APPLICATION OF FACTOR ANALYSIS TO PUBLIC SECTOR INTEGRITY IN INDONESIA

A) Warsono[1], B) Armen Yasir[2], C) Dian Kurniasari[1] , D) Widiarti[1], E) Ridwan Saifuddin[3]
[1]Department of Mathematics, University of Lampung, Indonesia
[2]Faculty of Law, University of Lampung, Indonesia
[3]BAPPEDA Kota Metro, Lampung, Indonesia

Abstract

The main purpose of this study is to analyze interrelationships among variables used on the survey of public sector integrity by Indonesia's Corruption Eradication Commission (Komisi Pemberantasan Korupsi, KPK). The nine variables include corruption experiences, corruption perceptions, working environments, administration systems, the behavior of individuals, corruption prevention efforts, integrity experiences, integrity potencies, and integrity total. Using factor analysis, the approach is to explain these variables in terms of their common underlying dimensions, well-known as factors. Technically, factor analysis involves condensing the information contained in a number of original variables into a smaller set of new composite factors with a minimum loss of information. The results show that based on eigen values the first factor alone accounts for 70.7% of the common variance. The second factor alone accounts for 13,4%. The common variance of the nine variables explained by two factors is 84.1%. Using the varimax rotation and based on values of factor loadings the first factor makes high contribution to the variance of corruption experiences, corruption perceptions, working environments, the behavior of individuals, integrity experiences, and integrity total variables. The second factor makes high contribution to the variance of corruption prevention efforts and integrity potencies variables. Similar results, also, are obtained by quartimax rotation and equamax rotation.

Keywords: Corruption Eradication Commission (KPK), Factor Analysis, Eigenvalues, Factor Loadings, Varimax Rotation, Quartimax Rotation, Equamax Rotation

1. INTRODUCTION

Originally introduced by Spearman (1904)[11] in the area of psychology, factor analysis is one of a number of statistical methods which comprise the branch of statistical theory known as multivariate analysis. Started as a controversial and difficult subject, factor analysis has emerged as one of the most fascinating and usefull data analysis tools and its applicability to many diverse areas such as social sciences, education, and biology. The general purpose of factor analytic techniques is to find a way to condense the information contained in a number of original variables into smaller set of new, composite dimensions or variates (factors) with a minimum loss of information. In meeting its purpose, factor analysis provides several key pieces of information about multivariate data: (1) identifi cation of inferred latent variables referred to as factors, (2) estimates of the amount of variance explained by each factor, and (3) the relationship of the original data to each factor [1, 5, 6, 7, 8, 9, and 10].

Meanwhile in order to support the efforts more effective aort nd efficient to combat and eradicate an extraordinary crime of corruption, Indonesia's Corruption Eradication Commission (Komisi Pemberantasan Korupsi, KPK) regularly conducts integrity surveys on public services in some institutions and local governments across the country [3]. These surveys involve a large number of variables that consist of observable and unobervable or latent variables. As discuss above that because of the prospect of factor analysis usefulness, it makes motivation of this study to examine the application of factor analysis to the area of law, especially to corruption survey data of public sector.

Hopefully, in terms of science application, this study might contribute to analyze suvey data of public sector integrity in Indonesia.

2. DATA OF PUBLIC SECTOR INTEGRITY AND POCEDURE OF FACTOR ANALYSIS

In order to demonstrate the application of factor analysis, this study uses subsets data of public sector integrity in 60 local government (Pemerintah Kota) in Indonesia published by KPK in 2011. The considered data consist of 9 variables that are x_1: Corruption Experiences; x_2:Corruption Perception; x_3: Working Environments; x_4: Administration systems; x_5: Behavior of Individuals; x_4: Corruption Prevention Efforts; x_7: Integrity Experiences; x_8: Integrity Potencies; and x_9: Integrity Total [3].

Suppose we make observations on p=9 variables $\mathbf{x} = (x_1, x_2, ..., x_9)'$ with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_9)'$ and variance-covariance matrix $\boldsymbol{\Sigma}$, the factor analysis model expresses each variable as a linear combination of underlying common factors $f = (f_1, f_2, ..., f_k)'$ with an accompanying residual $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_9)'$ and can be explained by:

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{Lf} + \boldsymbol{\varepsilon}$$

that implies

$$x_1 = \mu_1 + \lambda_{l1} f_1 + \ldots + \lambda_{1k} f_k + \varepsilon_1$$
$$x_2 = \mu_2 + \lambda_{21} f_1 + \ldots + \lambda_{2k} f_k + \varepsilon_2$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$x_9 = \mu_2 + \lambda_{91} f_1 + \ldots + \lambda_{9k} f_k + \varepsilon_9$$

The elements $f_1, f_2, ..., f_k$ are called the *common factors*; the number of factors k should be substantially smaller than p. The coefficient λ_{ij} is the weights called the *factor loadings*, so that λ_{ij} is the loading of the i^{th} variable on the j^{th} factor. The coefficient λ_{ij} is indicates the importance of the j^{th} factor f_i to the i^{th} variable x_i and can be used in interpretation of f_i. The variable $\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$ describes the residual variation specific to the i^{th} variable. The residual variables are called the *specific factors*. It is assumed that $E(\varepsilon_i) = 0$, $var(\varepsilon_i) = \psi_i$, $cov(\varepsilon_i, \varepsilon_k) = 0$, $i \neq k$, and $cov(f_i,\varepsilon_i) = 0$ [4, 6, 13, 15].

From the above factor model and under the assumptions, we have

$$E(\mathbf{f}) = \mathbf{0},\ cov(\mathbf{f}) = \mathbf{I},$$
$$E(£) = \mathbf{0},\ cov(£) = \vee$$
$$cov(\mathbf{f}, £) = \mathbf{0}$$
$$E(\mathbf{x}) = \boldsymbol{\mu},\ cov(\mathbf{x}) = \mathbf{LL}à + \vee$$
$$cov(\mathbf{x},\mathbf{f}) = \mathbf{L}$$
$$\sigma_{ij} = cov(x_i, x_j) = \lambda'_i \lambda_j = \lambda_{i1}\lambda_{j1} + \lambda_{i2}\lambda_{j2} + ... + \lambda_{ik}\lambda_{jk}$$

and

$$\sigma_{ii} = var(x_i) = \lambda'_i \lambda_i + \psi_i$$
$$= (\lambda^2_{i1} + \lambda^2_{i2} + ... + \lambda^2_{ik}) + \psi_i$$
$$= h_i^2 + \psi_i$$
$$= \text{communality} + \text{specific variance}$$

The quantity ψ_i, the contribution of the specific factor $£_i$, is called the *uniqueness* or *specific variance*, and the quantity h_i^2, the contribution of common factors, is called *communality of common variance*. Furthermore, λ^2_{i1} is the contribution of the 1^{st} common factor to the common variance, λ^2_{i2} is the contribution of the 2^{nd} common factor to the common variance, and so on [6, 8, 9, and 10].

The parameters of the factor analysis model, including the factor loadings and the error variances, are usually unknown and need to be estimated from the sample data. The sample covariance matrix is occasionally used, but it is much more common to work with the sample correlation matrix. In estimating the parameters, this study consider to use correlation matrix and principal factor method.

The factor loadings can be used to interpret the label of the factors in terms of the common elements that load highly on each factor. However, if the factor loadings obtained are difficult to interprate, it is customary to rotate these factor loadings. The interpretation will usually be clearer after rotation of the factor pattern that offers the most adequate interpretation of the variables under examination. For example, suppose the factor loadings corresponding the first two original variables are wether positively or negatively high for the first factor, The first common factor then can be interpreted as a linear combination of only these two variables. Factor rotations are broadly classified as either orthogonal, in which the

rotated factors are orthogonal to each other , or oblique, in which the rotated factors are not orthogonal to each other [5, 6, 8, 9, and 10].

In many areas of applications, orthogonal rotations are used commonly. Orthogonal rotation is the process of extracting so that the factor axes are maintained at 90 degrees. There are three popular orthogonal that varimax rotation, quartimax rotation and equamax rotation [1, 4,and 5]. Among them the variamx method proposed by Kaiser in 1958 [7] is the most popular of these methods and is often used to rotate principal components solutions. For comparison purposes, this study consider varimax rotation, quartimax rotation,and equamax rotation.

3. DATA ANALYSIS AND RESULTS

To demonstrate how to implement factor analysis this study use use data set published by Indonesia's Corruption Eradication Commission known as Komisi Pemberantasan Korupsi (KPK) [2]. Tabel 1 contains the unrotated component analysis factor matrix. The first row of numbers at the bottom of each column is the column variance (eigenvalues) of each factor and indicates the relative important of each factor in accounting for the variance associated with the set of variables. To determine the numbers of factors needed to explain correlations among variables, the most popular approaches are the eigenvalue-greater-than-one rule, the proportion of variance explained by the factors, and the scree plot that a plot of the eigenvalues associated with each of the factors extracted, against each factor. The first factor,

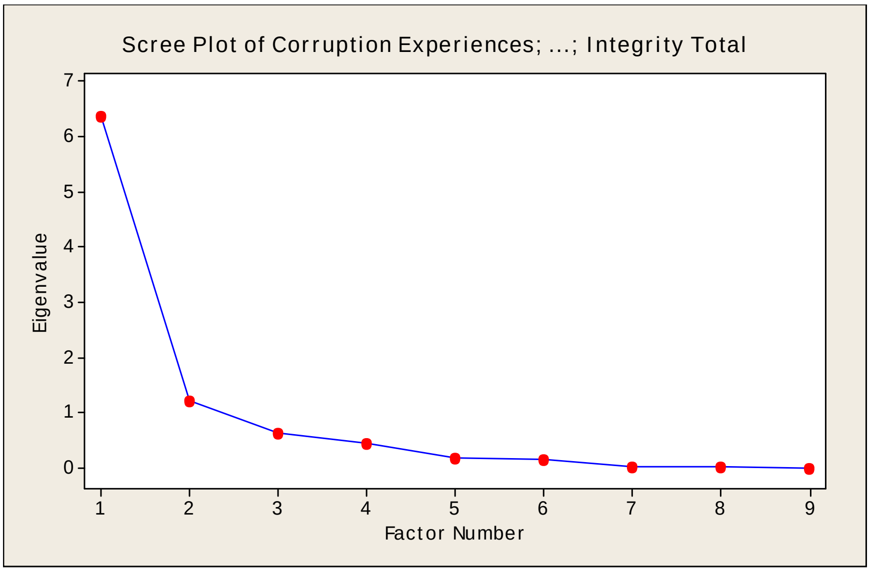
Table 1. Estimated unrotated factor loadings, eigenvalues, and communalities

Principal Component Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Factor4	Communality
Corruption Experiences	0,929	0,242	0,096	0,169	0,960
Corruption Perceptions	0,937	0,246	0,073	0,148	0,965
Working Environments	0,856	0,299	-0,011	-0,296	0,910
Administration Systems	0,665	-0,137	-0,730	0,034	0,995
Behavior of Individuals	0,847	-0,032	0,138	-0,461	0,950
Corruption Prevention Efforts	0,435	-0,851	0,200	0,103	0,964
Integrity Experience	0,941	0,167	0,100	0,223	0,973
Integrity Potencies	0,846	-0,478	-0,039	-0,100	0,956
Integrity Total	0,969	0,025	0,073	0,160	0,971
Eigenvalue	6,3636	1,2098	0,6234	0,4472	8,6440
% Var	0,707	0,134	0,069	0,050	0,960

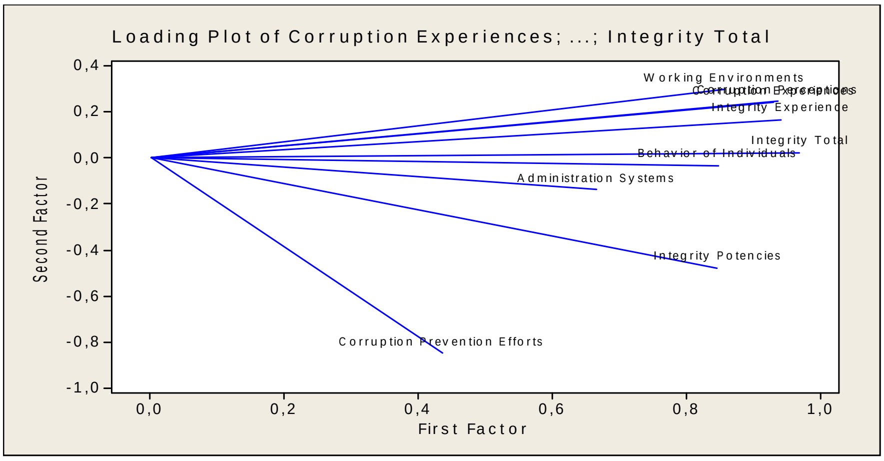
with eigenvalue of 6.3636, accounts for approximately 70.7% of the variance. The second factor, with eigenvalue of 1.2098, accounts for 13.4% of the variance explained. The remaining factors have eigenvalues less than 1. The cumulative percent of variance explained by the first two factors is 84.1%. Based upon the first two rules, therefore, we might consider the first and the second factor retained. As shown in Graph 1, moreover, the scree plot confirms our conclusion. The elbow of the scree plot is approximately at two factors.



Graph 1. Scree plot of KPK data set with 9 variables

Table 1 also presents unrotated factor loadings all of variables that extracted by the principal component method. Factor loadings represent the degree of association or correlation of each variable with each factor. Based on unrotated factor loadings, the first factor can be roughly interpreted as “General Integrity Conditions”, since it is positively high correlated with variable Integrity Total, Integrity Experience, Corruption Perceptions, Corruption Experiences, Working Environments, Behavior of Individuals, Integrity Potencies,and Administration Systems. The first factor can be labeled as a “Integrity Index” factor. Because it is negatively high correlated with variable Corruption Prevention Efforts, the second factor can be called “Corruption Prevention” factor.

Vector plot graph can be constructed from the factor loadings of Table 1, as shown below (Graph 1). This is a graphical expression of the information in the factor pattern. This graph presents clearly that the first factor is defined primarily by variable Integrity Total, Integrity Experience, Corruption Perceptions, Corruption Experiences, Working Environments, Behavior of Individuals, Integrity Potencies,and Administration Systems. The second factor is represented mainly by variable Corruption Prevention Efforts.



Graph 1. Vector of unrotated factor loading

Since the factor solution is not unique and to achieve a simpler factor structure that can obtain another factor solution by rotating the axes. This study considers to use orthogonal rotations that are varimax, quartimax, and equamax methods. In applied social sciences subject, orthogonal rotation is used most often, probably because it is the default in major statistical programs and the perception that orthogonally rotated solutions are more easily interpreted because the factor loadings represent correlations between the indicators and the latent factors.

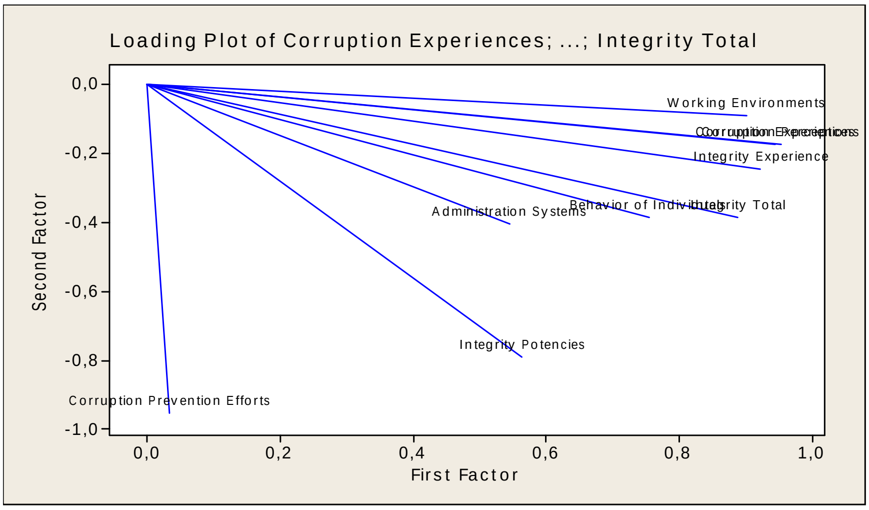
In the varimax rotation, the first factor recieves high factor from the variables Corruption Perceptions, Corruption Experiences, Integrity Experience, Working Environments, Integrity Total, Behavior of Individuals, Integrity Potencies,and Administration Systems, respectively (Table 2). Table 2, also, shows that the second factor recieves high factor from the variables Corruption Prevention Effots and Integrity Potencies.

Table 2. Varimax rotated factor loadings, eigenvalues, and communalities

Rotated Factor Loadings and Communalities
Varimax Rotation

Variable	Factor1	Factor2	Communality
Corruption Experiences	0,944	-0,174	0,922
Corruption Perceptions	0,953	-0,173	0,938
Working Environments	0,902	-0,092	0,822
Administration Systems	0,545	-0,405	0,461
Behavior of Individuals	0,754	-0,387	0,719
Corruption Prevention Efforts	0,034	-0,955	0,914
Integrity Experience	0,923	-0,246	0,913
Integrity Potencies	0,564	-0,791	0,945
Integrity Total	0,889	-0,387	0,940
Eigenvalue	5,4411	2,1323	7,5734
% Var	0.605	0.237	0.841

Graph 2 presents vector plot graph can be constructed from the factor loadings of Table 2. This graph presents clearly that the first factor is defined primarily by variables Corruption Perceptions, Corruption Experiences, Integrity Experience, Working Environments, Integrity Total, Behavior of Individuals, Integrity Potencies,and Administration Systems. The second factor is represented mainly by variables Corruption Prevention Effots and Integrity Potencies.



Graph 2. Vector of varimax rotated factor loading

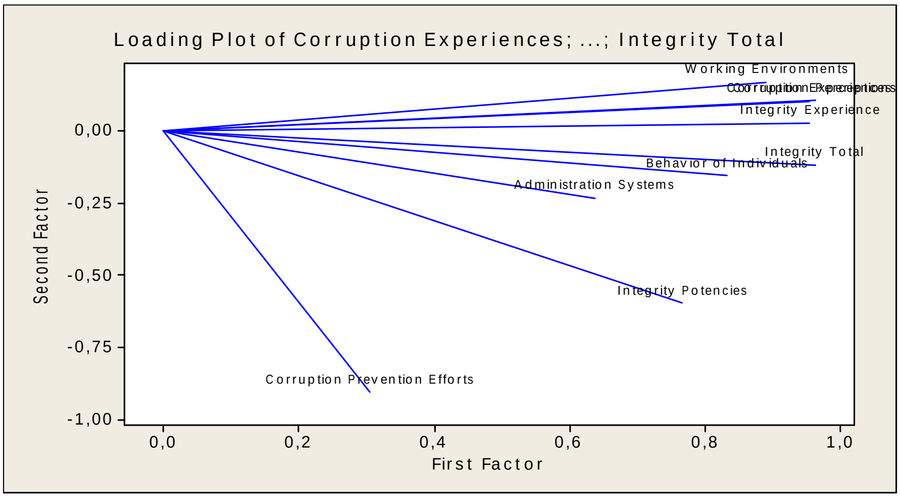
In the quartimax rotation, the first factor recieves high factor from the variables Corruption Perceptions, Integrity Total, Corruption Experiences, Integrity Experience, Working Environments, Behavior of Individuals, Integrity Potencies,and Administration Systems, respectively (Table 3). Based on Table 3, it can be interpretated that the second factor recieves high factor from the variables Corruption Prevention Effots and Integrity Potencies.

Table 3. Quartimax rotated factor loadings, eigenvalues, and communalities

Rotated Factor Loadings and Communalities
Quartimax Rotation

Variable	Factor1	Factor2	Communality
Corruption Experiences	0,955	0,103	0,922
Corruption Perceptions	0,963	0,106	0,938
Working Environments	0,891	0,170	0,822
Administration Systems	0,638	-0,233	0,461
Behavior of Individuals	0,833	-0,156	0,719
Corruption Prevention Efforts	0,305	-0,906	0,914
Integrity Experience	0,955	0,027	0,913
Integrity Potencies	0,767	-0,597	0,945
Integrity Total	0,962	-0,118	0,940
Variance	6,2524	1,3210	7,5734
% Var	0,695	0,147	0,841

Graph 3 presents vector plot graph can be constructed from the factor loadings of Table 3. This graph presents clearly that the first factor is defined primarily by variables Corruption Perceptions, Integrity Total, Corruption Experiences, Integrity Experience, Working Environments, Behavior of Individuals, Integrity Potencies,and Administration Systems. The second factor is represented mainly by variables Corruption Prevention Effots and Integrity Potencies.



Graph 3. Vector of quartimax rotated factor loading

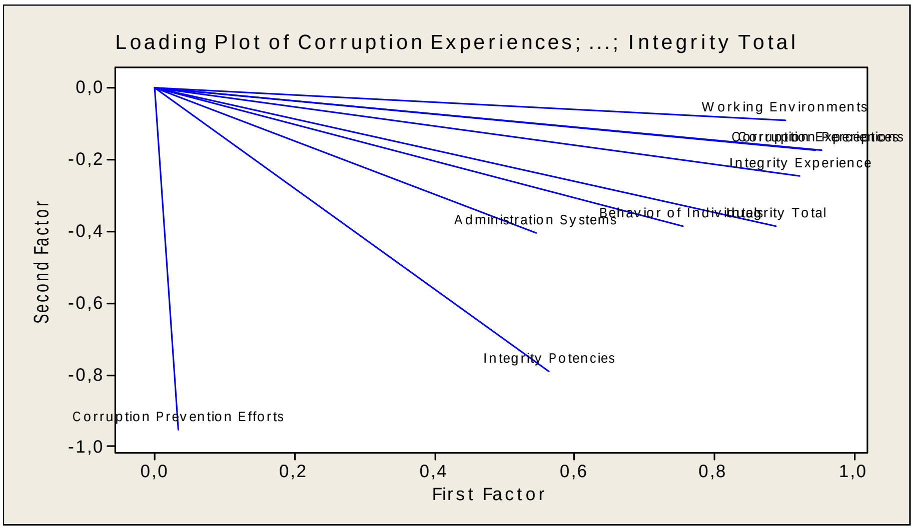
Results of the equamax are similar than those of the the varimax rotation. the first factor recieves high factor from the variables Corruption Perceptions, Corruption Experiences, Integrity Experience, Working Environments, Integrity Total, Behavior of Individuals, Integrity Potencies,and Administration Systems, respectively (Table 4). Table 4 presents that the second factor recieves high factor from the variables Corruption Prevention Effots and Integrity Potencies.

Table 4. Equamax rotated factor loadings, eigenvalues, and communalities

Rotated Factor Loadings and Communalities
Equamax Rotation

Variable	Factor1	Factor2	Communality
Corruption Experiences	0,944	-0,174	0,922
Corruption Perceptions	0,953	-0,173	0,938
Working Environments	0,902	-0,092	0,822
Administration Systems	0,545	-0,405	0,461
Behavior of Individuals	0,754	-0,387	0,719
Corruption Prevention Efforts	0,034	-0,955	0,914
Integrity Experience	0,923	-0,246	0,913
Integrity Potencies	0,564	-0,791	0,945
Integrity Total	0,889	-0,387	0,940
Variance	5,4411	2,1323	7,5734
% Var	0,605	0,237	0,841

Graph 4 presents vector plot graph can be constructed from the factor loadings of Table 4. This graph presents clearly that the first factor is defined primarily by variables Corruption Perceptions, Corruption Experiences, Integrity Experience, Working Environments, Integrity Total, Behavior of Individuals, Integrity Potencies,and Administration Systems. The second factor is represented mainly by variables Corruption Prevention Effots and Integrity Potencies.



Graph 4. Vector of equamax rotated factor loading

4. SUMMARY

Based on survey data of public sector in Indonesia published by KPK in 2011, the results of the factor analysis show that based on eigen values the first factor alone accounts for 70.7% of the common variance. The second factor alone accounts for 13,4%. The common variance of the nine variables explained by two factors is 84.1%. Using the varimax rotation and based on values of factor loadings the first factor makes high contribution to the variance of corruption experiences, corruption perceptions, working environments, the behavior of individuals, integrity experiences, and integrity total variables. The second factor makes high contribution to the variance of corruption prevention efforts and integrity potencies variables. Similar results, also, are obtained by quartimax rotation and equamax rotation.

REFERNCES

[1] Brown, Bruce L., Suzanne B. Hendrix, Dawson W. Hedges, and Timothy B. Smith. 2012. Multivariate Analysis for the Biobehavioral and Social Sciences A Graphical Approach. USA: John Wiley & Sons, Inc., Publication.
[2] Browne, Michael W. 2001. An Overview of Analytic Rotation in Exploratory Factor Analysis. *Multivariate Behavioral Research*, *36* (1), 111-150
[3] Direktorat Penelitian dan Pengembangan Kedeputian Bidang Pencegahan Komisi Pemberantasan Korupsi Integritas. 201. Sektor Publik Indonesia Tahun 2011. Fakta Korupsi dalam Layanan Publik. 2011. Diterbitkan oleh Direktorat Penelitian dan Pengembangan Kedeputian Bidang Pencegahan Komisi Pemberantasan Korupsi.
[4] Finch, W. Holmes. 2011. A Comparison of Factor Rotation Methods for Dichotomous Data. *Journal of Modern Applied Statistical Methods*, Vol. 10, No. 2, 549-570
[5] Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., and Tatham, R.L. 2006. Multivariate Data Analysis. New Jersey : Pearson Education, Inc.
[6] Johnson, R.A., and Wichern, D.W. 1982. Applied Multivariate Statistical Analysis. New Jersey : Prentice-Hall, Inc.
[7] Kaiser, H.F. 1958. The Varimax Criterion for Analytic Rotation in Factor Analysis. *Pyschometrika*, 23, 187-200.
[8] Khattree, R. And Naik, D.N. 2000. Multivariate Data Reduction and Discrimination With SAS Software. USA : John Wiley and Sons, Inc.
[9] Mardia, K.V., Kent, J.T., and Bibby, J.M. 1979. Multivariate Analysis. San Diego : Academic Press, Inc.
[10] Sharma, S. 1996. Applied Multivariate Techniques. USA : John Wiley and Sons, Inc.
[11] Spearman, C. 1904. General Intelligence, Objectively Determined and Measured. *American Journal of Psychology*, 15, 201-293.