MULTIVARIATE COMPARITIVE EXPLORATION OF COLLBORATIVE PRINCIPAL COMPONENT ANALYSIS AND SINGULAR VALUE DECOMPOSITION : A STUDY OF EMPLOYEE DATA IN AN INDIAN STATE

[1]Ann Baby, [2]A.Kannammal

Department of Computer Science,Rajagiri College of Social Sciences,Ernakulam, Kerala, India

Professor,Coimbatore Instiute of Technology,Coimbatore, India

Abstract:

In today’s technologically challenging era, establishments show honest concern towards the happiness and well-being of their employees. Job Satisfaction and Quality of Work Life of employees play a major role in the effectiveness and development of organisations. The betterment of quality of work life of human resources has become their topmost priority. In this paper, the latent constructs of Job Satisfaction and Quality of Work Life are identified with the help of a novel algorithm. This paper also performs a comparative study between the various dimensionality reductions strategies used for identification, in exploratory analysis. Based on data obtained from IT employees in Kerala, this paper proposes a Collaborative PCA algorithm to perform the factor analysis of the obtained data.A comparison between the proposed algorithm and the existing algorithm is proven in this research paper. A comparison with SVD is performed, which obtains similar results. A comparison with commercially available market ready statistical models and a complexity analysis of the same is alsoperformed.

Index Terms: PCA, SVD, Collaborative PCA, Exploratory Analysis, Job Satisfaction, Quality of Work Life

I INTRODUCTION

Employees are as the most basic strategic resources of every organization as success of organizations can only be tracked with the support of capable and accomplished manpower. Thus, IT organizations are spending considerable effort in the retention of committed employees. Every organizations need to focus on the work aspect of their employees and stimulate positive attitudes and behaviour through self-worth, self-esteem and positive identity at the workplace. The feel of content from their work lives is directly proportional to their usage of their full potential in achieving organizational goals.

II BACKGROUND

Dimensionality of a problem refers to the number of attributes orvariables thatis present in the data, which needs to be visualized[1]. The Curse of Dimensionality arise when analysing and organizing data in high-dimensional spaces. Data sets can range from simple single point data to multivariate data, from multidimensional databases. As the number of variables increase, the chance of noisy data in the data set increases. This can also lead to missing data. The dataset couldbelong to non-identical attribute sets, which are a combination of both nominal and continuous variables. As the dimensionality of the data increases, the magnitudes of the universe increases making the data sparse. To overcome this problem, Dimensionality Reduction techniques like Principal Component Analysis (PCA) and Singular Value Decomposition (SVD) are used.

Laws were legislated in the first half of the twentieth century to guard employees from job grievance and to eradicate dangerous working situations, monitored by the unionization crusade in the 1930's and 1940's[2]. The 1950's and the 1960's saw the growth of different models by psychologists suggesting an optimistic association among self-esteem and efficiency that enhanced human relations. In this paper, a comparative multivariate two-dimensional analysis using a newly proposed Collaborative PCA and SVD is performed on Job Satisfaction and Quality of Work Life of IT employees.

III ANALYSIS AND MODELLING

The variables for Job Satisfaction (JS) and Quality of Work Life (QWL) for the development of the algorithm has been adapted from various studies [3][4][5][6][7][8][9][10].Simon Easton's and Darren Van Laar's instrument("Work Related Quality of Life Scale Modified") is adopted[11]. Seventeen statements are used to measure the level of QWL and thirty two statements are used to measure the JS levels. The pilot study is done with twenty seven selected respondents for reliabilityanalysis of the created questionnaire. Based on the feedback from the pre-test, certain modifications, additions and deletions are carried out to ensure standardization in the questionnaire. As JS and QWL are variables of high dimensions, the Collaborative PCA and SVD are apt for dimensionality reduction.

IV THE PROPOSED MODEL

Primary data for the dataset are collected from two hundred and six employees by administering the questionnaires to IT employees in Kerala. To evaluate the level of quality of work life and job satisfaction among the employees, and to analyze the relationship between the quality of work life and job satisfaction, opinions of respondents are assessed in a five point Likert scale varying from "Strongly agree" to "Strongly disagree" and "Highly satisfied" to "Highly dissatisfied". The schematic of the

proposed model is detailed in Figure 1.

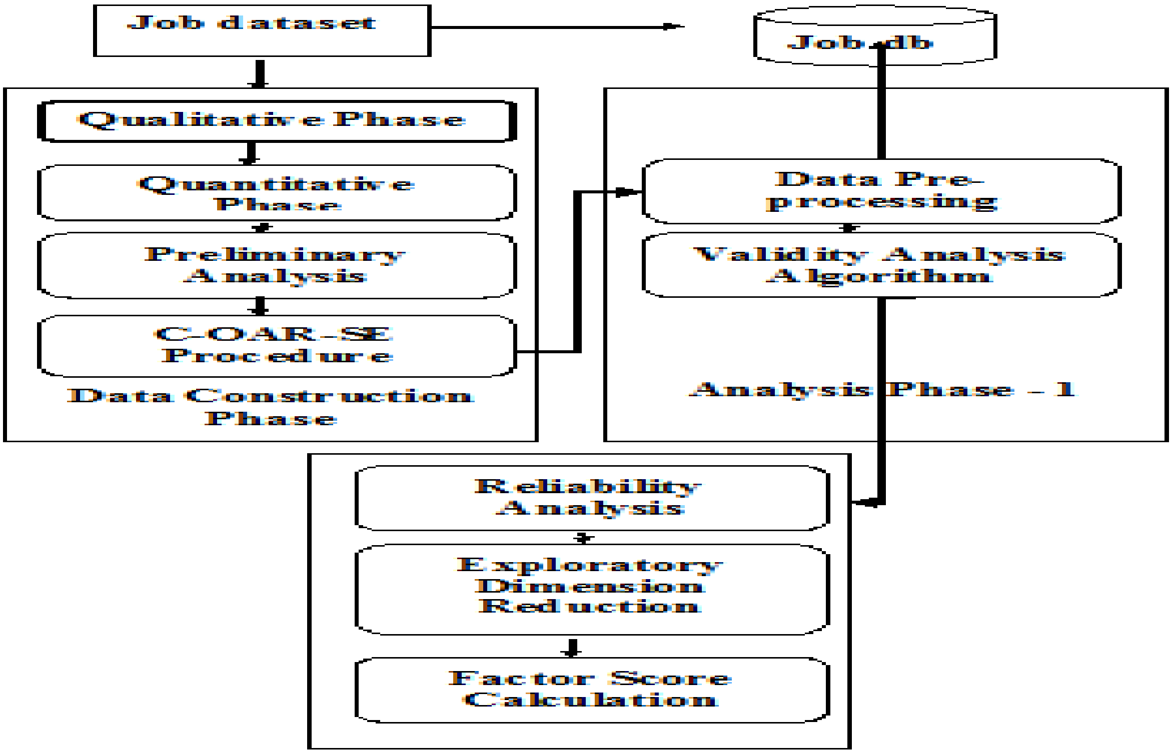

Figure 1. The Proposed Model

To ensure the inner consistency of the instrument, 'Cronbach's Alpha' reliability test was applied. Reliability values of greater than 0.75 are obtained for both the scales.Initially, data pre-processing and multi-level analysis (Analysis Phase-1) is performed, followed by a reliability analysis, dimensionality reduction and factor analysis (Analysis Phase-2).

In the data processing phase of Analysis Phase - 1, the first step is data cleansing, followed by data screening for missing values, and identification of outliers. The data set obtained from the tool generated is used to inspect the assumptions using dimensionality reduction techniques, after sufficient data pre-processing. Various consistency analysis procedures of content validity, face validity and construct validity are performed.

V MODEL ADEQUACY

To evaluate the appropriateness of the data for factor analysis, the following procedures are carried out, the results of which are summarized in Table 1.In the Analysis Phase –II, the Average Variance Extracted (AVE), the Fornell–Larcker criterion, 'Cronbach's Alpha (α)', Bartlett's Test of Sphericity and Kaiser-Meyer-Olkin (KMO) reliability tests [12] are incorporated into the reliability analysis algorithm.

Test Results		QWL	JS
KMO Measure		0.610	0.766
Bartlett's Test Measure	Approx. Chi-Square	1667.2	5373.2
	Degrees of freedom	136	496
	Significance	0	0

Table 1. Factor Analysis Adequacy of Job Dataset

The KMO value of sampling adequacy for the distinct variables is calculated. A score greater than 0.566 is advisable to get a sample good enough for sampling[13].KMO measure of sampling adequacy for the individual variables is analyzed to be 0.610 and 0.766 for QWL and JS respectively.**Error! Reference source not found.**2 describes the reliability coefficient associated with QWL and JS scales, number of items in the scale, its mean and standard deviation. Higher values of Cronbach's alpha indicate higher reliability. Reliability values are greater than 0.75 for both the scales. Therefore, the questionnaire meets the reliability requirements of minimum value of 0.75 as recommended.

Variab	Cronbach's	No: of
JS	.92	32
QWL	.85	17

Table 2. Reliability Statistics of JS and QWL

Exploratory data analysis is carried out to break the large database into different subgroups or clusters, via dimensionality reduction techniques. Dimensionality reduction of the data in the set is attempted by using Exploratory Centred Collaborative Principal Component Analysis (C-PCA) and Singular Value Decomposition (SVD). A comparison between the results generated by a covariance and correlation triggered PCA are studied. Factor matrix and composite measures are adjusted to achieve a meaningful factor solution by rotating the factors [14].

COLLABORATIVE PCA AND SVD

The proposed model uses a novel Collaborative PCA(C-PCA) algorithm. Collaborative PCA

follows the normal PCA algorithm. The algorithm for Collaborative PCA can be summarized as follows:Initiallya Normalized Dataset X(N,n) is obtained, following which the covariance

matrix of X, B is calculated. The Eigen vectors and Eigen values of the covariance matrix are calculated from where the components are chosen to form the feature vector, and obtain the principal components.The eigenvalues are ordered from highest to lowest to get the components in order of significance. The Eigen vectors with the highest Eigen values will be the principal components of the data set. The least significant Eigen Values are ignored and the eigenvector with the highest eigenvalue is considered as the principal component of the data. A feature vector matrix, U, is formed.

The new dataset is derived by computing F(N,p)=B(N,n) x U(n,p), where the reduced data set dimensionality will be p. A pattern matrix of the new matrix that will be a regression equation where the standardized observed variable is expressed as a function of the factors, is created. The loadings of this matrix will be the regression coefficients. TheCollaborative PCA is iterated until constructs load on only a single factor. A structure matrix is created that will contain the correlations between the variables and the factors.The correlation matrices are computed and examined. It shows that there are sufficient correlations to perform factor analysis. The overall significance correlation matrices are tested with Bartlett's Sphericity tests, which proves to be highly significant. This indicates a valid correlation between the items and proves goodness of fit of the data[13]. The existence of valid inter correlations between various items and goodness of fit to the data is thus proved.

The Rotated Component Matrix of QWL and JS load perfectly as there are no values < 0.5 across all cross loadings.For a scree representation, though there are 32variables, which represent JS, only those with Eigen values greater than one are considered significant. The scree plot obtained by Collaborative PCA reduction is depicted in **Error! Reference source not found.**2.The x-axis shows the proportion of variance for each principal component, while the upper line shows the cumulative variance explained by the first N components. The factors above the inflection point, where the curve starts to levels off at the elbow-point, are taken into consideration and the factors below the inflection point are discarded.For plots that have more than one inflection point, the factors corresponding to the Eigenvalues with maximum variance is retained.

In the given model, relationships between the Eigen values and variances are obtained using MATLAB, to obtain the percentage of variance. The Collaborative PCA algorithm helps in identifying the factors having latent roots or Eigen values greater than one (1), which are thus considered significant. Using these Eigen values for creating a cutoff is most reliable when the number of variables is between 20 and 50. Scree tests are used to identify the number of factors that can be extracted before the amount of unique or specific variance begins to dominate the common variance structure.

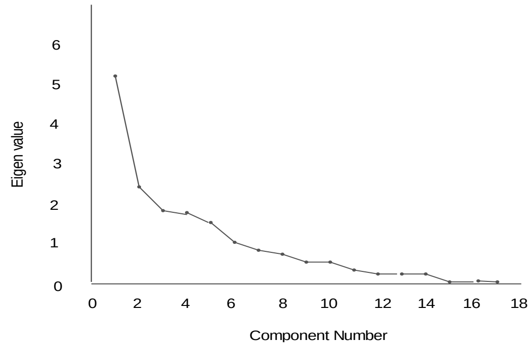

Figure 2. Scree plot of QWL using C-PCA

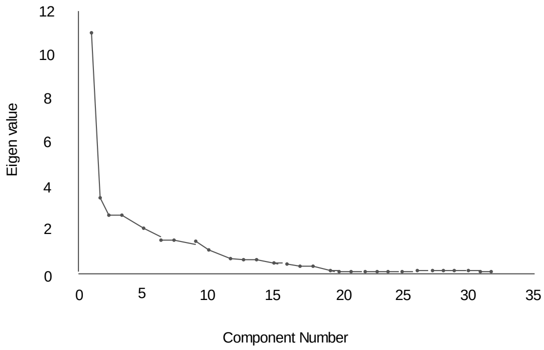

Figure 3. Scree plot of JS using C-PCA

From the above scree plot, nine factors are found to be highly influencing to achieve better job satisfaction for employees. The principal components are rescaled iteratively to identify factors loading with minimal error. Factors with latent root less than 1 are concluded to be insignificant and are ignored. The factors are thus grouped, identified and labelled as following.Factor 1 consisting of employees' opinion about Job-Quality, in-service training, Work exposure, Lack of motivation, Direction by managers and Research policies areidentified as "Employee Satisfaction and Engagement".The second factor comprises the employee opinion communication, political problems and feedbacks are identified as "Interpersonal Relationship and Recognition".The third factor consisting of employees' opinion regarding professional growth, job autonomy, rewarding system, career growth and bonus are named as "Career Growth and Support". The fourth factor including employee opinion about clear statement of projects, relation between staff and admin and continuity of programs is considered as "Internal Environment".The fifth factor comprising of employee opinion about training and tools, efficient support for family and transportation is considered as "Time Management and Resources".The sixth factor consisting of employee opinion about medical assurance and insurance is considered as "Stress and Strain".The seventh factor containing employee opinion about job specialization and motivation is considered as "Workplace environment".The eighth factor of employee opinion about technical help and opportunities for publishingis considered as "Facilities and Development". The ninth factor containing employee opinion about managerial style is considered as "Support of Top Management"[9][10][15].

Similarly, QWL consists of 17 variables (work environment,career growth,relationship behaviour,salary, security, bonus, incentives,organisation culture,training,facilities,workbalance,gender inclusive,time management, training,recognisation,health, and role conflict) which are reduced to six dimensions by the Collaborative PCA. Factor loadings with varimax rotation for QWL is carried out to investigate the relationships of a large number of items and to determine the feasibility of reducing them to a smaller set of factors.

The factors of QWL are identified and grouped into the following six factors. The first factor consisting of employees' opinion about work environment, career growth and relationship behaviour is identified as "Work life Balance". The second factor considers the employee opinion about salary, security and bonus and is identified as "Organization Support". The third factor consists of employees opinion regarding job designation, facilities, men /women equality, and roles managed is considered as "Gender Inclusiveness". The fourth factor comprises the employee opinion about recognitions, job autonomy and interpersonal skills, considered as "Recognition". The fifth factor covers the employee opinion about organization support and opportunities for career enhancement and is considered as "Career Advancement".The sixth factor contains the employee opinion about time management and health is considered as "Time Management"[9][10][15].

It is noted that in the SVD algorithm ten dimensions of JS are identified in contrast to Collaborative PCA, where only nine factors are obtained. The scree plot of SVD is shown in Figure 4. Meanwhile, the variables of QWL which has 17 variables, are reduced to six dimensionsas same as that of Collaborative PCA, which is shown in Figure 5.

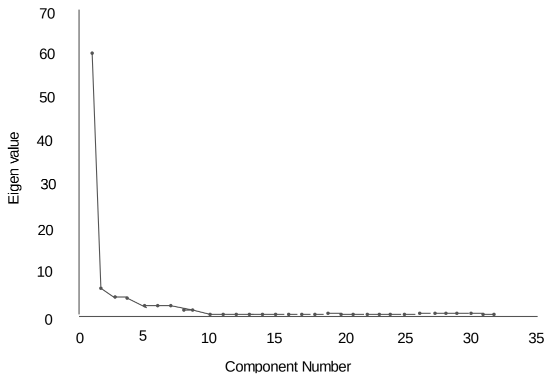

Figure 1. Scree plot of JS using SVD

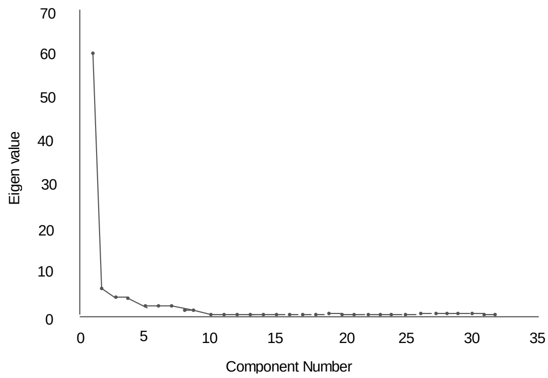

Figure 5. Scree plot of QWL using SVD

VI FINDINGS AND CONCLUSION

The time elapsed for performing Collaborative PCA and SVD, as per the algorithm in the paper is calculated, the results of which are summarized in Table 3. It is to be noted in this context that Collaborative PCA takes more time compared SVD.

Table 3 Time elapsed for C-PCA and SVD

Algorithm	**Time Elapsed**
C-PCA	1.9 seconds
SVD	0.8 seconds

As SVD and Collaborative PCA are mathematically related, the two algorithms essentially deliver similar result.

The results of the factor analysis of JS and QWL are compared IBM-SPSS, and depicted in Figure 6 and Figure 7 respectively. The earlier obtained nine factors are plotted with the results obtained from IBM-SPSS tool, which portrays a negligible error of10^{-5}.

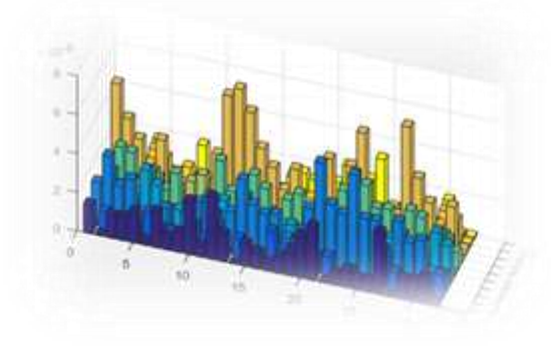

Figure 6. Result Comparison of JS

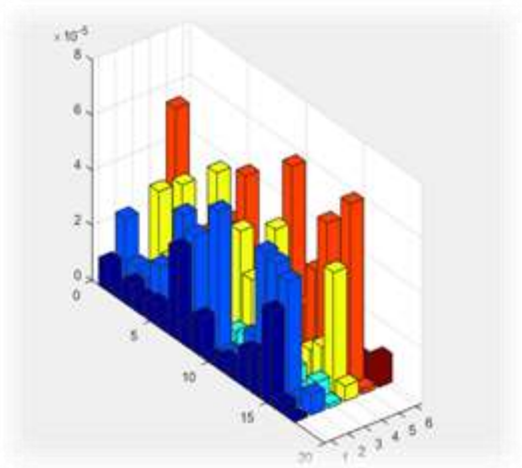

Figure 7.Result Comparison of QWL

VII CONCLUSION

A Comparative Multivariate Two-Dimensional Analysis Using Collaborative PCA and SVD on Job Satisfaction and Quality of Work Life of IT employees is performed in this paper.The job dataset justifies the Collaborative PCA algorithm, with the already existing job dataset models. The researched C-PCA algorithm reveals significant factorsof QWL and JS, the relationship between the same, and justifies existing literature with respect to results. The presence of QWL and JS in organizations, benefits both the employer and employee. It leads to improvement in the overall performance of the organization. It is concluded that Job satisfaction level among IT company employees is positively correlated with the quality of work life factors.

The Collaborative PCA for Dimensionality Reduction helps to perform the exploratory data analysis of the dataset. In contrast to SPSS and AMOS, MATLAB programming created and computed valid and robust models using statistical measures. It is also to be noted that dimensionality reduction models both the exogenous and endogenous variables. Relevant validity and reliability analysis procedures were performed to validate the analysis. A comparative research between Collaborative PCA and SVD was thus performed, with satisfactory results.

FUTURE WORK

The present study focusses on the dimensionality reduction using a collaborative PCA algorithm. The development of methods to develop an algorithm which can aid the presence or absence of a mediator/moderate variable would help in building up robust models. The forecasting of models can be optimized on a time-series forecast modelling, where models can be developed real-time. The emergence of large data sets, which would add up to the number of variables and constructs could be performed by a real- time analytical network analysis method. Application of a combined model generation and network path analysis method can help decision makers.

REFERENCES

[1] J. S. Yi, a. K. Youn and J. Stasko, “Toward a deeper understanding of the role of interaction in information visualization.,” IEEE transactions on visualization and computer graphics, vol. 13, no. 6, pp. 1224-31, 2007.

[2] F. Talib, Z. Rahman and M. N. Qureshi, “An empirical investigation of relationship between total quality management practices and quality performance in Indian service companies,” International journal of quality & reliability management, vol. 30, no. 3, pp. 280-318, 2013.

[3] P. A. Vlachos, N. G. Panagopoulos and A. A. Rapp, "Feeling good by doing good: Employee CSR-induced attributions, job satisfaction, and the role of charismatic leadership.," Journal of business ethics, vol. 118, no. 3, pp. 577-588, 2013.

[4] A. M. Grant, "Does intrinsic motivation fuel the prosocial fire? Motivational synergy in predicting persistence, performance, and productivity," Journal of applied psychology, vol. 93, no. 1, p. 48, 2008.

[5] F. T. Amstad, L. L. Meier, U. Fasel, A. Elfering and N. K. Semmer, "A meta-analysis of work–family conflict and various outcomes with a special emphasis on cross-domain versus matching-domain relations," Journal of occupational health psychology, vol. 16, no. 2, p. 151, 2011.

[6] V. Moniarou-Papaconstantinou and K. Triantafyllou, "Job satisfaction and work values: Investigating sources of job satisfaction with respect to information professionals.," Library & Information Science Research, vol. 37, no. 2, pp. 164-170, 2015.

[7] E. M. Skaalvik and S. Skaalvik, "Teacher self-efficacy and perceived autonomy: Relations with teacher engagement, job satisfaction, and emotional exhaustion," Psychological reports, vol. 114, no. 1, pp. 68-77, 2014.

[8] L. Lee, P. Wong, F. M. Der and A. Leung, "Entrepreneurial intentions: The influence of organizational and individual factors," Journal of business venturing, vol. 26, no. 1, pp. 124-136, 2011.

[9] F. Yuan and R. W. Woodman, " Innovative behavior in the workplace: The role of performance and image outcome expectations," Academy of Management Journal, vol. 53, no. 2, pp. 323-342, 2010.

[**10**] J. M. Millán, J. Hessels, R. Thurik and R. Aguado, "Determinants of job satisfaction: a European comparison of self-employed and paid employees," Small business economics, vol. 40, no. 3, pp. 651-670, 2013.

[**11**] E. C. Duyan, S. Aytaç, N. Akyıldız and D. Van Laar, "Measuring work related quality of life and affective well-being in Turkey," Mediterranean Journal of Social Sciences, vol. 4, no. 1, pp. 105-116, 2013.

[**12**] F. Murtagh and A. Heck, "Multivariate data analysis," Springer Science & Business Media, vol. Vol. 7, no. 131, 2012.

[13] J. W. Creswell and J. D. Creswell, Research Designs: Qualitative, Quantitative and Mixed Method Approaches., Sage Publications Inc., 2017.

[**14**] H. Abdi and L. J. Williams, "Principal Component Analysis," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 2, no. 4, pp. 433-459, 2010.

[**15**] N. Tanaya, K. S. Chandan and K. M. Pravat, Quality of Work Life plays the Mediating role in between Workplace Empowerment and Employee Commitment: A Study on Healthcare Employees of an Indian State, Ahmedabad: 1st IIMA International Conference on Advances in Healthcare Management Services, At Indian Institute of Management, 2015.

[**16**] L. S. Meyers, G. Glenn and J. G. Anthony, Applied multivariate research: Design and interpretation, Sage publications, 2016.

[**17**] P. Gill, W. Murray and M. H. Wright, The Levenberg-Marquardt Method Practical Optimization, London: Academic Press, 1981, pp. 136-137.

[**18**] J. R. Rossiter, "How to use C-OAR-SE to design optimal standard measures," European Journal of Marketing, vol. 50, no. 11, pp. 1924-1941, 2016.

[**19**] V. Venkatesh and Bala, "Technology acceptance model 3 and a research agenda on interventions," Decision Sciences, vol. 39, no. 1, p. 273–315, 2008.

[**20**] Y. Fang, I. Qureshi, H. Sun, P. McCole, E. Ramsey and K. H. Lim, "Trust, Satisfaction, and Online Repurchase Intention: The Moderating Role of Perceived Effectiveness of E-Commerce Institutional Mechanisms.," MIS Quarterly, vol. 38, no. 2, pp. 469-479, 2014.

[21] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," MIS quarterly, vol. 3, no. 1, pp. 319-340, 1989.

[22] O. N. Nelson and S. Queenie, “Consumer attitudes, system's characteristics and internet banking adoption in Malaysia,” Management Research News, vol. 29, no. 2, pp. 16 - 27, 2006.

[23] O. Turel, A. Serenko and P. Giles, “Integrating technology addiction and use: An empirical investigation of online auction users,” MIS Quarterly, vol. 35, no. 4, pp. 1043-1062, 2011.

[**24**] A. Arbore, I. Soscia and R. P. Bagozzi, “The role of signaling identity in the adoption of personal technologies,” Journal of the Association for Information Systems, vol. 2, no. 2, p. 86, 2014.

[**25**] V. Venkatesh, F. D. Davis and M. G. Morris, “Dead or alive? The development, trajectory and future of technology adoption research,” Journal of the Association for Information Systems, vol. 8, no. 4, p. 267–286, 2007.

[**26**] V. Venkatesh, Technology acceptance model and the unified theory of acceptance and use of technology, Wiley Encyclopedia of Management, 2014, pp. 273-315.

[**27**] L. S. Meyers, G. Glenn and J. G. Anthony, Applied multivariate research: Design and interpretation, Sage publications, 2016..

[**28**] J. F. Hair, G. T. Hult, C. Ringle and M. Sarstedt, A primer on partial least squares structural equation modeling (PLS-SEM), Sage Publications, 2016.