# Use of Factor Model to Study and Analyze the Secondary Education Situation in South Darfur State

*Paper prepared by:*
*Dr. Abdelgalal Osman Idris Abaker*[1]

## Abstract

This research is about secondary education in South Darfur State using factor analysis. The problem is: what is the underlying structure of the secondary education in South Darfur State. The study population is the secondary schools in the State. A sample size of 46 secondary schools was taken. The data was collected using stratified random sampling. The collected data was statistically analyzed using factor analysis. Principal component for extraction and varimax for rotation were used to analyze the correlation matrix using the statistical package for social science SPSS. The purpose of this research is to examine the underlying structure of secondary education in South Darfur State. After review the literature and the data were analyzed, the researcher discusses the results of analysis and come to some acceptable results as follows: Seven factors were extracted to explain 77.7% of the total variance at secondary education. The first two factors are general situation and degree of crowded of classrooms.

[1] Dr. Abdelgalal Osman Idris Abake'r, Professor Assistant of Statistics, Department of Economics, Faculty of Economics and Commercial Studies. University of Nyala. Sudan

## 1- Introduction

This research deals with applying factor analysis as a method of multivariate analysis in secondary education in South Darfur State (SDS). A large number of variables affect the secondary education in general and in South Darfur State in particular as a remote State with special properties. South Darfur State (SDS) is located at the south part of the Darfur region, in Sudan. The number of population in the State is more than four million[2].

The religious education in Darfur region started since the kingdom of Fur tribe and the modern education started after Darfur kingdom was invaded by Britain and then it became a part of Sudan[3]. During the first years of Salvation Revolution, the number of schools has increased at all levels[4]. From 1992 up to 2012, enough number of schools is seen in the State; the system of education has been changed to two levels, (Basic and secondary) instead of three[5]. The secondary level includes two types of education, the academic secondary education and technical secondary education. It extends for three years at the end of which the students sit for the Sudan certificate exams and compete for joining higher education institutions[6]. There are 262 secondary schools of all types in the State, including 102 government schools, and the rest are private schools

**Statement of the problem:** Secondary education in the State has been suffering from many problems; the situation harms the people and hinders their progress or development while the globalization dominates the world. These problems (factors) will be investigated in this research by using factor models. Many of the observed variables affect educational progress and the academic achievement. But the real features of secondary education are scientifically unidentified, so: What are the real features of secondary education in South Darfur State? What is the underlying structure of secondary education in the State? And to what extent could the educational variables be reduced?

---

[2] Central Bureau of Statistics. (2009) WWW.cbs.org.s

[3] Musa, M. A. (1992) The establishment and history of secondary education in Darfur state, Masters in education, University of Khartoum. PP 24 - 35

[4] ibid. P(34)

[5] Ibid. P(74)

[6] Ministry of General education, Educational statistics 2007-2008. P9

**This research aims:** to use the factor analysis to examine the underlying structure of the secondary education in the State. The hypotheses to be tested are that the main factors affect the situation of secondary education do not relate to the environment of the schools, to the teachers, to the students or to all of them.

There were some previous studies using factor analysis in education but no one of them was applied in the State.

## 2- Source of data

Analytic procedures will be used to analyze the data of educational variables collected from secondary schools in South Darfur State in 2009. Questionnaire was used to collect primary data from schools and ministry of education.. Statistical packages such as SPSS will be used to analyze the data.

The stratified random sampling is designed to collect numeric primary data about education from schools in the State. The population of the study is secondary schools in South Darfur State, while the frame of sampling is that of three classes in the State, which are 262 schools and the sample unit is school. Because of Darfur crisis it is too difficult to increase the sample size so the suggested sample size is 50 schools selected randomly (which represent more than 19% of population) and the response is (46) cases.

## 3- Factor Analysis (Statistical Tool)

Factor analysis (FA) is that branch of multivariate analysis which deals with the internal structure of matrices of covariance or correlation. Today (FA) is the most widely used of multivariate techniques, though not always appropriately. Also its use has been facilitated by the advent of electronic computers and is spreading to many disciplines[6].

The application of factor analysis techniques has been used mostly in the field of psychology. It bound up with psychological conceptions of mental factor[7]. In education as in psychology factor analytic techniques have been used extensively, particularly in the field of educational psychology. Since 1950 the applications of factor analysis in other fields have become very popular with the use of computers.

---

[6] D. N. Lawley, A. E. Maxwell, 1963. Factor analysis as statistical method, Butter Worth, Londo.n P(1)

[7] Harry H. Harman, 1976. Modern factor analysis 3rd edition. The University of Chicago Press.P(60)

## 3.1 Matrices and concepts in factor analysis

Matrix algebra provides a mathematical framework to deal with arrays of numbers or algebraic symbols. It is especially useful in the representation and development of methods for modeling and analyzing multivariate data, it allows us to express the common factor model for all individuals and multivariate simultaneously[9]. Data matrix $\underline{X}$ is a matrix of order $n \times p$ each element $X_{ij}$ is a score for the $i^{th}$ individual on the $j^{th}$ variable. Most factor analysis conducted under the condition that the factors are standardized[10]. Standardization eliminates influences of different measurements of data and simplifies interpretation of solution

## 3.2- Correlation matrix and Covariance matrix

A correlation matrix R is a symmetric matrix with units diagonal and off diagonal elements $r_{jj'}$ that satisfy: $-1 \leq r_{jj'} = r_{j'j} \leq 1 \ , \quad j \neq j'$ (This relation is necessary but not sufficient)[11].

Where $r_{jj'}$ represents simple correlation between variables $X_j$ and $X_{j'}$.

Covariance matrix is a symmetric matrix of order p, the diagonal elements $v_{jj}$ represents the variances of the variables $(S^2_{x_j})$, and the off diagonal elements $v_{jj'}$ , $(j \neq j')$ is the covariances of variables $X_j, X_{j'}$ .

There is a relation between the covariance and correlation matrix as below:

$$\Sigma = \underline{D}\left(S_{x_j}\right) \underline{R}\, \underline{D}\left(S_{x_j}\right) \qquad (1)$$

Such that $\underline{D}\left(S_{x_j}\right)$ is a diagonal matrix of order p, its diagonal elements are the elements of the covariance matrix[12].

## 3.3- Eigen values and eigenvectors

Eigenvalue represents the amount of variance accounted for by a factor[13]. Suppose S is a symmetric matrix of order P, if u is a column vector of order P and $\lambda$ is scalar such that $Su = \lambda u$ then, $\lambda$ said to be an eigenvalue

---

[9] Robert MacCallum, 2004. Factor analysis class notes, Ohio State University. P(19)

[10] Ledyard R. Tucker, Robert C. MacCallum, 1997. Exploratory factor analysis, University of Illinois, USA. P(67)

[11] Robert MacCallum Op., cit. P(24, 25)

[12] Amal Alsir Al-khidir, 2005. Statistical study about factors affect the academic achievement of secondary certificate students by using factor analysis (case study: Sudan University of science and technology). P(46)

[13] Joseph F Hair. et al, 2010. Multivariate data analysis A Global perspective, seven edition. Pearson Prentice Hall New Jersey. P(92)

of S, u said to be corresponding eigenvector of S. A symmetric matrix of order P has p eigenvalues that:

$$\lambda_1 \geq \lambda_2 \geq \ldots\ldots \geq \lambda_p \quad \text{——————————} \quad (2)$$

## 3.4- Factor analysis models

Matrix equations are useful for representing linear combination of random variables. Suppose the variables in matrix $\underline{X}$ are linear functions of the variables in matrix $\underline{Z}$ , so it could be represented by the following matrix equation:

$$X = \mu + AZ \text{ , A is coefficient of Z} \quad \text{——————————} \quad (3)$$

By using the rules of matrix multiplication we can write an equation for any variable in $\underline{X}$ , such as:

$$X_1 = \mu_1 + a_{11}Z_1 + a_{12}Z_2 + \cdots + a_{1n}Z_n$$

$$X_2 = \mu_2 + a_{21}Z_1 + a_{22}Z_2 + \cdots + a_{2n}Z_n$$

$$X_p = \mu_p + a_{p1}Z_1 + a_{p2}Z_2 + \ldots + a_{pn}Z_n \quad \text{——————————} \quad (4)$$

Such equations are commonly used to represent linear statistical models as factor analysis[14]. The observable random vector $\underline{X}$ with p components, has mean $\mu$ and covariance matrix $\Sigma$. The factor analysis postulate that $\underline{X}$ is linearly dependent upon a few unobservable random variables $F_1, F_2, \ldots, F_m$ called common factor and P additional sources of variation $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$ called error or specific factors. So the factor analysis model is:

$$X_1 - \mu_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \ldots + \lambda_{1m}F_m + \varepsilon_1$$

$$X_2 - \mu_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \ldots + \lambda_{2m}F_m + \varepsilon_2$$

. . . . . .

[14] Robert acallum. Op. cit. P(sg)

$$X_p - \mu_p = \lambda_{p1}F_1 + \lambda_{p2}F_2 + \ldots + \lambda_{pm}F_m + \varepsilon_p \quad (5)$$

Where: $\mu_i$ = ith variable mean, $F_j$ = jth common factor, $\varepsilon_i$ = ith specific factor

The coefficients $\lambda_{ij}$ called the loading of the ith variable on the jth factor, so the matrix $\Lambda$ is the matrix of factor loadings[15].

Or in matrix notation:

$$(X - \mu)_{p\times 1} = \underline{\Lambda}_{p\times m}\underline{F}_{m\times n} + \underline{E}_{p\times 1} \quad (6)$$

When the data is standardized, which is the usual case in common factor analysis the mean $\mu$ is no longer appears. So the formula could be:

$$\mathbf{X = \Lambda F + E} \quad (7)$$

## 3.5- Factor extraction methods

Principal component is one of the most commonly used in extraction factors, in principal component a set of P correlated variables is transformed to a smaller set of uncorrelated hypothetical constructs called principal components (PCs). Other methods such as maximum likelihood, un-weighted least squares, image analysis, alpha factor analysis, principal axis factoring, the least square method, centroid method and the diagonal method are used to extract factors.

## 3.6- Factor rotation

The term rotation means that the reference axes of the factors are turned about the origin until some other position has been reached. The final effect of rotating the factor matrix is to redistribute the variance from earlier factors to new ones to achieve a simpler, theoretically more meaningful factor pattern.[16]

If there is a matrix of estimated factor loading $\hat{L}$ obtained by any method, the matrix of rotated loading could be obtained by post-multiplication the matrix with transformation matrix $T$. Such that: $T'T = I$

[15] Richard A Johnson, Dean W Wichern 1998 Applied multivariate statistical analysis, Prentice Hall Inc. New Jersey. P(515)

[16] Joseph F Hair Jr etal Op cit P(112,113)

## 3.7- Criteria for number of factors

In the statistical procedure for determining the number of factors to extract; if the residual matrix has no significant variance, then the correct number of factors has already been extracted[17]. Currently, latent root criterion (Kaiser's criterion), a priori criterion, percentage of variance criterion and Scree test criterion (Cattell's criterion) are utilized criteria for the number of factors..

## 4- Results of Analysis

Sometimes we deal with a large number of variables which make it difficult to study or interpret the situation. And sometimes some of these variables have similar characteristics, on one hand that similar ones can be grouped using appropriate statistical techniques. On other hand, the variables may be affected by another type of variables (factors) which cannot be measured or dealt with without the original variables. The data would be analyzed using factor analysis followed by discussion in order to draw conclusions.

The variables of research are about school and its teachers and students which will be defined, for the purpose of analysis, as follow:

| variable | definition | variable | definition |
|---|---|---|---|
| X1 | No. of studying classes | X11 | Teachers average of salary |
| X2 | Date of establishment of the school | X12 | Percentage of teachers graduate from faculty of education |
| X3 | Age of headmaster | X13 | Average of periods for teacher per week |
| X4 | Years of experience of head master | X14 | Average distance of teachers home |
| X5 | No. students in the school | X15 | Average of students age |
| X6 | Average of academic records | X16 | Income of student's father per month |
| X7 | Average of teachers age | X17 | Average of past academic records |
| X8 | Teachers average years of training | X18 | Average No of persons per family |
| X9 | No. of toilet | X19 | Average distance of students home |
| X10 | Teachers average years of experience | | |

The correlation matrix was checked; the results refer that correlation matrix is not Identity matrix, singularity and multi-colinearity are not

[17] Richard L. Gorsuch, 1974. Factor analysis, U. B. Sounders Company Philadelphia. P(135)

problematic for these data and sample size is fit; therefore, factor analysis is appropriate.

The principal component is used as a method of extraction; and Kaiser Criterion (Eigen values equal to one or greater) for the number of factors. For rotation the Varimax method was chosen. The missing values are replaced with the means.

The final analysis was run for a selected nineteen variables by SPSS program. The results of factor analysis are below:

## 4.1- Factor extraction

First we show KMO measure of sampling adequacy and Bartlett's test as in Table (4-1).

Table (4-1): KMO and Bartlett's Test

| Kaiser-Meyer-Olkin Measure of Sampling Adequacy. | | 0.593 |
|---|---|---|
| Bartlett's Test of Sphericity | Approx. Chi-Square | 210.54 |
| | df | 171 |
| | Sig. | 0.021 |

The source: SPSS output

The value of KMO is 0.593, which indicates that the sample size is mediocre, so the sample size is sufficient. The P-value of Bartlett's test is equal 0.021, significant at level of 5%, so the correlation matrix is not Identity matrix; therefore, factor analysis is appropriate[18].

Table (4-2): Communalities

| | Initial | Extraction | | Initial | Extraction |
|---|---|---|---|---|---|
| $Z_1$ | 1 | 0.86781936 | $Z_{11}$ | 1 | 0.698072037 |
| $Z_2$ | 1 | 0.636289776 | $Z_{12}$ | 1 | 0.735894573 |
| $Z_3$ | 1 | 0.918759825 | $Z_{13}$ | 1 | 0.565988541 |
| $Z_4$ | 1 | 0.933886293 | $Z_{14}$ | 1 | 0.77847284 |
| $Z_5$ | 1 | 0.708071143 | $Z_{15}$ | 1 | 0.790384789 |
| $Z_6$ | 1 | 0.853950513 | $Z_{16}$ | 1 | 0.779835881 |
| $Z_7$ | 1 | 0.828323333 | $Z_{17}$ | 1 | 0.796550866 |
| $Z_8$ | 1 | 0.872794221 | $Z_{18}$ | 1 | 0.578436777 |
| $Z_9$ | 1 | 0.767103469 | $Z_{19}$ | 1 | 0.788753134 |
| $Z_{10}$ | 1 | 0.869356163 | | | |

The source: SPSS output

Table (4-2) shows the communalities before and after extraction. Before extraction communalities are all one. Communalities after extraction, reflect the common variance in the data structure, it is the proportion of variance explained by the underlying factors.

[18] Andy Field. 2005 Factor analysis using SPSS. http://www.sagepub.co.uk/field/multiplechoice.html.P(5,6)

Table (4-3) represents the extraction sums of squared loadings. In this table the eigenvalues associated with seven factors and the percentages of variance explained are shown.

Table (4-3): Extraction Sums of Squared Loadings

| | Extraction Sums of Squared Loadings | | |
|---|---|---|---|
| Component | Total | % of Variance | Cumulative % |
| 1 | 3.966958722 | 20.87873012 | 20.87873 |
| 2 | 3.316163851 | 17.45349395 | 38.33222 |
| 3 | 2.37282537 | 12.48855458 | 50.82078 |
| 4 | 1.461082231 | 7.68990648 | 58.51069 |
| 5 | 1.257506511 | 6.618455322 | 65.12914 |
| 6 | 1.231522161 | 6.481695584 | 71.61084 |
| 7 | 1.162684687 | 6.119393091 | 77.73023 |

The source: SPSS output

Table (4-4) represents the rotation sums of squared loadings. The eigenvalues after rotation are displayed in column two, the percentage variance in column three and the cumulative percentage of variance in column four.

Table (4-4): Rotation sums of the squared loadings

| Rotation Sums of Squared Loadings | | | |
|---|---|---|---|
| Component | Total | % of Variance | Cumulative % |
| 1 | 2.828418 | 14.88641 | 14.88641 |
| 2 | 2.758525 | 14.51855 | 29.40496 |
| 3 | 2.419263 | 12.73296 | 42.13793 |
| 4 | 2.04436 | 10.75979 | 52.89772 |
| 5 | 1.717636 | 9.040189 | 61.93791 |
| 6 | 1.526549 | 8.034467 | 69.97237 |
| 7 | 1.473992 | 7.757854 | 77.73023 |

The source: SPSS output

Rotation enables us to interpret the factor structure easily, and one of the consequences of rotation for these data is that the variances of the seven factors were redistributed. For instance, before rotation the first two factors accounted for 37% and the remaining accounted for 40%. However after rotation the factors from one to seven are account for (14.8%, 14.5%, 12.7%, 10.7%, 9%, 8% and 7.7%) respectively.

The Scree plot suggests four or seven factors which matching the results of Kaiser Criterion.

Table (4-5): Factor structure for secondary level

| | Factor 1 | Factor 2 | Factor 3 | Factor 4 | Factor 5 | Factor 6 | Factor 7 |
|---|---|---|---|---|---|---|---|
| 1 | $X_{16}$ | $X_{19}$ | $X_7$ | $X_9$ | $X_{12}$ | $X_{15}$ | $X_{15}$ |
| 2 | $X_{13}$ | $X_1$ | $X_{10}$ | $X_4$ | $X_5$ | $X_{17}$ | $X_6$ |
| 3 | $X_{12}$ | $X_5$ | $X_{11}$ | $X_3$ | $X_{11}$ | $X_{29}$ | $X_2$ |
| 4 | $X_{15}$ | $X_9$ | $X_8$ | | $X_{14}$ | | |
| 5 | $X_1$ | $X_{14}$ | | | $X_{19}$ | | |
| 6 | $X_{18}$ | | | | | | |
| 7 | $X_{17}$ | | | | | | |
| 8 | $X_8$ | | | | | | |

The source: prepared by researcher according the SPSS output

## 5- Discussion and conclusions

Interpretations of factors, deal with two main points: the first is to discuss the nature and origin of factors and the second is the correlation between factors.

We use the factor loadings to understand what the factors are, and to name them. High loadings (0.3)[19] show which variables are strongly influenced by the factor. Near zero loadings show which variables are not influenced appreciably by the factor. A characteristic that is shared by the variables that have high loadings will suggest what effect the factor has and enable the investigator to name that factor.

### 5.1- Interpretation of the factors

According to the Kaiser criterion seven factors were extracted to explain a proportion of 77.7% of the total variance of secondary education in South Darfur as shown in Table (4-4).The most important factors that explained most of the variance in the secondary education are: factor one, two and three.

First factor:

As shown in the results of orthogonal rotation of the factors as stated in Table (4-4), the first factor explains proportion of 14.8% of the total variance in secondary education. This factor is loaded by a number of eight variables. The monthly average income of the student's father loaded by -0.86, followed by average weekly periods of teacher by 0.71,

[19] Ibid P(4)

these two variables rank as the most important variables of this factor. The rest of the variables are the proportion of teachers who have graduated from faculties of education, student's age, the average past academic record of the student, the number of classrooms in the school, average years of teacher's training and the average number of family members of the student. The loadings are 0.58, 0.54, -0.47, 0.37, 0.33, 0.3, respectively. It is observed that the components of this factor are not related to a certain field; they involve variables relate to student and his family, the teacher and the school. This factor reflects the "general situation".

Second factor:

The second factor of the factor structure explains 14.5% of the whole variance of secondary education as shown in Table (4-4). The structure of this factor is composed of five variables. The distance between student's residence and school represents the most important variable; its loading is 0.86. The most unimportant variable is the distance between teacher's residence and school, which has only 0.4 of loading. In addition the number of classrooms 0.76, the number of students in school 0.73, and the number of toilets in school 0.66. The distance between the school and residence affects directly the access time to the school. Also the degree of crowdedness in the classrooms affects academic achievement. Therefore, this factor could be the "degree of classroom's crowdedness and the distance of residence".

Third factor:

Results of factor analysis after varimax rotation show that the third factor explains the proportion of 12.7% of the total variance as shown in Table (4-4). The structure of this factor is of four variables: all are related to the teacher. These variables are, the age of the teacher, the average years of teachers' experience, the teacher's salary and the average number of years of teacher's training. The loadings are 0.91, 0.84, 0.71, 0.34, respectively. As we know the teacher is the cornerstone of the educational process, because the well trained and consent teacher would lead to teacher's ability improvement and thus, the success of the educational process. The most suitable name for this factor is the "teacher".

Fourth factor:
Table (4-4) indicates that the fourth factor in the factor structure explains 10.7% of the total variance. Only three manifest variables are loaded on this factor. Two third of these variables relate to the administration. The loading of headmaster's experience reach to 0.92, beside the age of headmaster which reaches to 0.9. While the third variable loads only 0.31. The effective administrator can use the available possibilities and thus attains a good education. The most appropriate name for this factor is "administration".

Fifth factor:
The results of the rotation axis for the factors as in the Table (4-4) show that, the fifth factor explains 9% of the total variance. The structure of this factor consists of five observed variables. The greater loading which belongs to the teacher's residence distance from the school loads 0.75, followed by the number of family members of the student -0.61. Variables of this factor are related to the environment that surrounds the teacher and the student. The researcher calls this factor "environment round the teacher and student".

Sixth and Seventh factor:
The last two factors (the less important) explain 8%, 7.7% of the total variance; any one contains three manifest variables. The most important variable for the sixth factor is the student's previous achievement 0.73. Accordingly, this factor could be called the "background of the student and the school training". On the seventh factor, the academic achievement has great significance loading 0.88, so the researcher prefers to name this factor "The school academic achievement".

**5-2 Conclusions**
After collecting the data of education in SDS for this study; the researcher analyzed the data. After the results of the analysis were discussed some findings were drawn. These findings are considered as generalizable. Below are the findings:
The results of the factor analysis for the data show the underlying structure of the secondary education, which consist of seven unobservable variables (factors). About 77.7% of the total variance is explained by these factors.

The most important factors are: general situation, the crowded of classroom and the distance of residence. These three factors explain more than 42% of the total variance.

Eight manifest variables loaded on the first factor such as average of students' fathers income per month, weekly periods for the teacher, percentage of graduated teachers from faculties of education, average of student's age, past academic record, number of studying classes, average years of teachers' training and the number of individuals in each student's family.

The second important factor has five manifest variables in its structure that are: distance of students residence, number of studying classes, number of students in the school, number of toilets and distance of teachers residence.

Then four manifest variables, average of teacher's age, teachers' years of experience, teachers' salary, and teachers' years of training are loaded on the third factor.

## 6- Recommendations

On the basis of findings, the researcher recommended the following:

- The educational authorities should use the results of this research as the starting point to uncover the most important unobserved component of secondary education.
- The education authorities should make educational database for secondary education in South Darfur State.
- Statisticians and educational researchers should make join researches about the secondary education in the State using factor analysis.

- This research could be re-implicated by using another sampling unit in the field of education instead of school so as to gain more reliable results about unobserved variables (factors).

## 7- References

1- Amal Alsir Al-khidir, 2005. Statistical study about factors affect the academic achievement of secondary certificate students by using factor analysis (case study: Sudan University of science and technology).

2- Andy Field. 2005. Factor analysis using SPSS. http://www.sagepub.co.uk/field/multiplechoice.html
3- Central Bureau of Statistics , (2009) WWW.cbs.org.s
4- D. N. Lawley, A. E. Maxwell, 1963. Factor analysis as statistical method, Butter Worth, Londo.n
5- Harry H. Harman, 1976. Modern factor analysis 3rd edition. The University of Chicago Press.
6- Joseph F. Hair, et al, 2010. Multivariate data analysis A Global perspective, seven edition. Pearson Prentice Hall New Jersey.
7- Ledyard R. Tucker, Robert C. MacCallum, 1997. Exploratory factor analysis, University of Illinois, USA.
8- Ministry of General education, Educational statistics 2007-2008.
9- Musa, M. A. (1992). The establishment and history of secondary education in Darfur state, Masters in education, University of Khartoum.
10- Richard A. Johnson, Dean W. Wichern. 1998. Applied multivariate statistical analysis, Prentice Hall Inc. New Jersey.
11- Richard L. Gorsuch, 1974. Factor analysis, U. B. Sounders Company Philadelphia.
12- Robert MacCallum, 2004. Factor analysis class notes, Ohio State University.