**Scholars Journal of Physics, Mathematics and Statistics** **ISSN 2393-8056 (Print)**
Sch. J. Phys. Math. Stat. 2015; 2(1):44-56 **ISSN 2393-8064 (Online)**

(An International Publisher for Academic and Scientific Resources)

# Students' In-take Grade Index (SIGI): A Comparative Tool with CGPA in Determining the Level of Academic Impact during Tertiary Education

**John Komla Coker Ayimah**
Lecturer (Ho Polytechnic) & PhD Candidate (UG-Legon), Department of Statistics, Ho - Volta Region, Ghana

***Corresponding Author:**
John Komla Coker Ayimah
Email: johncokeroriginal@gmail.com

**Abstract:** Ranking tertiary institutions provides a good motivation, since it engenders healthy competition among stakeholders to strive for excellence. The issue of concern now might be finding a fair measure of the various criteria for this purpose. When institutions are endowed differently, then one fair criterion of assessing them would be to find an algorithm to measure their academic impart on the students by the level of output during the tertiary education. At the moment, there is the GPA which measures the output of the students during this period. What remains is to find a similar measure – a default estimate of students' output, using their in-take grades, so that the difference between the CGPA and the in-take grade index (SIGI) will provide some measure of the level of impact. A departure, positively or negatively, from the SIGI will determine how well the student was affected, giving the prevailing conditions specific to the institution. This is the objective of the study. The study proposed an algorithm using the procedures of performing principal component factor analysis to obtained weights specific to the six subject areas used to admit students into tertiary institutions, by subjecting the entire data of a group of students to factor analysis procedures. These weights are subsequently used in the final formula to compute the SIGI, after testing and ascertaining the fitness of the factor model. To test the algorithm, records of some 618 past students of Ho Polytechnic, taken randomly from the past three academic years, were used. It was revealed that the level of output during tertiary education has dropped or increased marginally than those who had the same increased tremendously. The model could be of great help to regulatory authorities in assessing tertiary institutions, so that appropriate intervention could be made for better academic performance. It could also be used as admission criteria when the grades of group of students are analysed to determine who to admit. Also, after studying the trend for some time within an institution, certain class of students could be given the chance if there is evidence from the model that they can do better when admitted. The model is hereby recommended to stakeholders in the educational front. The focus of subsequent studies would be to apply similar robust methods, used here, to diagnose the current GPA system as a measure of the student's academic performance in terms of its robustness, validity, fairness and accuracy. This is in line with proposing an alternate grading system which satisfies the criteria of robustness, validity, fairness and accuracy.
**Keywords:** Education; Factors; Grading; Academic Impart; Loadings; Performance.

## INTRODUCTION

Tertiary education is crucial in the life time of an individual. According to Chickering [1], quoted by Umar et al[2], the period of tertiary education represents a critical developmental period for both late adolescents and young adults. This is because; it is the period where the adolescent is exposed to many factors (both extraneous and pertinent) that influence life. Sadly, the social demand of such factors makes it at times very difficult for the vulnerable young–man/woman to overcome. Some of these factors are romantic relationships, joining clubs and other social activities. No matter how good or bad these may be to the youth, the overall determinant of how good or bad the student does at the tertiary level is his/her output, in terms of academic performance, measured as cumulative grade point average (CGPA).

Lots of researches have been done on some of these factors and their corresponding influence on academic performance. These social factors affect academic performance in terms of time demanded and the psychological state they may cause[2]. In the quest to provide an assessment of how well the student is doing or have done in the face of these factors, there is the need to set a benchmark for this measurement. The CGPA is one such internal measure for assessing the academic progress of the student. Another way might be to make a comparison with the student's in-take grades in determining whether the student's overall output has dropped or appreciated, percentage-wise. If we assume that all extraneous and pertinent factors prior to admission into a tertiary institution are similar to those during the tertiary education and common to specific group of students, given that the student's intelligence level is the same, then a good assessment of the student's level of output during the tertiary education can be determined by comparing his/her CGPA with a similar index/measure designed for the in-take grades. By this, one would be able to have an estimate of how well the student prevailed, given his in-take grades as the default level of output, as well as have a fair idea of how imparted the students was – in terms direct academic contacts and control over other social factors.

Having such an index would provide lots of benefits; it enables a good base for making inference about the general measure of the effects of all factors put-together during the tertiary education period. Also, it may form one of the bases for comparing and ranking tertiary institutions, Ghana since one goal of the institution is to find ways of mitigating the negative effect of these factors on students; so that, an institution which scores higher positive percentages, being the difference between comparing CGPA's to in-take grade indexes, is obviously doing well. There should some relationship between a student's performance during tertiary education and a previous academic performance at lower levels; in fact, Cohen-Schotanus et al.[3] quoted Ferguson et al. that variance in overall medical school performance was explained by previous academic performance. This is the motivation behind this study. The study seeks to introduce an algorithm for developing an index from a number of subject grades from the senior high schools, forming the in-take grades.

When grades, which are exams scores of students, are modelled, there may be correlations among subject grades. These correlations may be due to the presence of some factor(s) that may be common to all students under study. Therefore, studying the underlying reasons requires a statistical tool developed for such purpose. Factor Analysis is a multivariate tool designed for studying the reasons why correlations exist among course grades[4]. Multivariate data are common today and inability to apply them would no longer be an excuse, especially with the advent of inexpensive computer power to carry them out[5]. As a case study, the data obtained for simulation purposes, was taken from the Ho Polytechnic records of students for the past three academic years.

**Background of Research**

Ranking tertiary institutions provides good motivation, since it engenders healthy competition among stakeholders to strive for excellence. Recently, such rankings have sparked some concerns among the populace as to which criteria were used for the assessment. Also, there is the concern about whether institutions endowed differently have to be assessed using the same set of criteria for all. These concerns, undoubtedly, would arouse some reservations whenever such rankings come out; people may feel unfairly treated, since the ground for assessment may not seem even to them. In response to these concerns is the need to develop a robust index that can be used independently for students that enrol into an institution. The assessment would then be based on those institutions, give their conditions, are able to impart more positively the academic performance of their students. By this, it will come out more clearly those tertiary institutions that have made more interventions in students' life. The main questions of concern here are; do students went out better than they came in, in terms of their general output, given the prevailing compelling factors of influence? In percentage terms, has the student increased or decreased, in terms of his general output? There should be a mathematical model for used in determining these answers.

Just as an index would help in assessing institutions, so would the index be helpful to assess the output of gender, faculties, departments, year groups, etc. Cheesman et al[6] in their research about determinants of student performance at university: reflections from the Caribbean, concluded that gender and faculty, among other factors, are main determinants of student performance. In their research: the effect of social factors on students' academic performance in Nigerian tertiary institutions, Umar et al [2] found out that student cults are academic impediment; romantic relationships, have the highest impart and may be a psychological barrier to an effective learning process; excessive sporting activities and involvement in clubs and organizations may pose some threats too. Obviously, the tertiary student, who Wright [7] said is prone to stressors at the transformative stages of his/her life, has a daunting task of managing these factors and still maintain a good academic output.

As the student grapples with this task, more stress is generated; Romano [8] said all factors have a direct or indirect relationship with students' performance; these daily stressors do not cause anxiety by themselves, stress results from interactions between stressor and the individual's perception and reaction to them. In addition to these social factors, there are other environmental factors like availability of lecture halls, availability of other learning equipments, and even design of academic facilities, that can have some influence on the output of students. In fact, the inadequacies of such physical resources like lecture halls, halls of residence, laboratories, libraries and other academic resources translates to poor results because it breeds over crowdedness[9]. Again, Fabiyi and Uzoka[9]have observed that the planning and design of educational facilities for schools have imparted on educational outcomes. According to Earthman[10], school building design features and components have been proven to have a measurable influence on student's learning.

In the face of these compelling evidences, interest might be on how the student copes, by having an algorithm to measure the impart, assuming that an index$(I)$ for the student's in-take grades is the benchmark. With respect to comparing one institution to the other, it should be the case that an institution and its students which show greater number of positive impact are adjudged to have done better, given all the factors peculiar to the institution. The impart $(D)$ in this case should be the difference between the student's CGPA $(Output, O)$ and an index $(I)$ for the student's in-take grade, such that,

$$D = O - I \tag{1}$$

The method developed here for the index$(I)$ is with the assumption that all factors are common to some group of students admitted to a particular tertiary institution. From equation (1), a high level of impact$(D)$ would have resulted from a higher output $(O)$; a lower level of impact$(D)$ would also be as a result of a lower output$(O)$.

At the moment in Ghana, in most tertiary institutions, the student's performance is measured using the grade point average (GPA) system. The measure tells the semester-by-semester output of the student throughout the program, until the final GPA, cumulative grade point average is found. In order to have a good benchmark for comparing the level of output for the student during the tertiary education, there is the need to develop a similar measure for student's in-take grades, while assuming that the CGPA is an optimal tool for determining the student's performance. In doing so, an index for the in-take grades obtained by a student should be computed so that

1. There is a highest possible index (value) a student can get as there is also the highest possible GPA value.
2. A student's in-take grade index is a percentage of the highest possible index, since a student's GPA could also be converted to a percentage of the highest possible GPA.
3. In calculating this in-take grade, a weight is assigned to each subject area showing how well the students, collectively, performed in that subject. This weight is such that the higher the weight, the more "important" the subject or course is in assessing the students.

Following these measures, a single index should be obtained for the number of in-take grades considered for admitting the student. An index could be calculated for each student, and his/her percentage index found. The difference between the percentage GPA and the percentage in-take grade index would then form the bases for making the inference.

As was mentioned much earlier in the introduction, when variables describing a data are large (mostly greater than or equal to two) and correlates the most effective statistical tool that can best offer a good explanation of the data is the application of multivariate tools. Often when variables are correlates there may be optimal linear combinations of variables[11]. Rencher argued further that, multivariate inference is especially useful in curbing the researcher's natural tendency to read too much into the data: one such multivariate tool is factor analysis[11]. According to Rencher [11], if the original variables, $X_1, X_2, \dots, X_p$, are at least moderately correlated, the basic dimensionality of the system is less than $p$. Hence the goal of factor analysis is to reduce the redundancy among the variables by using a smaller number of factors, hence reducing the amount of noise in the data. In relating this to student's performance in $p$ number of subjects or courses, we can posit that there is a factor that is common to all students within a defined area which can explain why their scores are correlated. When the number of common factors is adequately determined, using appropriate validation tools, the loadings of each variable on the factors could be used as weights specific to the variables.

**Objective of the Study**

The main objective of the study is to model students' in-take grade (SIG) using factor loadings so that it could be used as comparative tool with CGPA in determining the level of output during tertiary education, hence the level of impart. The specific objectives to be covered are:

1. Run a principal component factor analysis and determine unique factor loadings attributed to some salient factors associated with each variable.
2. Use these unique factor loadings as weights to develop the student in-take grade index(SIGI),
3. Determine percentage SIGI and percentage CGPA for sampled students, using their records as test data.
4. Determine the distribution of the difference between the CGPA and the SIGI.

## RESEARCH METHODS

### *Sampling and Sample Size*

The data for simulation in this study were taken from the students' records department of Ho Polytechnic, Ho. Ho polytechnic is one of the ten polytechnics in Ghana and situated in the capital of the Volta region of Ghana. The data, which were taken from all departments and for three consecutive years, are for simulation purposes only. Systematic method of sampling (at an interval of 5 units) was employed to determine the individual sample units from the sample frames of the various year groups. The table below is the summary of the sampling results.

**Table 1: Sampling Statistics**

| Indicator | Proportionate Allocation | Percentage Allocation |
|---|---|---|
| *Department* | | |
| Accountancy(A) | 238 | 38.5 |
| Automobile Engineering(AE) | 28 | 4.5 |
| Building Technology(BT) | 75 | 12.1 |
| Electrical & Electronic Engineering(EEE) | 49 | 7.9 |
| Fashion Design & Modelling(FDM) | 19 | 3.1 |
| Hotel Catering & Institutional Management(HCIM) | 36 | 5.8 |
| Marketing(M) | 105 | 17.0 |
| Production Engineering(PE) | 11 | 1.8 |
| Statistics(S) | 22 | 3.6 |
| Secretaryship & Management Studies(SMS) | 35 | 5.7 |
| Total | 618 | 100 |
| *Sex* | | |
| Male | 451 | 73 |
| Female | 167 | 27 |
| Total | 618 | 100 |

*Source: Student records, 2010 to 2013 graduation list*

***Variables in the Research***

The variables in this study are the grades for the six subjects used in admitting a student into a tertiary institutions in Ghana, denoted as $X_j, j = 1,2, \dots ,6$, so that, $x_{ij}, i = 1,2,3, \dots ,618$, is a realization under a variable so that the mean vector is represented as, $(x_{i1}, x_{i2}, x_{i3}, \dots . x_{i12})', i = 1$; the student's final GPA, denoted as CGPA. For convenience,

*$X_1$=Grade for English*
*$X_2$=Grade for Mathematics (Core)*
*$X_3$=Grade for Science (Core)*
*$X_4$=Grade for First Elective*
*$X_5$=Grade for Second Elective*
*$X_6$=Grade for Third Elective*
*CGPA=Student Final GPA*

The in-take grades for the variables were quantified using the grade the student obtained as defined in the table below.

**Table 2: Quantifying the In-take Grades**

| S/N | Category | Value |
|---|---|---|
| 1 | A or A1 | 5 |
| 2 | B or B2 and B3 | 4 |
| 3 | C or C4, C5 and C6 | 3 |
| 4 | D or D7 | 2 |
| 5 | E or E8 | 1 |

*Source: As defined by the researcher*

The highest value of "5" for the highest category was determined so that it corresponds to the highest possible grade point for the GPA system. These categories were so defined because the cut-off grade for entry into the polytechnics during the periods considered for this research was E or E8 as in the Senior Secondary Schools' Certificate Examination (SSSCE) and the West African Senior Schools' Certificate Examination (WASSCE) respectively.

***Concepts behind the Study***

When data consist of two or more variables, it is frequently of interest to evaluate or study the inter-correlations among the variables. The pair-wise correlations among the indicator variables suggest the grouping of the variables into homogenous sets. The statistical tool which does this is called factor analysis. By factor analysis, we seek the most parsimonious factor model that best explains or accounts for the correlations among the indicators[11]. The objectives of factor analysis as Sharma [4] outlined are to use the computed correlation matrix to

1. Identify the smallest number of common factors that best explain or account for the correlations among the indicators.

2. Identify, via factor rotations, the most plausible factor solution.
3. Estimate the pattern loadings (factor loadings), communalities, and unique variances of the indicators.
4. Provide an interpretation for the common factor(s).
5. Estimate or compute the factor scores for use in a further analysis.

In order to achieve these objectives, two common methods may be used; namely, the principal component factoring (PCF) and the principal axis factoring (PAF). Again, Sharma [4], contented that in "most cases it really does not matter which of the two techniques is used", since the difference between the results may not be significant, even though they are conceptually different techniques. However, he further suggested that in cases where the researcher's interest is in seeking an "implicit" underlying factor model, the PAF should be used. The use of the principal component factoring technique posit that the initial variance a variable shares with others, which is the communality, is one (high), and hence the procedure goes ahead to subject the correlation matrix, with the estimated communalities in the diagonal, to a principal component analysis.

The initial assumption for using PCF suggests that, while the researcher seeks to understand the correlation among the variables, he/she should have it, as a hypothesis, that each variable shares a high amount of variance with the other. In dealing with a data such as the one being dealt with here, the scores of students in examination, it is proper to use the same assumption as for PCF. That is, while we seek to understand why scores or grades are correlated, we posit initially that each variable shares the same high variance with other variables in the data – this explains why this method is used here. This variance could be the level of difficult of the various examinations or the relative importance of a particular course to the candidate. So, principal component factoring is appropriate for this study. However, it is important to stress that principal component factoring is not principal component, since the former is an improvement of the latter because of the importance it holds over the same. This is because, incorporated in principal component factor analysis is the advantage of having to rotate the component matrix along new orthogonal axes, which is dealt with in the next section.

***Principal Component Factoring***

Now, assume in a principal component analysis technique where new variables which are linear combinations of the original variables are formed. If the $p$ indicator variables are $X_1, X_2, X_3, \dots, X_p$ , then the new components are,

$$y_i = \sum_{j=1}^{p} a_{ij} X_j \text{ , } where\ i = 1,2, \dots, p \qquad (2)$$

where the $a_{ij}$'s are the eigenvectors of the correlation matrix. Hence equation (2) suggests that there can be as many components as the number of original indicators available. The new variables $y_i$ are called principal components, which must satisfy the following conditions.

1. They must be uncorrelated; for the $y_i$'s to be uncorrelated, then
   $a_{i1}^2 + a_{i2}^2 + a_{i3}^2 + \cdots + a_{ip}^2 = 1, i = 1,2,3, \dots, p$ and
   $a_{1i}a_{j1} + a_{2i}a_{j2} + a_{3i}a_{j3} + \cdots + a_{pi}a_{jp} = 0, for\ all\ i \neq j$
2. The first new variable, $y_1$,must account for the maximum variance in the data, the second, $y_2$, must also account for the maximum variance that was not accounted for by the first.

By condition (2) above, the best criteria for retaining a desired number of components for further factoring would be to observe the eigen values of the respective components, since that measures the variance accounted for by the component. Additionally, Rencher[11] outlined a number of methods used to determine the number of factors to retain; among them is to choose $m$, the number of factors, to equal the number of eigenvalues greater than the average eigenvalue – for standardised data with correlation matrix, $R$, the average is one. A number of other methods are used in this regard. In fact, when a data set is successfully fitted by a factor analysis model, whichever the method used will almost give the same number of extracted value [11]. Once the factors to retain are known, we carry out the mathematics of factoring using the principal component method as follows.

From equation (2), the $k^{th}$ principal component, which is uncorrelated and constitutes orthogonal axes with another component, would be given as,

$$y_k = \sum_{j=1}^{p} a_{kj} X_j \qquad (3)$$

Rewriting this in vector notation gives,

$$y_k = (a_{k1} \quad a_{k2} \quad a_{k3} \quad a_{kp}) \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_p \end{pmatrix} = a_k' X$$

$\therefore y_k = a_k' X$ (4)

Now the variance of $y_k$, $Var(y_k) = Var(a_k' X)$.

$\therefore Var(y_k) = a_k' \Sigma a_k$ (5)

In this manner, $\Sigma$ is referred to as the variance – covariance matrix of $X$. We seek that $y_1$ account for the largest variation in the data and $y_2$ the second largest variation etc. From equation (5), we have

$$\therefore Var(y_1) = a_1' \Sigma a_1$$

Let $S_1^2 = Var(y_1)$, so that

$$S_1^2 = \Sigma a_1^2 + \boldsymbol{\ell}(1 - a_1^2)$$

And $\ell$ is a Lagrange multiplier. Taking the partial derivative of $S_1^2$ with respect to $a_1$ yield,

$$\frac{\partial S_1^2}{\partial a_1} = 2a_1\Sigma - 2a_1\ell = 2(\Sigma - \ell I)a_1$$

Where $I$ is the $p \times p$ identity matrix. For maximum $S_1^2$, $\frac{\partial S_1^2}{\partial a_1} = 0$

$\Rightarrow 2(\Sigma - \ell I) = 0, for\ a_1 \neq 0$. Hence $|\Sigma - \ell I| = 0$. Multiplying $2(\Sigma - \ell I)$ by $a_1'$ yield $a_1'\Sigma a_1 - a_1'\ell a_1 = 0$. Implying that, $a_1'\Sigma a_1 = a_1'\ell a_1$ and since $a_1' a_1 = 1$,

$\ell = a_1' \Sigma a_1$ (6)

Equation (6) is interpreted as the eigenvalue of the first principal component and, $a_1$, is the first eigenvector. In general, $\lambda_k = a_k' \Sigma a_k$ is the sample variance of $y_k$ and is the $k^{th}$ largest eigenvalue of $\Sigma$. One thing that is commonly done in multivariate analysis is to standardize the data so that the variance of each $p$ is one. So it makes sense to say that if there is a composite of all $p$ indicators, then the variance of that composite must not be less than one. Hence this becomes a good measure in retaining which components to carry further. That is why, even though there are a number of methods used in deciding on the adequate number of salient components to retain, the eigenvalue-greater-than-one rule was used here.

Now, in using the principal component method as estimation in factor analysis, we assume that the maximum possible component is equal to some maximum possible unobservable factors that can result when the correlation matrix is analyzed for possible regrouping. Suppose therefore, that the principal component analysis of a correlation matrix yields some component (as in equation 3), which now becomes some new factor, $f_j$, as

$$f_j = \sum_{i=1}^{p} a_{ij} X_i \qquad (7)$$

where the set of coefficients $a_{ij}, i = j = 1,2,3, \ldots, p$ form the $j^{th}$ eigenvector and, $\sum_{j=1}^{p} a_{ij}^2 = 1$. Suppose that the eigenvalues of $f_j$ is $\lambda_j$, then multiplying the right hand side of equation (7) by $\sqrt{\lambda_j}$, we have

$$f_j = \sqrt{\lambda_j} \sum_{i=1}^{p} a_{ij} X_i = \sum_{i=j=1}^{p} a_{ij} \sqrt{\lambda_j} X_i$$

$$\therefore f_j = \sum_{i=1}^{p} \beta_{ij} X_i \qquad (8)$$

Where, $\beta_{ij} = a_{ij}\sqrt{\lambda_j}$, is also called the vector of loadings of all variables $X_i$ on $f_j$. Hence we have some new construct, $f_j$, which are composites of the original variables, $X_i$. These factors also represent the underlying dimensions (construct) that summarize or account for the original set of observed variables. In matrix notation, we can rewrite equation (8) in the form,

$F = AX$ (9)

where $F$ is of dimension $p \times 1$, called the vector of factors; $A$ is of dimension $p \times p$, called the orthogonal matrix ($AA' = A'A = I$) of the indicator variables and $X$ is of dimension $p \times 1$, called the vector of indicators. From equation (9), we can have $X = A'F$ or

$$X_i = \sum_{j=1}^{p} \beta_{ij} f_j \qquad (10)$$

Equation (10) also suggests that, the existence of the original variables is actually because there exist some underlying construct that necessitate its existence. In accordance with the objectives of factor analysis, suppose that $m$ factors are used to explain the correlations among the original variables $X_i$, then equation (10) will now be written as,

$$X_i = \sum_{j=1}^{m} \beta_{ij} f_j + e_i \qquad (11)$$

That is, the remaining $p-m$ factors are represented by $e_i$ and this value is used to determine the variance specific to any original indicator, $X_i$. In matrix notation, equation (11) is written as $X = \Lambda f + \psi$. After extraction, the factor model needs to be labeled or interpreted; factor loading provides the key to understanding the nature of the particular factor. The factor loadings are correlations between the original variables, $X_i$, and the factors, $f_j$, they are therefore very important in factor analysis. Like in equation (4), $f_j = a_j^{'} X$ and $X_i = a_i^{'} X$

$$Cov(X_i, f_j) = Cov(a_i^{'} X, a_j^{'} X) = a_i^{'} \Sigma a_j$$

But $\lambda_j = a_j^{'} \Sigma a_j$,multiply through this by $a_j$, we have

$$a_j \lambda_j = a_j a_j^{'} \Sigma a_j$$
$$\Rightarrow a_j \lambda_j = \Sigma a_j$$

Hence

$$Cov(X_i, f_j) = a_i^{'} a_j \lambda_j = \lambda_j a_i^{'} a_j$$

$$\therefore Cov(X_i, f_j) = \lambda_j a_{ij} \qquad (12)$$

By definition, the correlation between the $X_i$ and $f_j$ is

$$\rho_{X_i, f_j} = \frac{Cov(X_i, f_j)}{\sqrt{Var(X_i)}\sqrt{Var(f_j)}}$$
$$\rho_{X_i, f_j} = \frac{\lambda_j a_{ij}}{\sqrt{\delta_i^2}\sqrt{\lambda_j}}$$

$$\therefore \rho_{X_i, f_j} = \frac{a_{ij}}{\delta_i}\sqrt{\lambda_j} \qquad (13)$$

Equation (13) is called the loading of $X_i$ on $f_j$. In interpreting the factors, we consider those with high loadings, $\rho_{ij}$, of all $X_i$'s on a particular factor $f_j$. Comparing equations (8), (11) and (13), it is seen that, $\beta_{ij} = \delta_i \rho_{ij}$, with $\rho_{ij}$ being "modified" by $\beta_{ij}$'s, such that $\beta_{ij}$ and $\delta_i \rho_{ij}$ have a maximum possible value of one. Hence the $\beta_{ij}$ 's will be maximised quickly when using the principal component method of extraction in factoring than just doing a principal component analysis, where the $\rho_{ij}$ 's are the loadings.

Now the sum of squared loadings indicates what percentage of the variance in the original variables is explained by a factor. The total variance explained by the $m$ extracted factors is the sum of their respective variances. Hence a high loading indicates the variable's importance in understanding the factor label. Once the loadings are obtained, they may be rotated. The concept of rotation is necessary because it presents the opportunity of taking a look at the data from different perspectives (being the new axes). So the perspective, being the type of rotation, that present the maximum opportunity of having a better view of the data help in the interpretability of the factor model. In fact, if we can achieve a rotation in which every point is close to an axis, then each variable would have loaded highly on one factor and lowly on the others – this removes ambiguity and such a state is called simple structure [11]. This state, Rencher, stressed, simplifies factor interpretability because we can observe which variables are associated with each factor hence reducing the complexity of a variable to one.

An objective of factor extraction is to obtain new components/factors that are uncorrelated so that the uniqueness of these could be measured in subsequent analysis, if necessary. In such cases orthogonal factor rotation is required. In accordance with objective-1 of this study, to identify unique factor loadings attributed to some salient factors that are associated with each variable (subject area), orthogonal factor rotation will be the appropriate method to use. Going by the objective of factor rotation, the objective of orthogonal rotation, as well as the objective of this study, varimax rotation method, among others, will be appropriate to use here. Thus the varimax method attempts to make the loadings either large or small, on one factor, to facilitate interpretation[11].

A factor model can be said to be a "good one" when the amount of variance the indicators share with other variables is appreciably high and the specific variances as low as possible, such that the following equation is satisfied.

$$\varphi = res_{ij} = R - \Lambda\Lambda' \qquad (14)$$

where $\varphi$ is the matrix of unique variances(residual matrix), $R$ is the correlation matrix based on the $p$- indicator variables and $\Lambda\Lambda'$is called the reproduced correlation matrix based on the $m$-factor model. For an $m$-factor adequacy test, then the residual matrix must be close to zero as much as possible for the factor model to be statistically significant.

***The Grade Point Average (GPA)***

The grade point average is a system of grading adopted by most tertiary institutions in the world. Of particular interest to this study is the GPA system used by the public polytechnics in Ghana. Incorporated in the computation of the GPA is a weight associated with each course or subject of study so that the final output is a weighted average. The system also has the highest grade point of five, so that, a student's GPA could be deemed as a percentage of this "five". Therefore, in the GPA system, the student's raw examination scores are replaced by some grade points of which five is the highest. The aggregate of the GPA is the cumulative grade point average (CGPA) for a student.

***The Student In-take Grade Index (SIGI)***

The focus of this study is to develop a formula, an index, which could be used in comparison with the GPA to determine the level of output for a student during his/her tertiary education. The development of the index is based on the aforementioned theories and concepts discussed under section 2.3 above. Assume, from equation (11), that the basic equation of the factor analysis is obtained. Then for each $f_j, 1 \le j \le m$, those variables, $X_i, 1 \le i \le p$, with high loadings, $\beta_{ij}$, called the indicators of $f_j$ would be found. Assume further that the factor model is a good one and all variables have a complexity of one. Then their rotated factor loadings, as defined in equation (13) can be used as weights of $f_j$'s such that the higher the $\beta_{ij}$'s of a variable on a particular factor, the higher the weight of the variable. So, if there are $p$ indicator variables (in this case the six in-take grades), then

$$\sum_{i=1}^{p} \beta_{ij} \le p \qquad (15)$$

Which seeks to say that, the sum of all loadings associated with the variables cannot exceed the number of variables in the model, since the highest possible loading (correlation) is one. From equation (15), it implies that, the greater $\sum_{i=1}^{p} \beta_{ij}$, the better, since the highest possible weight might be sought for each indicator – hence we seek to maximise $\sum_{i=1}^{p} \beta_{ij}$. If so, then for all $f_j$'s, there should be some $\beta_i^*, 1 \le i \le p$, to be called harmonised loadings, being the largest $\beta_{ij}$ for each $X_i$ so that

$$\sum_{i=1}^{p} \beta_{ij} < \sum_{i=1}^{p} \beta_i^* \qquad (16)$$

Now the study sought to posit an index, $I$, such that in an extreme case where all the $\beta_i^*$ are unity,

$$\max_{1 \le i \le p} I = \sum_{i=1}^{p} \beta_i^* = p \qquad (17)$$

Hence for the six grades considered for admission into a tertiary institution in Ghana, with the grade categorisation as in Table 2, the student's in-take grade index (SIGI) could be determined by the formula

$$I_i = \sum_{j=1}^{p} \frac{x_j \beta_j^*}{\omega}, i = 1,2, \dots, N \qquad (18)$$

where $x_j \in X_j$ and $\omega$, a constant, is the highest possible value of the entries, $x_{ij}$'s in the array of the data set.

***Data Analysis and Procedures***

The analysis of the data was done with the help of the SPSS – Statistical Package for Social Sciences, version 19, from IBM SPSS Statistics, R version 3.1.2 and Microsoft Excel and an Excel add-in Real Statistics Resource Pack, from Real Statistics Using Excel. The algorithm for obtaining the SIGI are

1. Perform factor analysis using the principal component method of extraction.
2. Obtain a good factor model for $\beta_{ij}$'s and determine the harmonised rotated factor loadings, $\beta_j^{*\prime}s$.
3. Compute the SIGI's for each individual, $I_i$.
4. Convert the individual's CGPA and SIGI to percentages.

5. Find the difference between the percentages, called $D$.
6. Analyse the $D$ by determining the distribution of $D$.

***Limitations to the Use of SIGI***

The main limitation to the use of the SIGI is when the data are not highly correlated enough, hence factor analysis will not be possible. This is to say that the SIGI is influenced by high redundancy in the data set.

***Delimitation of the Study***

Obtaining the factor loadings are of paramount interest in the use of SIGI; once there is evidence of multicolinearity in the data set, and the variables are multi-dimensional, finding an appropriate label for the unobservable construct underlying the data set is not the goal of SIGI. However, arriving at a good factor model is paramount for determining the harmonised loadings for use in the computation of SIGI.

**Simulation Results and Discussions**

The data obtained for this study is for simulation purposes only; to test equation (18), the SIGI, using the algorithm outlined above. The simulation is therefore organised accordingly.

***Principal Component Factoring of Data***

The first procedure in performing a principal component factor analysis is the assessment of the correlation matrix to ascertain that there exist significant pair-wise correlations among the indicator variables.

**Table 3: Correlation Matrix**

| | | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
|---|---|---|---|---|---|---|---|
| Correlation | $X_1$ | 1.000 | | | | | |
| | $X_2$ | .117 | 1.000 | | | | |
| | $X_3$ | .331 | .296 | 1.000 | | | |
| | $X_4$ | .244 | .164 | .203 | 1.000 | | |
| | $X_5$ | .286 | .208 | .187 | .660 | 1.000 | |
| | $X_6$ | .298 | .240 | .217 | .541 | .712 | 1.000 |
| Sig. (1-tailed) | $X_1$ | | | | | | |
| | $X_2$ | .002 | | | | | |
| | $X_3$ | .000 | .000 | | | | |
| | $X_4$ | .000 | .000 | .000 | | | |
| | $X_5$ | .000 | .000 | .000 | .000 | | |
| | $X_6$ | .000 | .000 | .000 | .000 | .000 | |

*Source: SPSS analysis of original data*

The correlation table suggests that the data is appropriate for factor analysis, since the correlations are significant. The eigen value-greater-than-one rule and the scree plot also suggest an initial two factor model as indicated in the scree plot below.

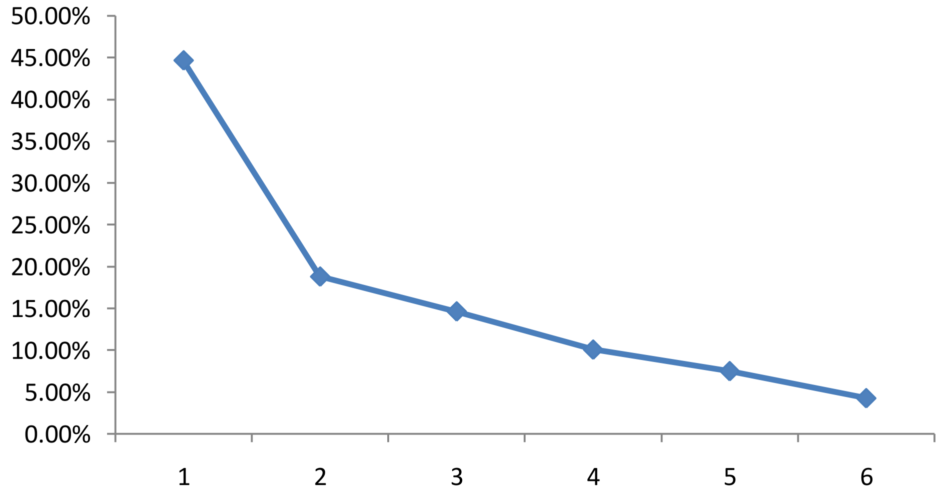


**Fig-1: Scree Plot Depicting Number of Initial Factors to Retain**

Table 3 and Figure 1 are indicating that a two-factor model is adequate in modelling the in-take grades of the category of students included in this study. Parallel analysis from R also suggests that the number of factors = 2 and the number of components = 2. The next steps in the algorithm will seek to confirm this position before the weights are obtained for computing the SIGI.

***Obtaining a Good Factor Model for $\beta_{ij}$'s***

To obtain a good factor model, or to confirm the two factor model, we calculate the residual or error matrix, as in equation (14) above, $\varphi = res_{ij} = R - \Lambda\Lambda'$. The values in this matrix should be close to zero as much as possible for the confirmation process to be achieved.

**Table 4: Residual Matrix**

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
|---|---|---|---|---|---|---|
| $X_1$ | **0.602** | | | | | |
| $X_2$ | -0.283 | **0.561** | | | | |
| $X_3$ | -0.158 | -0.257 | 0.300 | | | |
| $X_4$ | -0.054 | 0.001 | 0.054 | 0.301 | | |
| $X_5$ | -0.045 | 0.023 | 0.016 | -0.103 | 0.168 | |
| $X_6$ | -0.053 | 0.017 | -0.007 | -0.174 | -0.068 | 0.262 |

*Source: SPSS analysis of original data*

With the exception of the two entries (bold-faced), about 90% of the entries are values closer to zero, which is an indication that the two factor model is adequate for the data. The $\beta_{ij}$ 's and $\beta_j^{*'}s$ in this case are in the table below.

**Table 5: Rotated Factor Matrix**

| | $\beta_{ij}$ 's | | $\beta_j^{*'}s$ |
|---|---|---|---|
| | 1 | 2 | Harmonised Loadings |
| $X_1$ | .277 | .567 | 0.567 |
| $X_2$ | .101 | .655 | 0.655 |
| $X_3$ | .057 | .835 | 0.835 |
| $X_4$ | .827 | .122 | 0.827 |
| $X_5$ | .901 | .144 | 0.901 |
| $X_6$ | .832 | .213 | 0.832 |

*Source: SPSS analysis of original data*

By the varimax rotation option, we obtain loadings for each variable which are the weights assigned to each variable when the correlation matrix of the raw grades of the 618 students were analysed. From the loadings, the harmonised loadings, $\beta_j^{*'}s$, are deduced too.

***Computing the SIGIs***

From equation (18), the SIGI for the $i^{th}$ student, $I_i = \sum_{j=1}^{p} \frac{x_j \beta_j^*}{5}$, is calculated for the first-ten students as follows.

**Table 6: Computing SIGI**

| | $X_j$ | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | SIGI |
|---|---|---|---|---|---|---|---|---|
| S/N | $\beta_j^*$ | 0.567 | 0.655 | 0.835 | 0.827 | 0.901 | 0.832 | $\boldsymbol{\sum \beta_j^* = 4.617}$ |
| 1 | | 1 | 1 | 1 | 4 | 4 | 4 | 2.46 |
| 2 | | 1 | 1 | 1 | 4 | 1 | 1 | 1.42 |
| 3 | | 1 | 1 | 1 | 4 | 3 | 3 | 2.11 |
| 4 | | 1 | 1 | 3 | 4 | 4 | 4 | 2.79 |
| 5 | | 2 | 1 | 4 | 4 | 3 | 5 | 3.06 |
| 6 | | 2 | 1 | 3 | 4 | 3 | 3 | 2.56 |
| 7 | | 1 | 1 | 2 | 4 | 3 | 3 | 2.28 |
| 8 | | 2 | 1 | 3 | 4 | 4 | 3 | 2.74 |
| 9 | | 3 | 1 | 1 | 5 | 4 | 4 | 2.85 |
| 10 | | 2 | 1 | 1 | 1 | 3 | 1 | 1.40 |

*Source: Excel output of original data*

If it is assumed that all students admitted into the polytechnic have something in common, then the values in the SIGI column seek to measure the student's output level before enrolling in the polytechnic.

**Table 7: CGPAs Verses SIGIs**

| *S/N* | *SIGI* | *CGPA* | *%SIGI* | *%CGPA* | *D* |
|---|---|---|---|---|---|
| 1 | 2.46 | 2.02 | 53.27 | 40.40 | -12.87 |
| 2 | 1.42 | 2.71 | 30.75 | 54.20 | 23.45 |
| 3 | 2.11 | 2.66 | 45.76 | 53.20 | 7.44 |
| 4 | 2.79 | 2.65 | 60.50 | 53.00 | -7.50 |
| 5 | 3.06 | 2.35 | 66.28 | 47.00 | -19.28 |
| 6 | 2.56 | 1.16 | 55.45 | 23.20 | -32.25 |
| 7 | 2.28 | 1.78 | 49.38 | 35.60 | -13.78 |
| 8 | 2.74 | 1.97 | 59.35 | 39.40 | -19.95 |
| 9 | 2.85 | 2.71 | 61.76 | 54.20 | -7.56 |
| 10 | 1.40 | 2.63 | 30.26 | 52.60 | 22.34 |

*Source: Excel output of original data*

Again, if it is assumed that all students admitted into the polytechnic are predisposed to some common factors during their academic pursuit, then CGPA may give an optimal measure of their output during tertiary education. If the two assumptions are anything to go by, then the difference between the two measures should tell the level of output during tertiary education. The values under column *D* in Table 7, is therefore an indication of this level of output. The distribution of the differences can then be analysed separately as follows.

***The Distribution of the Differences***

The distribution of the differences for all the 618 students considered for this study would subsequently be required for making comparisons specific to departments, institutions, etc - here the distribution of differences is made specific to departments. Before any comparisons are made, the distribution of the differences, in general, is presented in the histogram below.

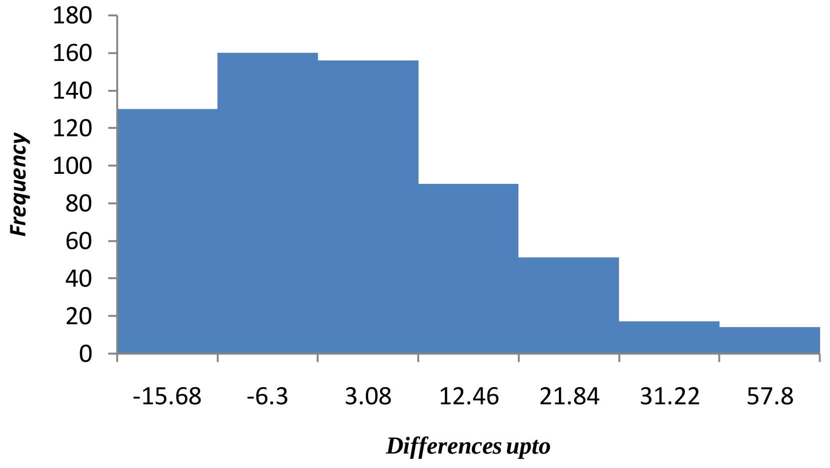


**Fig-2: Distribution of Differences**

The histogram of the differences, which is skewed to the right, suggests that there were more students, over the three year period, whose level of output during the tertiary education has dropped or increased marginally than those who had the same increased tremendously. Hence scale for assessing their level of out during tertiary education could be defined as follows. This scale could then be used to determine the distribution of differences specific to departments.

**Table 8: Definition of Scale for Difference**

| *Scale* | *Limits(At least)* | *Description* |
|---|---|---|
| 1 | $-48.11$ | Dropped drastically(DD) |
| 2 | $-6.3 = 12.46 - 2(9.38)$ | Dropped quite drastically(DQD) |
| 3 | $3.08 = 12.46 - 9.38$ | Dropped(D) |
| 4 | $12.46$ | Better(B) |
| 5 | $21.84 = 12.46 + 9.38$ | Improved(I) |
| 6 | $31.22 = 12.46 + 2(9.38)$ | Improved quite tremendously(IQT) |
| 7 | $40.14 = 12.46 + 3(9.38)$ | Improved tremendously(IT) |
| $Mean\ Difference, \mu = 12.46$ | | $Minimum\ Difference = -48.11$ |
| $Standard\ deviation, \sigma = 9.38$ | | $Maximum\ Difference = 57.50$ |

*Source: Excel output of original data*

The limits in Table 8 were obtained using the number of standard deviations from the absolute mean differences. With these, the distribution of differences specific to departments is obtained using the chi-square contingency table.

**Table 9: Distribution of Differences Specific to Departments**

| | Level of Output | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Department | DD | DQD | D | B | I | IQT | IT | Total |
| A | 171 | 14 | 4 | 45 | 3 | 1 | 0 | 238 |
| | 71.8% | 5.9% | 1.7% | 18.9% | 1.3% | .4% | .0% | 100.0% |
| AE | 9 | 7 | 3 | 7 | 1 | 0 | 1 | 28 |
| | 32.1% | 25.0% | 10.7% | 25.0% | 3.6% | .0% | 3.6% | 100.0% |
| BT | 13 | 22 | 10 | 17 | 7 | 5 | 1 | 75 |
| | 17.3% | 29.3% | 13.3% | 22.7% | **9.3%** | **6.7%** | **1.3%** | 100.0% |
| EEE | 12 | 8 | 10 | 18 | 0 | 1 | 0 | 49 |
| | 24.5% | 16.3% | 20.4% | 36.7% | .0% | 2.0% | .0% | 100.0% |
| FDM | 3 | 5 | 1 | 9 | 1 | 0 | 0 | 19 |
| | 15.8% | 26.3% | 5.3% | 47.4% | 5.3% | .0% | .0% | 100.0% |
| HCIM | 10 | 11 | 5 | 6 | 1 | 2 | 1 | 36 |
| | 27.8% | 30.6% | 13.9% | 16.7% | **2.8%** | **5.6%** | **2.8%** | 100.0% |
| M | 57 | 8 | 6 | 33 | 1 | 0 | 0 | 105 |
| | 54.3% | 7.6% | 5.7% | 31.4% | 1.0% | .0% | .0% | 100.0% |
| PE | 4 | 4 | 1 | 2 | 0 | 0 | 0 | 11 |
| | 36.4% | 36.4% | 9.1% | 18.2% | .0% | .0% | .0% | 100.0% |
| S | 4 | 4 | 3 | 10 | 1 | 0 | 0 | 22 |
| | 18.2% | 18.2% | 13.6% | 45.5% | 4.5% | .0% | .0% | 100.0% |
| SMS | 7 | 7 | 8 | 9 | 2 | 1 | 1 | 35 |
| | 20.0% | 20.0% | 22.9% | 25.7% | **5.7%** | **2.9%** | **2.9%** | 100.0% |
| Total | 290 | 156 | 90 | 51 | 17 | 10 | 4 | 618 |
| | 46.9% | 25.2% | 14.6% | 8.3% | 2.8% | 1.6% | .6% | 100.0% |

*Source: SPSS analysis of original data*

Observing Table 9 will bring out many more deductions from the data specific to various departments as well as the general output levels of the departments put together. For example, over the period under review, there were improvements in output levels for students in the Building Technology; Hotel Catering and Institutional Management; and Secretaryship and Management Studies departments than there were in other departments. On a whole, there were more drastic drops, quite drastic drops and general drops than there were better, improve, quite improve or tremendous improvements.

## CONCLUSION, IMPLICATION AND RECOMMENDATION

The study proposed an algorithm that can be used to calculate a student's in-take grade index, dubbed SIGI. The index used the procedures of performing principal component factor analysis to obtained weights specific to the six subject areas used to admit students into tertiary institutions, by subjecting the entire data of a group of students to factor analysis procedures. These weights are subsequently used in the final formula to compute the SIGI, after testing the fitness of the factor model. The model is deemed robust since it has a validation test incorporated. The model, the index formula, is then used to compute an index specific to each individual; the comparison between the student's in-take grade and the final CGPA is obtained by converting the two measures to percentage of their highest possible values; the difference between the two are then used for making comparisons.

With reference to the data used to test the model, it was found that, during the past three years of tertiary education in Ho Polytechnic, using data obtained for some 618 students, the level of output during tertiary education has dropped or increased marginally than those who had the same increased tremendously. The outcomes reached have a number of implications;

1. That the common factors do not have enough positive impact on many of the students, or
2. That the common factor rather had a lot more negative impact on many students.

The algorithm proposed would help tertiary institutions to frequently assess the level of output of their students on year-on-year basis so that appropriate interventions could be made for better academic performance. The model could be of great help to regulatory authorities in assessing tertiary institutions. The model is hereby recommended to stakeholders in the educational front.

## DIRECTION FOR FURTHER STUDY

The focus of a subsequent study would be to apply similar robust methods, used here, to diagnose the current GPA system as a measure of the student's academic performance in terms of its robustness, validity, fairness and accuracy. This is in line with proposing an alternate grading system which satisfies the criteria of robustness, validity, fairness and accuracy.

## REFERENCES


1. Chickering AW; Education and identity . San Francisco : Jossey-Bass. 1969. Retrieved November 10, 2014 from http://www.clarku.edu/undergraduate/parents/chickering.cfm.
2. Umar SS, Shaib DN, Yakubu NA, Bada O; The Effect of Social Factors on Students' Academic Performance in Nigerian Tertiary Institutions. Library Philosophy and Practice, 2010; 1522-0222: 1 – 12.
3. Cohen-Schotanus J, Muijtjens AMM, Reinders JJ, Agsteribbe J, Van Rossum HJM, Van Der Vleuten CPM; The predictive validity of grade point average scores in a partial lottery medical school admission system. Blackwell Publishing Ltd., Medical Education, 2006; 40: 1012-1019.
4. Sharma S; Factor Analysis. In: Applied Multivariate Techniques (1st Edition). New York: John Wiley & Sons, Inc, 1996.
5. Rencher AC; Factor Analysis. In: Methods of Multivariate Analysis (2nd Edition). New York: John Wiley & Sons, Inc., 2002.
6. Cheesman J, Simpson N, Wint AG; Determinant of Student Performance at University: Reflections from the Caribbean. University of West Indies. 2006. Retrieved November 7, 2014 from http://www.mona.uwi.edu/opair/research/student-performance-paper-revised.pdf.
7. Wright JJ; Reported Personal Stress Sources And Adjustment Of Entering Freshmen. Journal of Counseling Psychology, 1967; 14 (4): 371-373.
8. Romano JL; Psycho Educational Interventions for Stress Management and Well-Being. Journal of Counselling and Development, 1992; 71:199-202.
9. Fabiyi A, Uzoka N; State of physical facilities in Nigerian universities: Implication for repositioning tertiary institutions for global competition. *2009;* Retrieved November, 2014 from http://researcharticlebase.blogspot.com/2011/11/the-influence-of-academic-facilities-on.html.
10. Earthman GI; School Facility Conditions And Student Academic Achievement. Los Angeles, CA: UCLA's Institute for Democracy, Education & Access (IDEA). 2002.
11. Rencher AC; Introduction: Why Multivariate Analysis. In: Methods of Multivariate Analysis (2nd Edition). New York: John Wiley & Sons, Inc. 2002.