaching. *Proceedings of the 10th International Conference*, 1, 1, 16–23.

[4] Flégl, M. (2014). Comparison of research engagement of Ph.D. students at various study programs at CULS Prague: an introductory study. *Journal on Efficiency and Responsibility in Education and Science*, 7, 3-4, 66-73.

[5] Hrubý, M. (2016). Feedback improvement of question objects. *International Journal of Continuing Engineering Education and Lifelong Learning*, 26, 2, 183– 195.

[6] Otavová, M. & Sýkorová, I. (2014). Analysis of scores from mid-term and final test by a contingency table. *in Efficiency and Responsibility Education Proc. of the 11th Int. Conf.*, pp. 527–533, Sept. 3-7, 2014.

[7] Otavová, M., & Sýkorová, I. (2016). Differences in results obtained by students of different faculties. *Journal on Efficiency and Responsibility in Education and Science*, 9, 1, 1-6.

[8] Krejčí, I., Kvasnička, R., & Dömeová, L. (2011). Introducing system dynamics at CULS Prague. *Journal on Efficiency and Responsibility in Education and Science*, 4, 4, 187-196.

[9] Brožová H., & Rydval J. (2014). Analysis of exam results of the subject´applied mathematics for IT. *Journal on Efficiency and Responsibility in Education and Science*, 7, 3-4, 59-65.

[10] Flégl, M. (2013). Innovation of doctoral studies at the FEM CULS Prague. *Journal on Efficiency and Responsibility in Education and Science*, 6(4), 265-280, 2013.

[11] Ječmínek, J., Kukalová, G., Moravec, L., & Filipová. D. B. (2018). Tax courses exams result at FEM CULS Prague evaluation. *Proceedings of the 15th International Conference*, 1, 1, 132–139.

[12] Zvára, K., & Anděl, J. (2001). The relationship between admission results and academic success at the Faculty of Mathematics and Physics. *Advances in Mathematics, Physics and Astronomy*, *46*(6), 304–312.

[13] Wahyuni, D.A.S., Kumala, F.Z., dan Prabowo, A. (2024). Rekomendasi pemilihan tes pilihan ganda menggunakan analisis kesintasan. *Skewness: Jurnal Statistika, Aktuaria, dan Sains Data*, 1, 82-99.

[14] Farthing, D. W., Jones, D. M., & McPhee, D. (1998). Permutational multiple-choice questions: An objective and efficient alternative to essay-type examination.

Proceedings of the 3rd Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE '98), Dublin, Ireland, 81–85.

[15] Premadasa, I. (1993). A reappraisal of the use of multiple-choice questions. *Medical Teacher*, *15*(2–3), 237–242.

[16] Zhao, Y. (2005). Algorithms for converting raw scores of multiple-choice question tests to conventional percentage marks. *International Journal of Engineering Education*, *21*(6), 1189–1194.

[17] Zhao, Y. (2006). How to design and interpret a multiple-choice-question test: A probabilistic approach. *International Journal of Engineering Education*, *22*(6), 1281–1286.

[18] Klůfa, J. (2015c). Comparison of the ways of acceptance students at university. *Journal on Efficiency and Responsibility in Education and Science*, *8*(3), 72–76.

[19] Klůfa, J. (2014). Multiple choice question tests: Advantages and disadvantages. *Mathematics and Computers in Science and Industry*, *10*(2), 91–94.

[20] Klůfa, J. (2015a). Dependence of the results of entrance examinations on test variants. *Procedia - Social and Behavioral Sciences*, *174*(1), 3565–3571.

[21] Klůfa, J. (2015b). Analysis of entrance examinations: Efficiency and responsibility in education. In *Proceedings of the 12th International Conference on Efficiency and Responsibility in Education and Science* (pp. 250–256).

[22] Esmaeeli, B., Shandiz, E. E., Noriiziasl, S., Shojaei, H., Pasandideh, A., Khoshkholgh, R., Fazli, B., & Ahmadi, F. B. (2021). The optimal number of choices in multiple-choice tests: A systematic review. *Medical Educational Bulletin*, *2*(3), 253–260.

[23] Klůfa, J. (2018). Probability aspects of the multiple-choice question tests. *International Journal of Economics and Management Systems*, *3*(1), 60–64.

[24] Klůfa, J. (2020). Probability aspects of entrance exams at university. *Mathematics and Statistics*, *8*(2), 142–151.

[25] Walpole, R. E., & Myers, R. H. (1986). *Ilmu peluang dan statistika untuk insinyur dan ilmuwan* [Probability and statistics for engineers and scientists]. Penerbit ITB.