International Journal of Business Analytics and Intelligence
7 (1) 2019, 46-52
http://publishingindia.com/ijbai/

Submitted: 09 October, 2018
Accepted: 29 December, 2018

The Statistical Analysis of Examination and Evaluation of Results of International Trade Course

Samik Shome*, Urmi Khatri**

Abstract

The statistical analysis of any examination results are important the professors to understand about the magnitude of knowledge students have incurred from their courses. Based on the examination results of International Trade course of International Business specialization students in the third term of MBA 2017-2019 batch of IMS Unison University, the quantitative analysis for several parameters including difficulty, discrimination, reliability and ANOVA are investigated. The results indicate that the distribution of examination scores approximate to normal distribution. The results stated that the exam paper belongs to a moderate level, which was then qualified by the discrimination and reliability tests done in the study. Thus it was concluded that the design of the examination paper was good and dependable.

Keywords: International Trade, Difficulty Index, Discrimination Index, Reliability Test, ANOVA

Introduction

The analysis of the examination results is done by any faculty or professor, interchangeably used in this paper, to know the understanding and grasping ability of the student for a particular subject. On the basis of the scores, student's conceptual clarity of that particular subject can be judged. In order to gain the most benefit from examination, faculty at each institution need to develop their own understanding of the process (Palomba and Banta, 1999). According to Hatchings and Marches (1990), the meaning of examination for a college graduate is best captured by responding to some fundamental questions: (a) What should they know or be able to do and value? (b) Have the graduates of our institutions acquired this learning? (c) What are the contributors of the institution and its programs to student growth? (d) How can student learning be improved? and, (e) When individuals involved in assessment become confused about its purposes, it helps to return to these questions.

Nowadays, in several colleges and universities, the examination analysis has been developed as an exercise after each examination. This exercise helps to identify the trend of marks the class received and its reviewing and reflection on academic standard. This exercise can be done by using statistical techniques and models on the results of the students. In other words, the analysis of the results helps the professors to understand the students' knowledge of a particular subject and also the quality of paper prepared for examination. This type of analytical exercise can serve two purposes simultaneously: first, it helps to know the ability of the students to understand the subject; and, secondly, it also helps to modify the examination paper, if required, and make it more adequate and proficient next time. Hence, through this work out, one gets the loop holes of the teaching method and examination system or pattern of the college.

In this backdrop, the central argument of this paper is to analyze the quality of a question paper in terms of its design and standard. The analysis and conclusion drawn are established on the basis of the marks students received in the course.

* Associate Professor, Institute of Management, Nirma University, Ahmedabad, Gujarat, India. Email: samik@nirmauni.ac.in
** Doctoral Scholar, Institute of Management, Nirma University, DPM Cubicles, Ahmedabad, Gujarat, India.
Email: urmi.khatri@nirmauni.ac.in

Literature Review

Basic statistical analysis of examination results is a method typically used nowadays in many universities as a part of academic audit and requirement for accreditation purpose. However, more detailed analysis can be done individually by professors to examine their realizations of expectations after teaching a course. Every course a faculty teaches has its own preferences and hopes that create its own view. Several researchers and scholars in the last few years have proposed means and the underlying analytical framework behind the learning pattern of students from different courses. One of the pioneering research in this area by Tversky and Kahneman (1981) discussed that the concept of framing effect the professors made up of their own views of the subject. They also showed how the framing effect influence the way in which the information is interpreted. Individual decisions are influenced by the presented information and by the problems formulation (Druckman, 2001).

Yang, et. al (2013) analyzed the examination results of materials research methods course for 100 students of Wuhan University of Science and Technology in China to evaluate the validity, reliability, difficulty, and discrimination power of the examinations. The study concludes that the design of the examination paper was good and dependable. Kaspříková (2011, 2012) did such analysis for mathematics course at University of Economics in Prague and JarkovskáKučera, Vostrá Vydrová, and Varvažovská (2012) for distance programs at Czech University of Life Sciences, Prague.

Borožová and Rydval (2014) used difficulty index, discrimination index and Cronbach's alpha to study the examination results of Applied Mathematics for Information Technology (IT) for a sample size of 615 students. They observed that the students found the theoretical questions to be the hardest among all the sections in the question paper. Again, Yang, et. al (2013) in analysis of the examination in anesthesiology for medical students used the Rasch model to estimate the parameters which creates a difficulty in tests for students. Similarly, Ramya, et. al (2017) used the Apriori algorithm technique to examine the performance of the students in an examination.

The Study and Its Methodology

The current study is done on a paper of International Trade offered in a two year full time MBA program in IMS Unison University, Dehradun (India). Out of several major streams, including Finance, Marketing and Human Resource Management (HRM), the program also offers a specialization on International Business (IB). The course of International Trade is a part of IB specialization which is offered in the third trimester of first year of the program. For the 2017-19 batch, 75 students opted for this course of International Trade. The examination paper was of 100 marks divided into three parts. The first part consists of five very short type questions, the second part includes five theoretical questions and, finally, the third part consists of two large practical examples. The questions of the tests cover all the topics of the course.

- Five very short type questions – these questions have a form of a brief question that requires a written answer not longer than a two sentences. For example, the students have to write the basic definitions or any two functions or the simple principle. Maximum score of each question is 2 points.
- Five theoretical questions – these questions have a form of description questions, which have to be discussed in details and the process or results have to be interpreted. Maximum score of each example used to be 10 points.
- Two essay or long questions – this part of the paper has a form of a case-study or scenario question, which is used to prove that students can understand and integrate key concepts of the course, apply theory to a practical context, and demonstrate the ability to analyze and evaluate obtained effects. These types of questions are practical in nature and the students need to attempt any two questions out of three. Maximum score used to be 40 points (20 points for each answer).

A summary of the question paper format is provided in Table 1.

Table 1: Question Paper Format of International Trade

Part	Compulsory questions	Total questions	Marks per question	Total marks
Part 1	5 compulsory	5 questions	2 marks each	10
Part 2	5 compulsory	5 questions	10 marks each	50
Part 3	2 compulsory out of 3	2 questions	20 marks each	40
Part (1+2+3)		12 questions		100

Source: Question paper of International Trade of IMS Unison University

To analyze the examination results, the data was used from student actual scores; the data was collected on number of students, their enrolment numbers and the numbers they scored in the examination. Now to identify the test quality and the level of difficulty the students faced in answering the question paper of International Trade, this paper followed the techniques used by Jacobs (1991), Miller (2012), Wells and Wollack (2003) and Yang, et. al (2013):

- Difficulty Index of the test,
- Discrimination Index of the test,
- Reliability of the test.

In addition, this paper also used the Analysis of Variance (ANOVA) to check the difference in the average marks of the students.

Difficulty Index

Any examination should not be either too easy or too tough. The question paper is required to prepare by keeping two objectives in mind: (a) the examination should be able to assess student's knowledge of the subject; and, (b) it should be able to measure student's knowledge.

The ratio of the average scores to item *i*, to the full scores of item *i*, gives the difficulty index of item *i*. It can also be understood as the proportion of a student who answers the examination question correctly. The formula is as follows:

$$Pi = \frac{Ai}{Ni}$$

where, Pi = Difficulty index of item i, Ai = Average scores to item i, Ni = Full scores of item i for the whole script. The average difficulty index P can be calculated by the formula as below:

$$P = \frac{1}{100}\sum_{i=1}^{N} PiNi$$

Interpretation: P > 0.75 implies exam is easy; P < 0.45 implies exam is difficult; 0.45 < P < 0.75 implies exam is average or passable.

Discrimination Index

Each examination should be able to discriminate between a well prepared student and a not so one. It is expected that the students who are prepared should answer the questions well in the examination. There should be some questions which the not so well prepared students will find difficult to answer. Hence, the discrimination index D gives a variation in the student's performance. In other words, from discrimination index highly intellectual students can be discriminated from less intellectual students.

$$D = \frac{P_H - P_L}{100}$$

where, P_H = Average score for the 27% of students with highest test marks,

P_L = Average score for the 27% of students with lowest marks.

Interpretation: According to R. L. Ebel (1972), if D > 0.39, the quality of the exam paper is excellent. When 0.30 < D < 0.39, the exam paper is qualified. If 0.20 < D < 0.29, it indicates that the quality of the exam paper is passable and has possibility for improvement. The exam paper should be discarded if D is less than 0.20.

Reliability of the Test

Cronbach's alpha is used for estimating the reliability of the test. This alpha was developed by Lee Cronbach (1951) to provide a measure of the internal consistency of a test or scale. Internal consistency describes the extent to which all the items in a test measure the same concept or construct and hence it is connected to the inter-relatedness of the items within the test.

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\in S_I^2}{S_X^2}\right)$$

where, k = Total of item, S_I^2 = Variance of scores for item i, and, $S_{X[}^2$ = Variance of scores for script.

It is also consider as the heart of quality control of examination. Although the value can range between 0 and 1, but it generally falls in the range of 0.60 to 0.80. Table 2 explains the internal consistency with respect to the alpha value.

Table 2: Alpha Value with Respect to Internal Consistency

Internal Consistency	*Cronbach's Alpha*
Excellent	0.9 - 1.0
Good	0.8 - 0.9
Acceptable	0.7 - 0.8
Questionable	0.6 -0.7
Poor	0.5 - 0.6
Unacceptable	0.0 - 0.5

Source: Cronbach (1951)

The examination paper is acceptable if the alpha value is greater than 0.7 and is totally unacceptable if it is below 0.5. The paper is excellent if the alpha value is between 0.9 and 1.

ANOVA

Analysis of variance (ANOVA) is a collection of statistical models used to analyze the differences among group means and their associated procedures (such as 'variation' among and between groups). It is developed by statistician and evolutionary statistician and biologist Ronald Fisher (1918). This paper used ANOVA to find out the difference in the average marks of the students, in all the three parts (i.e. Part 1, Part 2 and Part 3) of the test.

Results and Discussions

The descriptive statistics of the scores of 75 students is presented in Table 3. It can be inferred that a few students scored full marks in Part A. It can also be observed that some of the students could not score any marks in Part A and Part C. One possibility of this phenomenon may be that these students may have left the questions in those parts.

Table 3: Descriptive Statistics

Parameters	*Part A*	*Part B*	*Part C*
Minimum	00	10	00
Median	06	28	28
Mean	5.89	27.65	27.45
Maximum	10	43	36

Source: Calculated by the authors

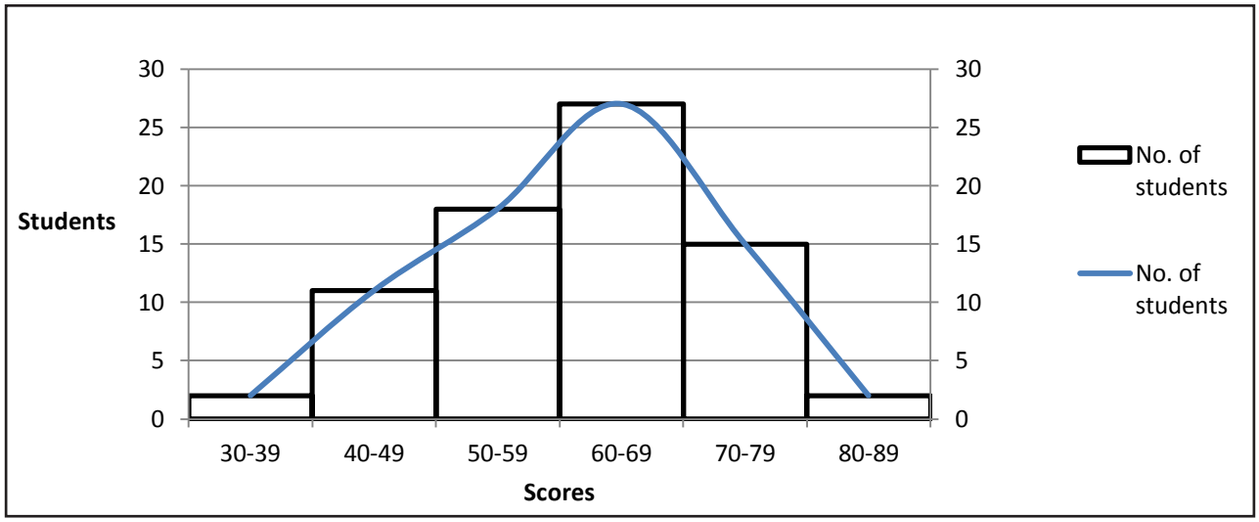
The average marks scored is almost same in Part B and C. However, the maximum score was 10 more in Part C than in Part B. This shows that the students have been able to perform comparatively better in Part C. Thus, it can be inferred that the students were more comfortable in answering essay type questions than the shorter once. The distribution of marks is presented in Table 4.

Table 4: Distribution of Marks

Part A		*Part B*		*Part C*		*Total (Part A + B + C)*	
Marks Interval	*No. of Students*	*Marks Interval*	*No. of Students*	*Marks Interval*	*No. of Students*	*Marks Interval*	*No. of Students*
00	04	00	00	00	01	30 – 39	02
0 – 2	03	0 – 10	01	0 – 10	00	40 – 49	11
2 – 4	10	10 – 20	11	10 – 20	04	50 – 59	18
4 – 6	27	20 – 30	36	20 – 30	48	60 – 69	27
6 – 8	21	30 – 40	25	30 – 40	22	70 – 79	15
8 – 10	10	40 – 50	02	–	–	80 – 89	02
Total	**75**	**Total**	**75**	Total	75	Total	75

Source: Calculated by the authors

Maximum students (27 altogether) have scored between the range of 60-69 marks followed by 15 students between 70-79. There are only two students who scored above 80 marks. However, there are also two students who scored in between 30-39. The distribution of the total marks of all students is depicted in Fig. 1.



Source: Calculated by the authors

Fig. 1: Distribution of Overall Marks

Since the pass mark is 30 out of 100, hence, no students has failed in this subject of International Trade. The results show that the scores are approximate to normal distribution.

Difficulty Index

The difficulty index of each part and the overall question paper of International Trade is presented in Table 5.

Table 5: Difficulty Index of Examination for Part 1, 2, 3 and overall

Particulars	*Part 1*	*Part 2*	*Part 3*	*Total*
Total score	10	50	40	100
Difficulty	0.5893	0.5531	0.6833	0.61
Quality	Average	Average	Average	Average/ Passable

Source: Calculated by the authors

By comparing the different parts of the examination paper, it can be seen that its difficulty index *P* ranges in between 0.55 to 0.69 (Table 5). The easiest part was the essay type or scenario questions in which majority of the students did well. Very short answer questions, as well as the theoretical were the most difficult parts. It implies that although the students were able to relate the basic concepts with the case based practical questions, but it is also evidenced that the ability of students for mastering basic definitions and handle theories was deficient. It shows the students were not able memories definitions and theories and write them but are enough knowledgeable apply the concepts in the practical scenario.

As the overall value of the difficulty index *P* of examination paper is 0.61 it implies that the level of the paper is moderate and therefore it is not difficult for students to pass this paper.

Discrimination Index

The analysis of the discrimination index *D* for the examination results shows that the value is 0.74 (Table 6).

Table 6: Calculation of D value

Particulars	*Value*
P_H	74.2
P_L	46.35
D	0.742666667

Source: Calculated by the authors

According to Ebel's rule, the exam paper is highly qualified and the question paper is designed very well. Although there is always scope for improvement in any examination paper, but this paper also qualifies to be highly acceptable at its current form.

Reliability of the Test

The reliability of the test i.e., Cronbach's alpha is conducted to study the internal consistency of a question paper. The consistency can be achieved by setting a question paper with an appropriate combination of definitions, theories, and practical application questions to determine the level of knowledge of the students. For this paper, the quality is proved to be good as the result of Cronbach's alpha is 0.71.

Table 7: Calculation of Cronbach's Alpha

Particular	*Values*
Variance	39.609
Total variance	130.43243
K	75
Cronbach's α	0.71

Source: Calculated by the authors

ANOVA

The ANOVA test is conducted to find out whether there is any significant difference in average marks scored in the three parts of the examination papers. For that, the scores are converted from absolute value to per cent form and then the single factor ANOVA test is run. The hypothesis of the test is as follows:

H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: Not all means are equal.

The result of the ANOVA test is presented in Table 8.

Table 8: ANOVA of the Exam Statistics of Part A, B and C

Summary						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Part 1	75	4420	58.93333	606.955		
Part 2	75	4148	55.30667	176.5939		
Part 3	75	5147.5	68.63333	184.509		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	7121.069	2	3560.534	11.03405	0.000027	3.036524
Within Groups	71636.28	222	322.6859			
Total	78757.35	224				

Source: Calculated by the authors

*Note: All the scores are normalized to 100 for ANOVA

As seen earlier in other analysis, the average marks are relatively higher in Part 3 i.e. 68.63 compared to the other two parts. The average is lowest in Part 2 which is 55.31 followed by 58.9 in Part A. It can also be seen that the variance in marks scored is more in Part 1.

Since, the decision rule of ANOVA test is: Reject H_0 if, $F > F\ crit$, here $11.03405 > 3.036524$. Therefore, H_0 is rejected. It means that the average scores of the three parts of the paper are not equal. It is found that Part C, the average score is higher than the rest of the parts. One probable reason for the students to score more in Part 3 may be because of the availability of choices in the questions. In Part 3, the students got an opportunity to attempt two questions out of three of 20 marks each. This ultimately had a positive effect on the increase in overall average scores of the examination.

Concluding Observations

The statistical analysis of any examination results is an essential for the professors to understand about the magnitude of knowledge students have incurred from their course. It can also act as a feedback mechanism about the

quality of examination papers in terms of its design and standard. This will help the professors to make necessary changes in the questions and improve the standard of the examination.

This paper studies the results of International Trade examination offered in International Business (IB) specialization in a two year full time MBA program in IMS Unison University, Dehradun (India). The analysis suggests that the distribution of examination scores approximate to normal distribution. Several parameters for the examination paper including difficulty index, discrimination index and reliability were calculated. The values are 0.61, 0.74 and 0.71, respectively. The difficulty of the exam paper belongs to moderate level, therefore it is not difficult for students to pass this examination. Thus, it has qualified both the discrimination and reliability. The ANOVA result shows that Part C of the question paper was relatively more scoring as compared to any other parts of the paper. To conclude, the statistical analysis of the result reveals that the design of the examination paper was good and consistent and the standard is dependable.

References

Beveridge, J. R., She, B. D., & Givens, G. H. (2001). Parametric and nonparametric methods for the statistical evaluation of human Id algorithms. In *Proc. 3rd Workshop on the Empirical Evaluation of Computer Vision Systems*.

Borožová, 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.

Cable, D. M., & Judge, T. A. (1996). Person–organization Fit, Job Choice Decisions, and Organizational Entry. *Organizational Behavior and Human Decision Processes*, *67*(3), 294–311.

Cable, D. M., & Judge, T. A. (1996). Person–organization fit, job choice decisions, and organizational entry. *Organizational Behavior and Human Decision Processes*, *67*(3), 294–311.

Cronbach, L. J. (1951). Coefficient Alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Druckman, J. N. (2001). Evaluating framing effects. *Journal of Economic Psychology*, *22*(1), 91–101. http://dx.doi.org/10.1016/S0167-4870(00)00032-5

Ebel, R. L. (1972). *Essentials of educational measurement*. Upper Saddle River, NJ: Prentice Hall.

Fisher, R. A. (1918). The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, *52*, 399–433.

Hutchings, P., & Marchese, T. (1990). Watching assessment: Questions, stories, prospects. *Change: The Magazine of Higher Learning*, *22*(5), 12–38.

Jacobs, L. C. (1991). *Test Reliability*. Retrieved from http://www.indiana.edu/~best/test_reliability.shtml

Jarkovská, M., Kučera, P., Vostrá Vydrová, H., & Varvažovská, P. (2012). Analysis of students' results in distance-studies centres. *Journal on Efficiency and Responsibility in Education and Science*, *5*(2), 78–91. http://dx.doi.org/10.7160/eriesj.2012.050203