

Implementation of Factor Analysis in Education Industry

Ritu Bhasin¹, Vamsikrishna A²

^{1,2} *Symbiosis Statistical Institute, ICC Trade Tower, Pune, India*

E-mail: ¹ritu.bhasin@ssi.edu.in, ²vamsikrishna.a@ssi.edu.in

Abstract

This research aims to examine the factors that affect the choice of a student opting for a new course. This research uses exploratory factor analysis in determining the factors that play a significant role in the study. The data was collected using a questionnaire and the reliability of the questionnaire was checked through reliability statistics. The results indicated that there are five factors that affect the choice of a student while choosing a new course, namely, course structure and diversity, affordability and credibility, Brand Value, Competition and the trend and interests. The total variance of these factors is 69.694 percent, with only a 30.306 percent loss of information, we may reduce the complexity of the data set by using these components, which accounts for 69.697 percent of the variability in the original 15 variables.

Keywords— Factor Analysis, Reliability, KMO Test, Bartlett's Test, Factor Loading, Education Management

I. INTRODUCTION

With the increase in competition in the field of education, many colleges in India are coming up with new courses. These educational institutes are catering to the needs of the students and shaping the new courses to attract the students. We can see a surge in the number of students opting for higher education. With this surge comes a higher demand for admissions to courses. Many students are enrolling themselves for courses that are new or have been begun in the past few years.

The authors of [1] (Kinanti, E.S., Ritchi, H. and Handoyo, S.) explained the determinants of service performance in higher education using exploratory factor analysis. The author Roberts, P emphasizes how academic perceptions play a key role in shaping the academic curriculum [2]. Paul M. Biner, Raymond S. Dean, and Anthony E. Mellinger emphasized student performance in a distance learning program [3]. Rummel, R.J. said that factor analysis can be used to examine a topic area, structure a domain, map foreign concepts, classify or reduce data, identify linkages, test hypotheses,

formulate theories, control variables, or draw inferences.[4]. The authors of [5] (Ford, J.K., MacCallum, R.C. and Tait, M.) have highlighted the fact that although factor analysis is a strong tool, researchers still lack a systematic approach to apply it in empirical work.

Factor analysis is a tool that helps in converting a large dataset into a smaller one by identifying the factors that have a significant role in the study. The two types of factor analysis are Exploratory Factor Analysis, which is a tool intended to generate a new theory and Confirmatory Factor Analysis, a tool to test an existing theory [6] (Matsunaga, M.). It is used to assess multicollinearity and prove or disprove theories that are already existing [7] (Williams, B., Onsman, A. and Brown, T.). Factor analysis can be applied to various fields, including psychology, education, health sciences, data science. It is also used in assessing groundwater quality [8] (Liu, C.W., Lin, K.H. and Kuo, Y.M.).

The youth nowadays have been trying to break the stereotypes of securing admissions in courses which offer basic graduation / post-

graduation in science, arts, commerce, law and engineering. The students have shifted their choice to choosing courses that are new. It can be due to multiple reasons, like high competition for existing courses, new courses provide a new field to explore, infrastructure is new, a good way to establish relations with the heads of the institute and many more. The main aim of this paper is to understand the primary factors that affect a student's decision, while choosing a new course, that has been started less than 5 years back, over an already existing course. The steps that can be taken by the colleges to incorporate these factors into the courses.

II. METHODOLOGY

We use factor analysis and specifically exploratory factor analysis with data reduction to identify structures from studies that have not been widely tested. The purpose of this study is to highlight the significant factors that affect the choice of a student for a new course.

2.1 Population

All the students who are currently enrolled in a course, be it Graduation or Post Graduation were eligible to participate in the survey designed. Those respondents whose course was new, i.e., had started less than 5 years back, were filtered out as that was the interest of our

study. Snowball sampling was used to conduct the survey. The survey was conducted in august 2021.

2.2 Questionnaire

The author Zhang, C., Yang, L., Liu, S., Ma, S., Wang, Y., Cai, Z., Du, H., Li, R., Kang, L., Su, M. and Zhang, J. used a questionnaire to collect the demographic information and to measure the severity of insomnia, respondents were asked to mark on a 4-point Likert scale [9]. A questionnaire was designed to study the factors that affect the choice for a new course over an existing course. It was a validated 15-item self-administered questionnaire. The questionnaire ascertained the age of the respondents, the current course they were enrolled in, if their course has been started less than 5 years back. The identity of the respondent was kept anonymous.

The responses were collected on a 4-point Likert scale, ranging from 'Not at all important' to 'highly important'. The respondents were asked to assess the important factors while choosing to secure admission in a new course. The total responses received were 71. Amongst the responses, 25 responses were excluded from our analysis as they did not fulfil the criteria of study. Hence, the final response for our study is 46.

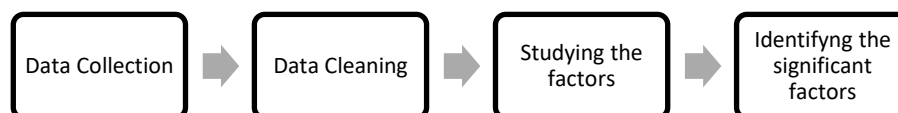


Fig 1: Methodology Flow Chart

III. RESULTS AND DISCUSSION

3.1 Reliability

Since the questionnaire is self-made, we need to check the reliability of the questionnaire. We evaluate Cronbach's alpha reliability statistics. If the reliability statistics > 0.7 , it means we can proceed with the questionnaire, as they are reliable and usable.

Figure 2 shows the reliability statistics is 0.748 i.e. 74.8% , hence our questionnaire is reliable and usable. We can proceed with the analysis.

Reliability Statistics	
Cronbach's Alpha	N of Items
.748	15

Fig 2: Reliability Statistics

3.2 KMO and Bartlett's Test

To begin with the exploratory factor analysis, adequacy data must be tested before. Kaiser-Meyer-Olkin (KMO) and Bartlett's test are terminologies in EFA that are used to measure the adequacy of sampling and measure the

homogeneity of indicators as a condition of construct correlation. To conduct exploratory factor analysis, the KMO value must be greater than 0.5 and Bartlett's test must be less than 0.05.

The values in Figure 3 show that the KMO is $0.568 > 0.5$. The results of the (KMO) and Bartlett tests are summarised in the above table. The KMO value is 0.568, and the Bartlett test is significant with $p \text{ value} = 0.000 < 0.5$, according to the results. It suggests that the data sufficiency conditions for factor analysis have been met. KMO values that fall within the range suggest that the data is sufficient to conduct an Exploratory Factor Analysis. Hence Bartlett test shows a significant value. Therefore, the data is sufficient to carry out further analysis.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.568
Bartlett's Test of Sphericity	Approx. Chi-Square	239.277
	df	105
	Sig.	.000

Fig 3: KMO and Bartlett's Test

3.3 Total Variance Explained

After conducting the adequacy data test, the data was extracted to discover the component pattern matrix. The Principal Component Analysis (PCA) extraction method and the Kaiser Varimax Normalization rotation method were used. The EFA results show that there are five initial factors identified as service performance and 12 elements of factor.

The degree of variation associated with the factor is represented by the eigenvalue. As a result, only factors with an eigenvalue greater than 1.0 are taken into account. Only five factors with eigenvalues greater than one are represented in the table, with factor 1 accounting for 25.194 percent of the variation in the data, factor 2 for 15.455 percent, and so on. See Figure 5.

The figure 4 shows the total variance explained explains approximately 70% of the variability in the original 15 variables, so we can use these components to minimise the complexity of the

data set while only losing 30% of the information.

The rotation keeps the cumulative proportion of variance explained by the extracted components, but it now spreads the variation more evenly across the components. Because of the huge differences in individual totals, the rotated component matrix should be easier to understand than the unrotated matrix.

Total Variance Explained						
Component	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.779	25.194	25.194	2.653	17.685	17.685
2	2.318	15.455	40.649	2.001	13.340	31.025
3	1.809	12.060	52.708	1.988	13.254	44.279
4	1.516	10.104	62.812	1.966	13.104	57.383
5	1.033	6.884	69.697	1.847	12.314	69.697
Extraction Method: Principal Component Analysis.						

Fig 4: Total Variance Explained

3.4 Scree Plot

The scree plot is a graph that shows the Eigenvalues in relation to all of the elements. The graph can be used to determine how many factors to keep. The point of interest is where the curve starts to flatten. We check the component number against the eigenvalue 1.

In figure 5 we observe that the eigenvalue was 1 against the component number 5. Hence, we considered 5 factors in our study.

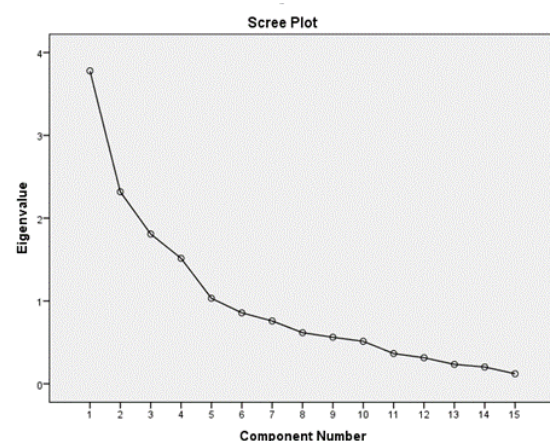


Fig 5: Scree Plot

3.5 Factor Loading

The factors in the component matrix are extracted. Once we decide how many factors to consider, we have to analyse whether a factor may be related to one or more variables in our data. The most common orthogonal rotation is varimax. The varimax rotated solution provides the simplest interpretation of the structure [10] (Pohlmann, J.T.). Rotation increases high item loadings while reducing low item loadings, resulting in a more comprehensible and streamlined solution. We have used varimax rotation in our analysis and get the rotated component matrix. For each of the five factors, the values which are greater than 0.5 are extracted. Those factors have a significant role in our analysis.

The figure 6.1 and 6.2 show the final variables under each factor and factor loading and factor name.

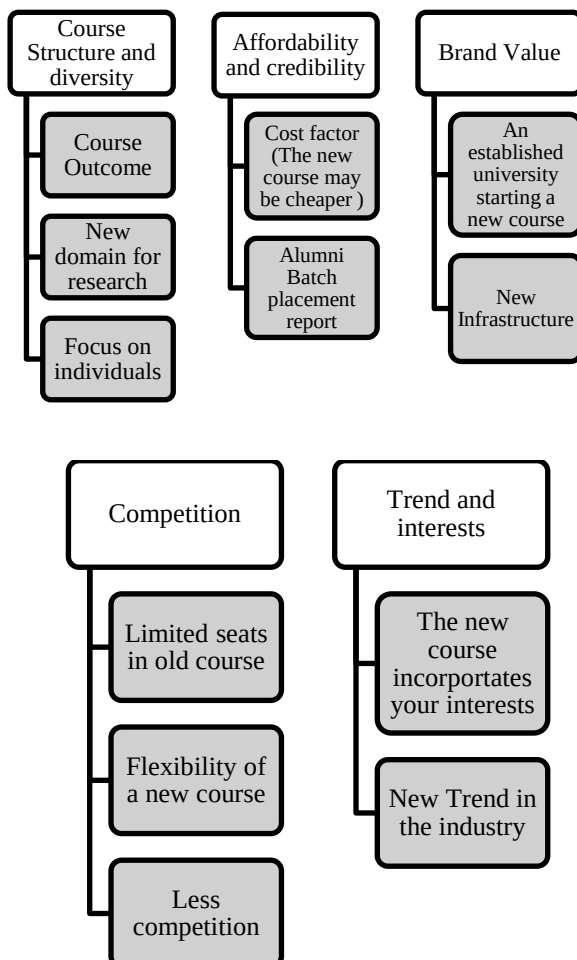


Fig 6: Key factors for choosing a new course

Factor 1 - Course structure and diversity: Factor 1 consists of 3 elements based on the course outcome, research opportunities and focus on individuals. This factor has a variance of 25.194%. It means that 25.194% of the factors in choosing a new course is determined by the course structure and diversity.

Factor 2 - Affordability and credibility: Factor 2 consists of 2 elements based on the cost factor and placement report. This factor has a variance of 15.455%. It means that 15.455% of the factors in choosing a new course is determined by the affordability and credibility.

Factor 3 - Brand Value: Factor 3 consists of 2 elements based on the new infrastructure and an established university starting a new course. This factor has a variance of 12.060%. It means that 12.060% of the factors in choosing a new course is determined by the brand value of a college.

Factor 4 - Competition: Factor 4 consists of 3 elements based on the limited seats in an old course, flexibility and less competition. This factor has a variance of 10.104%. It means that 10.104% of the factors in choosing a new course is determined by the competition.

Factor 5 - Trend and Interests: Factor 4 consists of 2 elements based on incorporating the individual interests and some new trend that has emerged in the market. This factor has a variance of 6.884%. It means that 6.884% of the factors in choosing a new course is determined by the interests of the individuals and trends in the market.

IV. CONCLUSION

The results showed that course structure and diversity played a crucial role in determining what are the factors that are important while choosing a new course. The course outcome gives a student a clear idea of the topics that the course is going to cover, which plays a major role in choosing any course, be it new or old. The new course would also open new horizons of research for students and gives a lot to explore in that field. And being a new course, it would focus on each individual. Apart from course structure and diversity other factors also play a key role. The primary factors that affect

the choice of a student while enrolling to a new course are course structure and diversity, affordability and credibility, brand value, competition and the trend and interests.

These results can be used in the education industry when making decisions about starting a new course as they are the main determining factors that a student looks for while enrolling to a newly started course.

The questionnaire was reliable and usable with reliability statistics as 0.784. The total variance of these factors is 69.697%, with only a 30.306% loss of information, we may reduce the complexity of the data set by using these components, which accounts for 69.694 % of the variability in the original 15 variables.

Out of the five factors extracted, factor 1 had 25.194% variance, factor 2 had 15.455% variance, factor 3 had 12.060% variance, factor 4 had 10.104% variance and factor 5 had 6.884% variance.

TABLE I. FACTOR LOADING

Factor Number	Factor Name	% Variance
Factor 1	Course Structure and Diversity	25.194%
Factor 2	Affordability and Credibility	15.455%
Factor 3	Brand Value	12.060%
Factor 4	Competition	10.104%
Factor 5	Trend and Interests	6.884%

REFERENCES

1. Kinanti, E.S., Ritchi, H. and Handoyo, S., 2020. Factor Analysis of Service Performance in Higher Education Institutions. *Journal of Accounting Auditing and Business-Vol*, 3(1).
2. Roberts, P., 2015. Higher education curriculum orientations and the implications for institutional curriculum change. *Teaching in Higher Education*, 20(5), pp.542-555.
3. Biner, P.M., Dean, R.S. and Mellinger, A.E., 1994. Factors underlying distance learner satisfaction with televised college-level courses. *American Journal of Distance Education*, 8(1), pp.60-71.
4. Rummel, R.J., 1967. Understanding factor analysis. *Journal of conflict resolution*, 11(4), pp.444-480.
5. Ford, J.K., MacCallum, R.C. and Tait, M., 1986. The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel psychology*, 39(2), pp.291-314.
6. Matsunaga, M., 2010. How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *International journal of psychological research*, 3(1), pp.97-110.
7. Williams, B., Onsman, A. and Brown, T., 2010. Exploratory factor analysis: A five-step guide for novices. *Australasian journal of paramedicine*, 8(3).
8. Liu, C.W., Lin, K.H. and Kuo, Y.M., 2003. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Science of the total environment*, 313(1-3), pp.77-89.
9. Zhang, C., Yang, L., Liu, S., Ma, S., Wang, Y., Cai, Z., Du, H., Li, R., Kang, L., Su, M. and Zhang, J., 2020. Survey of insomnia and related social psychological factors among medical staff involved in the 2019 novel coronavirus disease outbreak. *Frontiers in psychiatry*, 11, p.306.
10. Pohlmann, J.T., 2004. Use and interpretation of factor analysis in The Journal of Educational Research: 1992-2002. *the Journal of Educational research*, 98(1), pp.14-23.