Original Research Article

Neutrosophic-Principal Component Analysis of Causes of Performance Gap among Private and Public School Students in the Basic Education Certificate Examination

Abstract
The decade - old academic achievement discrepancies at the basic school level that are widely emphasized are still being worked on heavily in current research. In this study, we present a novel Neutrosophic-Principal Component Analysis and Two-way Neutrosophic ANOVA to analysing the causes of Performance Gap among private and public school students in the Basic Education Certificate Examination. One-hundred and eighty-nine respondents from the Ada East and West Districts of Ghana were involved in the study. We present a modified Neutrosophic regression equivalence of the classical least squares to solve problems of indeterminacy. The results showed that of the total 87% variability, three main components—student characteristics, instructor characteristics, and administrative-logistic features—accounted for 36%, 31%, and 20%, respectively. The remaining 13% of the variability, which was attributed to random effects by the Neutrosophic PCA technique, was tested using a two-way Neutrosophic ANOVA where the results identified that interaction between the identified factors was a contributing factor that needed to be explained.Our results reveal that Neutrosophic-PCA is a potential method to lessen human response errors, which are frequently tainted with ambiguous, conflicting, imprecise, indeterminate or uncertainty.When there is just a slight difference between the two options, the forced selection caused by the traditional Likert method will be eliminated by our new approach.When studying respondents' opinions, we recommend researchers to employ Neutrosophic- PCA in order to minimize any bias that may result from the responses' lack of precision.

Keywords. Neutrosophic, Principal Component Analysis, BECE, Deneutrosophication, Performance Gap, Basic School.

UNDER PEER REVIEW

1. Introduction

Disparities in academic achievements among basic school learners learning under same treatmental and learning conditions have been studied and found to differ [1-5] .Teaching Methods and teachers' effectiveness in using content knowledge, professional values and skills have also been found to control basic school learners' attitude and believes about the subjects they learn at school [6-12].

In a society where parents continue to promote sending their children to private schools to improve their academic performance, the age-old issue of maximizing performance of public school pupils in the Basic School Certificate Examination is still an issue [13]. According to the Ghana Pre-Tertiary School Performance Inspection Report for the 2019–2020 Academic Year, learning outcomes in private schools were superior to those in public schools. The survey also discovered that although public school head teachers spend a lot of time preparing for their institutions, they rarely carry out their plans or oversee their execution.

Private school head teachers, on the other hand, spent less time documenting plans and more time focusing on monitoring to achieve academic objectives. The report further noted that Teacher-absenteeism was twice as much in the public school than private school. Pupil absenteeism was also 3-fold higher in public than private basic schools [14]. Stakeholders may better provide for and improve performance in both sectors when the mechanisms causing this gap are identified and well understood.

[31] analysed the Trend of Junior High School Pupils' Performance in the Basic Education Certificate Examination (BECE) in Ghana. They found private schools to have performed 5.1% and 8.1% higher in the years 2014 and 2016 in two districts alone. [15] conducted research on junior high school academic achievement in Ghana's public and private schools. According to the study's findings, there are a variety of factors that contribute to private schools in the Tamale Metropolis performing academically better than their public counterparts. These include the fact that private schools had better resources and students whose parents had higher socioeconomic standing, meaning they were more engaged in their children' education. The academic performance of children in private schools was linked to the stringent internal control of the school heads/proprietors because teacher motivation was fairly low in both public and private selected schools.

Although considerable effort has been put into identifying the factors that contribute to performance gaps between private and public elementary schools in scientific literature, such as [15] and [16], little has been done to carry out high dimensional analysis that aims to quantify the contribution of the factors by the variation explaining each element. Few recent works that used High dimensional analysis to study learner disparities in performance did not incorporate Neutrosophic analysis to handle errors due to indeterminacy [17, 18, 33].

The most used psychometric tool for getting responses from survey participants is the Likert scale Due to its ordinal nature and closed format; it is typically linked to information distortion and information loss problems. In general, responses depend on the respondents' emotions and are therefore inconsistent, imprecise, and ambiguous. Neutosophy is a notion that is used to genuinely represent the answers by properly accounting for inconsistency, uncertainty, imprecision and indeterminate information which is mostly associated with Likert scale methods [19-21].

A useful method for categorizing respondents and targeting them appropriately is clustering according to their input. Indeterminate Likert scaling performs better in capturing the responses when dealing with real-world settings. As they must always select a Likert scale that can convey a range of emotions, respondents may be asked to select from a list of options that range from strongly agree to strongly disagree in order to communicate their opinions on the things being measured by the questionnaire. Respondents may select the option that garners the most support, "strongly agree," while dismissing any slight or insignificant level of dissent. Conversely, respondents may select "agree," which would result in the least amount of disagreement. Thus, the specific perception of the respondent expressing the precise degree of agreement or disagreement is not measured by the Likert scale used in the questionnaire [22-25].

Neutrosophic Likert scaling will eliminate the need to select the most popular alternative or a forced selection, which is not necessarily true if there is only a slight difference between the two options. Human emotions, which are frequently ambiguous, conflicting, imprecise, or indeterminate in nature and necessitate neutrosophic analysis to handle such issues, are the main source of the responses from the respondents [26-30].

The following properties of the factors taken into account in this study necessitate the use of principal component analysis. First, there is a good likelihood that some of the numerous variables we have are measuring the same underlying entity. Once more, they are probably highly associated. As a result, we must include components in our assessment scale that we believe well accurately represent the construct while excluding those that do not. Second, a new measurement scale might be required, but we do not even know whether the elements we have included properly capture the construct that interests us.

Therefore, in order to determine if your factors are adequately representative of the construct of interest or whether they should be removed from the new measurement scale, we must determine whether the construct of interest "loads" onto all (or just some) of your factors. Finally, we may want to determine whether our measurement scale, such as a questionaire, can be summarized to include fewer items because some of those items may be superfluous (i.e., more than one item may be measuring the same construct), and/or we may want to create a measurement scale that is more likely to be completed [32].

The idea of fuzzy sets was developed to deal with object uncertainty. The extension of fuzzy sets includes fuzzy topological space, intuitionistic fuzzy sets in topological structure space, vagueness in topological structure space, rough sets in topological space, the theory of hesitation, neutrosophic topological space, etc. A family of parameters known as a "soft set" is also a set. Soft sets are combined with different topological structures to create fuzzy soft topological space, intuitionistic fuzzy soft, and neutrosophic soft topological space.Numerous mathematical and non-mathematical disciplines, including operations research, physics, data science, etc., have applications for topological space. It can be challenging to apply the notion of topology in real-world situations at times because to elemental uncertainties, inconsistencies, and incompleteness of knowledge. Intuitionistic fuzzy soft topological space was developed to address several issues that arise in fuzzy soft topology, and fuzzy soft topological space was first introduced to overcome the challenge in classical sets that deals with object uncertainty.In some situations where an object's value is ambiguous, the previous techniques cannot be applied. Neutrosophic set has been developed as a result to deal with the uncertainty, inconsistency, and incompleteness [33-36].

Our approach in this current work sought to combine two powerful approaches connected to survey studies, Principal Component Analysis and Neutrosophic Statistics. Indeterminacy and high correlation issues, which are inherent to the nature of the data in this kind of study, will be alleviated as a result.

2. Method

For the purpose of minimizing errors of indeterminacy and uncertainty, we provide a modified Neutrosophic variant of the traditional least squares regression. The Neutrosophic PCA combined technique has the ability to eliminate respondent bias while directing us to focus on the components that account for the majority of the underlying problem variations without sacrificing generality. Overall, 189 responses from both public and private schools were used.

UNDER PEER REVIEW

2.1 **Construction of Principal Components**

For a random vector, say, X , with domain $\boldsymbol{R^m}$ **,** will have a mean and covariance matrix of μ_X and $\sum_X$, respectively. $\lambda_1 > \lambda_2 > \cdots > \lambda_m > 0$ for an array of eigenvalues of $\sum_X$, so that the i -th eigenvalue of $\sum_X$ represents the largest i -th eigenvalue. Again, suppose a vector α_i denotes the i -th eigenvector of $\sum_X$ corresponding to the i -th eigenvalue of $\sum_X$. We wish to derive principal components (PCs) form by considering the maximization of $var[\propto_1^T X] = \propto_1^T \sum_X \propto_1$, with respect to $\propto_1^T \propto_1 = 1$ (a typical optimization problem).
The Lagrange multiplier approach is then applied to solve the problem.
To that end,

$$L(\alpha_1, \phi_1) = \alpha_1^T \sum_X \alpha_1 + \phi_1(\alpha_1^T \alpha_1 - 1) \qquad (1)$$

$$\frac{\partial L}{\partial \alpha_1} = 2\sum_X \alpha_1 + 2\phi_1\alpha_1 = 0 \Rightarrow \sum_X \alpha_1 = -\phi_1\alpha_1 \Rightarrow \text{var}[\alpha_1^T X] = -\phi_1\alpha_1^T \alpha_1 = -\phi_1 .$$

Since $-\phi_1$ represent the eigenvalue of $\sum_X$, with α_1 denoting the respective normalized eigenvector, $\text{var}[\alpha_1^T X]$ is maximized when α_1 chosen as the initial eigenvector of $\sum_X$. To this end, $z_1 = \alpha_1^T X$ is referred to as the first PC of X , with α_1 representing the vector of coefficients for z_1, where $\text{var}(z_1) = \lambda_1$.

To get the second PC, $z_2 = \alpha_2^T X$, we shall maximize $\text{var}[\alpha_2^T X] = \alpha_2^T \sum_X \alpha_2$ on condition that z_2 is not correlated with z_1. But $\text{cov}(\alpha_1^T X, \alpha_2^T X) = 0 \Rightarrow \alpha_1^T \sum_X \alpha_2 = 0 \Rightarrow \alpha_1^T \alpha_2 = 0$, which we will solve by maximizing $\alpha_2^T \sum_X \alpha_2$, on condition that $\alpha_1^T \alpha_2 = 0$, and $\alpha_2^T \alpha_2 = 1$.
We again make use of the Lagrange multiplier approach.
To that end,

$$L(\alpha_2, \phi_1, \phi_2) = \alpha_2^T \sum_X \alpha_2 + \phi_1\alpha_1^T \alpha_2 + \phi_2(\alpha_2^T \alpha_2 - 1) \qquad (2)$$

$$\frac{\partial L}{\partial \alpha_2} = 2\sum_X \alpha_2 + \phi_1\alpha_1 + 2\phi_2\alpha_2 = 0$$

$$\Rightarrow \alpha_1^T (2\sum_X \alpha_2 + \phi_1\alpha_1 + 2\phi_2\alpha_2) = 0 \Rightarrow \phi_1 = 0$$

$$\Rightarrow \sum_X \alpha_2 = -\phi_2\alpha_2 \Rightarrow \alpha_2^T \sum_X \alpha_2 = -\phi_2 .$$

As $-\phi_2$ is the eigenvalue of $\sum_X$, where α_2 is the respective normalized eigenvector, we are able to maximize $\text{var}[\alpha_2^T X]$ when we select α_2 as the second eigenvector of $\sum_X$. As a result, $z_2 = \alpha_2^T X$ becomes the second PC of X , where α_2 represents the vector of coefficients for z_2, and $\text{var}(z_2) = \lambda_2$. Per the above results, we can deduce that the i -th PC $z_i = \alpha_i^T X$ is constructed α_i is chosen as the i -th eigenvector of $\sum_X$, which will then have the variance λ_i. We can conclude by the above results that PCA are the only set of linear functions of original data that are uncorrelated and have orthogonal vectors of coefficients.

PCA relies on either covariance matrix or the correlation matrix. The linear combination weights directly originate from combination eigenvectors of correlation matrix or covariance matrix.

Recall that for m variables, the $m \times m$ covariance or correlation matrix will contain the following sets:

$$\text{m eigenvalues } -\{l_1, l_2, \ . \ . \ . \ , l_p\}$$
$$\text{m eigenvectors } -\{\mathbf{e}_1, \mathbf{e}_2, \ . \ . \ . \ , \mathbf{e}_p\}.$$

We form each principal component (PC) when we consider the values of the elements of the eigenvalues as the weights of the linear combination.

Assuming that the k-theigenvector $\mathbf{e}_k = \left(e_{1k}, \ e_{2k}, \ \ldots, \ e_{pk}\right)$, then the PCs $Y_1, \ \ldots,$ are produced by

$$Y_1 = e_{11}X_1 + e_{21}X_2 + \ . \ . \ . \ + e_{m1}X_m$$

$$Y_2 = e_{12}X_1 + e_{22}X_2 + \ . \ . \ . \ + e_{m2}X_m ...$$

$$Y_m = e_{1m}X_1 + e_{2m}X_2 + \ . \ . \ . \ + e_{mm}X_m \qquad (3)$$

The algorithm for Construction of Principal Components was proposed in the work of[7-8].

2.2 Variability within cells, variability resulting from the interaction of the two factors, variability among the levels of the two factors, and overall variability are the four components of a two-way ANOVA (error variability). The first three sources of variability—variability owing to the first component, variability due to the second factor, and variability due to interaction—are compared to the error variability using three different statistical tests (based on the F statistic). The resulting p-value of each test assists in our assessment of the significance of that specific influence. Non-metric data is measured using binary, nominal, or ordinal scales; the spacing between the scale values has no practical significance. Metric data, however, might be discrete or continuous.

$$One-Way\ ANOVA \qquad\qquad Two-Way\ \ ANOVA$$
$$SS_y = SS_x + SS_{error} \qquad\qquad SS_y = SS_{x1} + SS_{x2} + SS_{x1x2} + SS_{error}$$

Since our model has interaction factors coupled with the fact that our dataset has both metric and nonmetric factors, two-way ANOVA is appropriate based on Model, Interaction and Main effect hypotheses.

Model Hypothesis: Ho: it is not suitable to perform two-way ANOVA; H1: It is suitable to perform two-way ANOVA. A significant P-value ($p<0.05$) will lead to the rejection of the null hypothesis to conclude that Two-way ANOVA is appropriate. Interaction Hypothesis: Ho: no Interaction effect exist between factors; H1: interaction effect exist between factors. A significant P-value ($p<0.05$) will lead to the rejection of the null hypothesis to conclude that there is interaction between factors.

$$\text{Main Effect Hypothesis}$$
$$Factor1\ (say, Gender) \qquad\qquad Factor\ 2(say, \text{Re}ligion)$$
$$H_o : \mu_1 = \mu_2 = \mu_3 ... = \mu_k \qquad\qquad H_o : \mu_1 = \mu_2 = \mu_3 ... = \mu_k$$
$$H_1 : \mu_1 \neq \mu_2 \neq \mu_3 ... \neq \mu_k \qquad\qquad H_1 : \mu_1 \neq \mu_2 \neq \mu_3 ... \neq \mu_k$$

The Statistical model is:

UNDER PEER REVIEW

$$y_{ijk} = \underbrace{\mu}_{\text{Overall mean}} + \underbrace{\beta_i}_{\text{row effect}} + \underbrace{\propto_j}_{\text{column effect}} + \underbrace{\ell_k}_{\text{treatment effect}} + \underbrace{(\propto \ell)_{jk}}_{\text{interaction effect}} + \underbrace{e_{ijk}}_{\text{error}} \quad (4)$$

where:

y_{ijk} = *observation on the unit in the* i^{th} *row* j^{th} *column given the* k^{th} *treatment*

μ = *the overall mean*; β_i = *the ith block effect*; $\propto_j$ = *is the effect of the jth* level of factor A_1

ℓ_k = *is the effect of the jth* level of factor A_2;

$(\alpha\ell)_{jk}$ = *is the interaction effect of factor* A_1 *and* A_2 *at the j and k levels respectively*

3. Analysis

3.1 Analysis of the Indeterminacy Component By Neutrosophic Regression

a. The Modified Least Squares Neutrosophic Regression
The Neutrosophic Least-Squares Lines that approximate the neutrosophic bivariate data (x_1, y_1), (x_2, y_2), … , (x_n, y_n) have the same formulae as in classical statistics but the calculations are based on sets rather than just numbers. Hence the classical statisctics are modified to reflect the Neutrosophic Least-Squares statistics as follows;

Table 1.Neutrosophic Modified Classical Regression Sums and Sums of Squares

Neutr. Obs.	x	y	x^2	xy	y^2
i	a	[a,b]	a^2	(a^2,ab)	$[a^2,b^2]$
ii	(a,b)	(c,d)	(a^2, b^2)	(ad,bc)	(c^2,d^2)
iii	k	{m,n}	k^2	{km,kn}	$\{m^2,n^2\}$
Sum	$\sum x$	$\sum y$	$\sum x^2$	$\sum xy$	$\sum y^2$

Where, the sum of X and the sum of Y are computed as;

$$\sum x = (a+a+k, a+b+k) = (2a+k, a+b+k)$$

$$\sum x = (a+c, b+d) + \{m,n\} = \{(a+c, b+d)+m, (a+c, b+d)+n\}$$

$$= \{(a+c+m, b+d+m), (a+c+n, b+d+n)\}$$

$$= \{(\boldsymbol{a+c+m}, b+d+m), (a+c+n, \boldsymbol{b+d+n})\} = (\boldsymbol{a+c+m}, \boldsymbol{b+d+n})$$

The sum of x-squared is computed as

$$\sum x^2 = (a^2+a^2+k^2, a^2+b^2+k^2) = (2a^2+k^2, a^2+b^2+k^2)$$

UNDER PEER REVIEW

The sum of the product of X and Y is computed as

$$\sum xy = \left(a^2+ad, ab+bc\right)+\{km, kn\} = \left\{\left(a^2+ad, ab+bc\right)+km, \left(a^2+ad, ab+bc\right)+kn\right\}$$
$$= \left\{\left(\mathbf{a^2+ad+km}, ab+bc+km\right), \left(a^2+ad+kn, \mathbf{ab+bc+kn}\right)\right\}$$
$$= \left(\mathbf{a^2+ad+km}, \mathbf{ab+bc+kn}\right)$$

The modified Neutrosophic means of X and Y are

The Neutrosophic mean of x_{N_i} is modified as $\bar{x}_N = \frac{\sum x_N}{n_N} = \frac{(2a+k, a+b+k)}{n_N}$

The Neutrosophic mean of y_{N_i} is modified as $\bar{y}_N = \frac{\sum y_N}{n_N} = \frac{(\mathbf{a+c+m}, \mathbf{b+d+n})}{n_N}$

Then the neutrosophic line of best fit becomes;

The y – intercept is modified as $a_N = \bar{y}_N - b_N \bar{x}_N$

The gradient is modified as $b_N = \frac{\sum x_N y_N - \left[(\sum x_N)(\sum y_N)/n_N\right]}{\sum x_N{}^2 - \left[(\sum x_N)^2/n_N\right]}$.

The neutrosophic predicted value of y_N, is given by $\hat{y}_N = a_N + b_N x_N$.

with $\bar{x}_N$ the neutrosophic average of x_N, and $\bar{y}_N$ the neutrosophic average of y_N. One uses the circumflex accent ^ above y in order to emphasize that $\hat{y}$ is a prediction of y.

The modified Neutrosophic Residual Sum of Squares, denoted by NSSResid, given by:

$$NSS\text{Resid} = \sum\left(y_N - \hat{y}_N\right)^2 = \sum y_N{}^2 - a_N \sum y_N - b_N \sum x_N y_N$$

and the modified Neutrosophic Total Sum of Squares, denoted by

$$NSS\text{To} = \sum\left(y_N - \bar{y}_N\right)^2 = \sum y_N{}^2 - \frac{\sum (y_N)^2}{n_N}$$

The modified Neutrosophic Coefficient of Determination is:

$$R_N{}^2 = 1 - \frac{NSS\,\text{Re}\,sid}{NSSTo} = 1 - \frac{\sum y_N{}^2 - a_N \sum y_N - b_N \sum x_N y_N}{\sum y_N{}^2 - \frac{\sum (y_N)^2}{n_N}}$$

and represents the proportion of variation in y_N, when considering a linear relationship between variables x_N and y_N

UNDER PEER REVIEW

We now demonstrate the method of Neutrosophic set using part of our data in the following analysis:

Table 2. Sample Estimations using Neutrosophic Modified Classical Regression Statistic

Neutr. Obs.	x_N	y_N	$x^2{}_N$	xy_N	${y_N}^2$	Neutrosophic Predicted value
1	2	[10,16]	4	[20,32]	[100,256]	(28.6248, -88.283)
2	[5,6]	5	[25,36]	[25,30]	25	(37.3904, -27.683)
3	1	7	1	7	49	(26.6834, -108.483)
4	(7,8)	(12,14)	(49,64)	(84,112)	(144,196)	(40.2732, 12.717)
5	9	{15,20}	81	{135,180}	{225,400}	(42.2146, 53.117)
6	4	6	16	24	36	(32.5076, -47.883)
7	(20,26)	(120,190)	(400,676)	(2400,4940)	(14400,36100)	(75.2184, 275.317)

$$\sum x = (54, 52)$$

$$\sum y = (245, 188)$$

$$\sum x^2 = (852, 602)$$

$$\sum y^2 = (36679, 15362)$$

$$\sum xy = (5235, 2785)$$

$$b_N = \left(\frac{3838}{190}, \frac{895}{461}\right) \approx (20.2, 1.9414)$$

$$\overline{x}_N = \left(\frac{\sum x_N}{n_N}\right) = \frac{(54,52)}{7} = (7.7, 7.43), \qquad \overline{y}_N = \left(\frac{\sum y_N}{n_N}\right) = \frac{(245,188)}{7} = (35, 26.857)$$

Where $a = \overline{y} - b\overline{x}$

We have $a_N = (24.742, -128.683)$ thus, the neutrosophic least-squares line is:

thus, the neutrosophic least-squares line is: $\hat{y}_N = a_N + b_N x_N$

UNDER PEER REVIEW

$$\hat{y}_N = (24.742, -128.683) + (20.2, 1.9414) \cdot x_N$$

Neutrosophic Predicted Values are computed as

$\hat{y}_{N_i}$ = (24.742, −128.683) + (20.2,1.9414) $\cdot$ x_N, for i = 1, 2, …, 7

Hence,

$\hat{y}_{N_1}$ = (28.6248, −88.283)
$\hat{y}_{N_2}$ = (37.3904, −27.683)

$\hat{y}_{N_3}$ = (26.6834, −108.483)
$\hat{y}_{N_4}$ = (40.2732, 12.717)
$\hat{y}_{N_5}$ = (42.2146, 53.117)
$\hat{y}_{N_6}$ = (32.5076, −47.883)
$\hat{y}_{N_7}$ = (75.2184, 275.317)

The modified Neutrosophic residuals are presented as;

$$y_1 - \hat{y}_1 = [10,16] - (28.6248, -88.283) = (-18.6248, 104.283)$$
$$y_2 - \hat{y}_2 = 5 - (37.3904, -27.683) = (-32.3904, 32.683)$$
$$y_3 - \hat{y}_3 = 7 - (26.6834, -108.483) = (-19.6834, 105.483)$$
$$y_4 - \hat{y}_4 = (12,14) - (40.2732, 12.717) = (0.717, -28.2868)$$
$$y_5 - \hat{y}_5 = (15,20) - (42.2146, 53.117) = (-22.2146, -24.4799)$$
$$y_6 - \hat{y}_6 = 6 - (32.5076, -47.883) = (-26.5076, 53.883)$$
$$y_7 - \hat{y}_7 = (120,190) - (75.2184, 275.317) = (-155.317, 114.7816)$$

3.2 Deneutrosophication

We now execute Deneutrosofications, which will convert the neutrosophic data into classical data by taking the midpoint of each set, as we have overcome the indeterminacy problem.This is outlined in the table 3 below:

Table 3. Deneutrosopicated Values

Midpoint of Neutrosophic Predicted Value	**Midpoint of Neutrosophic Residual**
-18.9026	-5.9026
53.5888	-48.5888
7.70605	-0.47255
75.9819	-62.9819
97.62495	-80.12495

UNDER PEER REVIEW

41.2957	-35.2957
227.0198	-72.0198

By substituting the set representations of the coefficients "a" and "b" with their respective midpoints, this approach also converts the neutrosophic least-squares line into a classical least-square line.

The equation $\hat{y}_N = (24.742, -31.723) + (20.2, 2.1931) \cdot x_N$ becomes
$\hat{y}_N = -3.5 + 11.19655 \cdot x_N$

The residuals of **Neutrosophic Sum of Squares (RNSS)**, is represented by **RNSS;**

$$\boldsymbol{RNSS} = \sum (y_N - \hat{y}_N)^2 = \sum y_N{}^2 - a_N \sum y_N - b_N = (42.8291)^2 + ... + (-20.2677)^2 = 5007.996396$$

Similarly, the expression for the **Neutrosophic Total Sum of Squares**, is

$$NTSS = \sum (y_N - \bar{y}_N)^2 = \sum y_N - \frac{(\sum y_N)^2}{n_N} = (36679, 15362) - \frac{(245, 188)^2}{7} = (35957.6939, 14137)$$

We now compute the **Neutrosophic Coefficient of Determination (NCD)**, using:

$$R^2{}_{NCD} = 1 - \frac{\boldsymbol{RNSS}}{NTSS} = 1 - \frac{5007.996396}{(35957.6939, 14137)} = 1 - \left(\frac{5007.996396}{35957.6939}, \frac{5007.996396}{14137} \right) = (0.8607, 0.6458)$$

Thus, the neutrosophic approximate linear relationship between x and y accounts for between 86% and 65% of the sample variation.

After Deneutrosophication, we obtain a perfect linear relationship between variable and response showing that the indeterminacy in the data has been resolved. This ensures the data is now commensurate with classical modelling that will lead to lowered bias in estimates.

UNDER PEER REVIEW

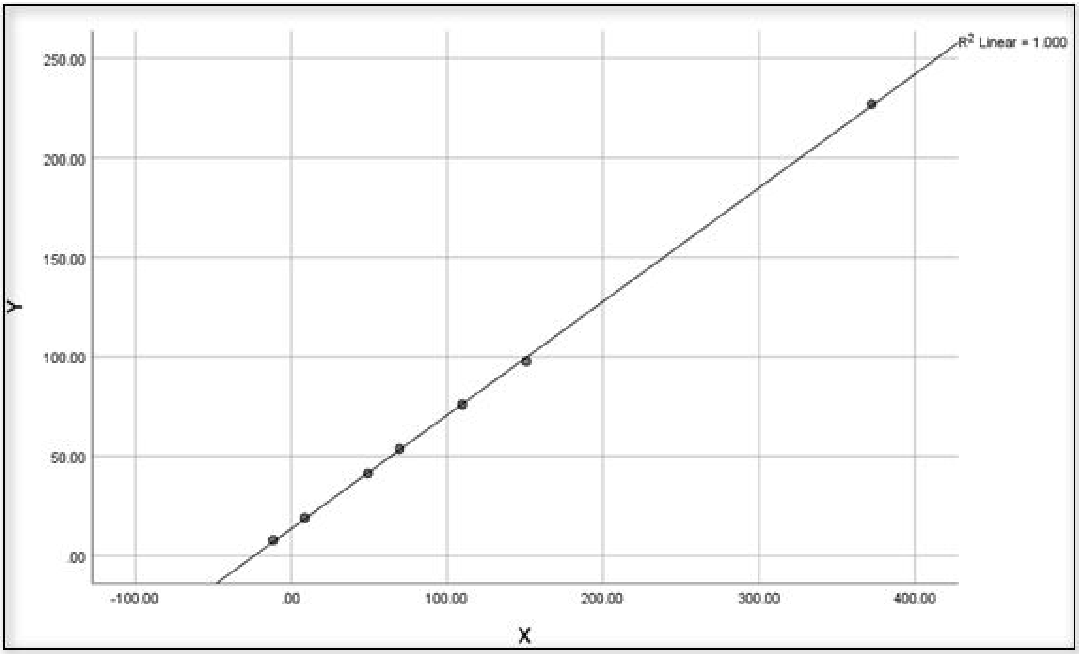

Figure 1. Test of Correlation after Deneutrosophication of variables

3.3 Principal Component Analysis following Deneutrosophication

Table 4. KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	Bartlett's Test of Sphericity		
	Approx. Chi-Square	df	Sig.
0.806	1741.892	406	.000

Using Bartlett's Test of Sphericity, the degree of the variables' interdependence is determined as shown in Table 4. Given that the null hypothesis variables are uncorrelated then, it is assumed that the population matrix is an identity matrix. The Bartlett's Test in the table above yielded a Chi Square value of 1741.892 with a DF of 406 and significance set to 0.00000. The null hypothesis is rejected, thus that the correlation matrix cannot be an identity matrix.

The significance demonstrates that our correlation matrix for our measured variables considerably deviates from an identity matrix, in accordance with the idea that the matrix should be treated as factorable. This confirms that the available data are more than adequate for the Bartlett's sphericity test.

Table 5.**Comparison of Parallel Analysis (Monte Carlo PA Output) and Kaiser'sEigenvalue> 1 Rule**

Factor	**Random order from parallel analysis**	eigenvalue from PCA	**Decision**
1	2.461511	20.265	Accept
2	2.333521	6.506	Accept
3	2.235278	5.913	Accept
4	2.154729	5.226	Accept
5	2.077648	4.752	Accept
6	2.012439	4.473	Accept
7	1.945982	3.757	Accept
8	1.889200	3.476	Accept
9	1.830679	3.363	Accept
10	1.783094	3.093	Accept
11	1.729416	2.851	Accept
12	1.684089	2.708	Accept
13	1.633839	2.586	Accept
14	1.589818	2.328	Accept
15	1.547742	2.270	Accept
16	1.506302	2.099	Accept
17	1.465871	1.856	Accept
18	1.429441	1.795	Accept
19	1.395169	1.719	Accept
20	1.354644	1.657	Accept
21	1.316386	1.551	Accept
22	1.283549	1.467	Accept
23	1.247764	1.420	Accept

UNDER PEER REVIEW

24	1.215083	1.184	Reject
25	1.183768	1.170	Reject

A 25-factor divergence from linearity was caused by the associated Eigen values, as can be seen by looking at the Scree plot result in Figure 2. This test concludes that 25criteria should be considered in the analysis of the data. However, this approach is renowned for its component of subjectivity. Factors must only be preserved if their eigenvalues are greater than 1, according to the Kaiser's eigenvalue >1 criterion. In order to do that, 17 components was suggested by Kaiser's Eigenvalue approach. When these methods are contrasted with the parallel analysis method, a better choice is made.

With 189 observations and 70 components indicator variables, parallel analysis was carried out. The default setting for the percentile Eigen value was set to 95 and 100 correlation matrices were generated. The Eigen values extracted from the data set were compared with the correlation matrices of the parallel analysis that were produced randomly. The factors that met the threshold for retention were those Eigen values (from the data set) above those of the Monte Carlo PA Output. In order to do that, 23 factors were approved and kept as shown in Table 5 and Table 6.

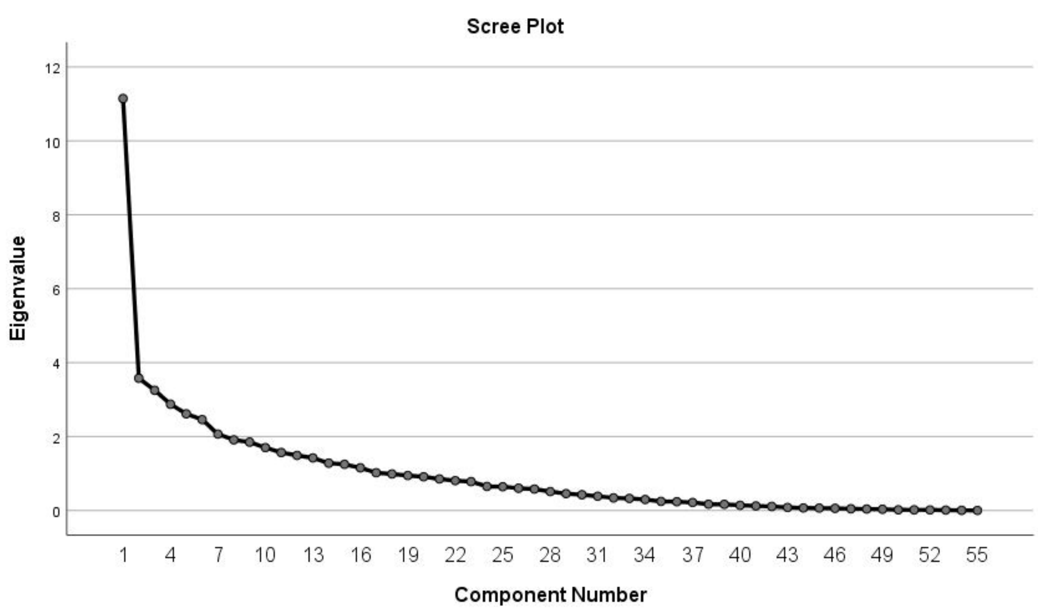

Figure2. Scree plot test

Table 6 . Total Variance Explained after using Multiple Extraction Approaches

Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %

1	11.146	20.265	20.265	11.146	20.265	20.265
2	3.579	6.506	26.771	3.579	6.506	26.771
3	3.252	5.913	32.684	3.252	5.913	32.684
4	2.874	5.226	37.910	2.874	5.226	37.910
5	2.614	4.752	42.662	2.614	4.752	42.662
6	2.460	4.473	47.135	2.460	4.473	47.135
7	2.066	3.757	50.892	2.066	3.757	50.892
8	1.912	3.476	54.367	1.912	3.476	54.367
9	1.850	3.363	57.730	1.850	3.363	57.730
10	1.701	3.093	60.823	1.701	3.093	60.823
11	1.568	2.851	63.673	1.568	2.851	63.673
12	1.489	2.708	66.381	1.489	2.708	66.381
13	1.422	2.586	68.968	1.422	2.586	68.968
14	1.281	2.328	71.296	1.281	2.328	71.296
15	1.249	2.270	73.566	1.249	2.270	73.566
16	1.154	2.099	75.665	1.154	2.099	75.665
17	1.021	1.856	77.521	1.021	1.856	77.521
18	.988	1.795	79.317	.988	1.795	79.317
19	.946	1.719	81.036	.946	1.719	81.036
20	.912	1.657	82.693	.912	1.657	82.693
21	.853	1.551	84.244	.853	1.551	84.244
22	.807	1.467	85.711	.807	1.467	85.711
23	.781	1.420	87.132	.781	1.420	87.132

Multiple extraction techniques, including the Scree Test, Kaiser Criterion, and parallel analysis, were used to prevent over- and under-extraction problems. Kaiser's Eigen value larger than 1 rule and the Scree Test recommended keeping 17 and 25 factors respectively when used in isolation. But the parallel analysis method which makes decision by considering the decision made by the Kaiser's Eigen value recommended preserving 23 elements. Studies in the past that contrasted the three approaches to determining the number of elements to Compared to the Scree test and Kaiser's Eigen value, researchers discovered that the parallel analysis' conclusions were more reliable. PCA with 23 components was therefore imposed.

Rotation was used to create a solution that is more comprehensible, streamlined, and frugal by maximizing high item loadings and minimizing low item loadings. Orthogonal Varimax, the most popular rotation approach, was employed to create uncorrelated factor structures. By increasing the size of the large loadings and decreasing the size of the minor loadings inside each component, it aims to reduce the complexity of the components. About 20.3% of the overall variance is explained by the first component. Additionally, the second factor accounts for 6.5% of the variance in all 21 other factors. The 23 variables together explained around 87.1% of the variance as outlined in Table 6.

Table 7. Rotated Component Matrix

	Component		
	1	2	3
Influence of Private Home Tutor	0.888		
Well organised, disciplined with time management due to school culture that translates into academic life	0.895		
I always want to be in class	0.864		
My teacher pays attention to everyone especially pupils with special needs	0.930		
Prep Time Culture	0.851		
Class competition	0.893		
Engages in Holiday Classes	0.610		
Pupil well managed and supervised at home by parents to focus on academics as a continuity from school	0.881		
Religious denomination		0.943	
My teacher provides me with enough learning activities		0.678	
Enhanced parent-teacher relationship		0.701	
I do perform well in school because my teacher teaches well		0.662	
My teacher treats everybody equally in the class		0.844	
My teacher is always regular in school		0.695	
My teacher gives me prompt feedback for my class exercises		0.784	
Sex of Class Teacher		0.852	
meaningfully communicates progress clearly to parents and learners		0.885	
Identifies and remediates learners' difficulties or misconceptions		0.876	
CondusiveTeaching /Learning treatment			0.922
Daily Quality Supervision of Head teacher and Teacher by superiors			0.956

UNDER PEER REVIEW

Timely Provision of Books and learning materials by Parents/Stakeholders for Pupils			0.940
Type of School			0.804
Concern and parents' support parents towards their Pupils' Academic Output			0.885

The Rotated Component Matrix results from Table 7 indicate strong loadings for variables that account for the notable variance explaining the discrepancy in basic student performance between private and public schools. To manage the pupil-characteristic variations component of this mismatch, eight pupil variables were found. To put it another way, favourably influencing these variables can put students in a position where they can respond appropriately and exhibit high levels of performance. Ten different, but controlled, teacher traits can improve BECE scores if they are handled. But it was also discovered that five elements could influence administrative and logistical factors.

Table8. Variance Explained by Pupil Characteristics

Pupil Factor	**Variance Explained (%)**
My teacher pays attention to everyone especially pupils with special needs	4.211901
Well organised, disciplined with time management due to school culture that translates into academic life	4.053388
Class competition	4.044331
Influence of Private Home Tutor	4.021686
Pupil well managed and supervised at home by parents to focus on academics as a continuity from school	3.989983
I always want to be in class	3.912992
Prep Time Culture	3.854116
Engages in Holiday Classes	2.762645
Total Variance Explained	**30.85104**

UNDER PEER REVIEW

According to Table 8, student characteristics account for approximately 31% of the overall variance causing achievement inequalities between basic students in private and public schools. Given that this factor has the largest variance score of all the pupil factors, teachers must devote enough attention to students with a range of academic abilities, as can be seen from the table. According to the second leading variance score, when students are well-organized, disciplined with time management because of school culture that translates into their academic lives, it puts them in the best possible position to make personal decisions that enhance their academic life. Thirdly, basic school children are substantially more likely to perform better in the BECE examination due to the effect of home tutors.

Table9. Variance Explained by Teacher Characteristics

Teacher Factor	Variance Explained (%)
Religious denomination	4.27077686
meaningfully communicates progress clearly to parents and learners	4.008099174
My teacher provides me with enough learning activities	3.07061157
I do perform well in school because my teacher teaches well	2.99814876
Identifies and remediates learners' difficulties or misconceptions	3.967338843
Sex of Class Teacher	3.858644628
My teacher treats everybody equally in the class	3.822413223
My teacher gives me prompt feedback for my class exercises	3.550677686
Enhanced parent-teacher relationship	3.17477686
My teacher is always regular in school	3.147603306
Total Variance Explained	**35.86909**

According to Table 9, 10 qualities of teachers are to blame for the BECE achievement gaps between students in private and public schools. The sum of these characteristics accounts for around 36% of the variance overall, which is the highest of the three key elements determined to regulate the discrepancies. The teacher's religious affiliation is the main element contributing to the unpredictability of the teacher-factor. It indicates that religious discipline has some bearing on how instructors carry out their duties. The second important teacher-

UNDER PEER REVIEW

related feature is that parents are more likely to take action to raise their children's achievement when teachers meaningfully and clearly communicate progress to parents and students. Thirdly, when teachers Identify and remediate learners' difficulties or misconceptions it can positively affect their performance in BECE.

Table10. Variance Explained by Administrative-Logistic Characteristics

Administrative Logistic Factor	Variance Explained (%)
Concern and parents' support parents towards their Pupils' Academic Output	4.329653
Timely Provision of Books and learning materials by Parents/Stakeholders for Pupils	4.25719
CondusiveTeaching /Learning treatment	4.175669
Daily Quality Supervision of Head teacher and Teacher by superiors	4.008099
Type of School	3.641256
Total Variance Explained	**20.41186777**

Table 10's inferences show that there are five administrative-logistical factors that together account for around 20% of the overall variance that governs achievement inequalities. The most important component of this component is the timely provision of books and educational resources for students by parents and other stakeholders. The second most important component of this component is the favourable teaching and learning treatment or infrastructure. The daily quality oversight of the head teacher and teachers by superiors is the third important component of this component that, if managed properly, can raise performance levels in BECE.

Table 11. Reduced Two-Way Deneutrosophic PCA Model for Between-Subject Effects

Dependent Variable	Type III Sum of Squares	Mean Square	F	P-Value
Teacher Factor	1.956	2.616	12.373	.001
Pupil Factor	2.925	2.249	10.637	.000
Administrative-Logistic Factor	1.396	1.073	5.076	.007
Int. Teacher-Pupil	2.106	0.692	3.275	.013

UNDER PEER REVIEW

Int. Teacher-Admi	1.696	0.556	2.632	.036

Df SS MS

Residuals 185 39.11 0.2114

The Teacher factor, Pupil factor, Administrative-Logistic component, as well as the interaction effects between Teacher-Pupil and Teacher-Administration, were statistically significant ($p < 0.05$) as shown in Table 11.

4. Discussion

Overall, the study concluded that the three elements—student factors, teacher factors, and administrative-logistic factors—are what cause differences in BECE performance between students in public and private basic schools. The overall variance explained by these three factors was around 87%, with the teacher factor accounting for the highest variation—roughly 36%—while the student and administrative-logistic factors each explained 31% and 20% of the total variance respectively. The finding attributed a difference of about 13% to causes of performance disparity to random effects not explained specifically. Some part of this unexplained variation maybe due to some interaction effects between factors which was not addressed by the neutrosophic-PCA method. The ability of a follow-up modeling approach such as generalized linear mixed effect model in a future work will help to explain the interaction component from the 13% percent which will allow us to comprehend the true variation resulting from the random effect, which is not directly determined by the factors taken into account in the data scope of this investigation.

To ascertain whether the independent variables, which included the Teacher factor, Pupil factor, Administrative-Logistic factor, and two interacting effects, had a meaningful impact, a two-way neutrosophic ANOVA test was conducted. According to the outcome, all independent variables were statistically significant ($p < .05$) in relation to students' BECE performance.

5. Conclusions

This research work sought to use Neutrosophic-Principal Component combined technique to identify and explain the factors that are responsible for the prevailing achievement gaps that is yet to be fully elucidated in scientific literature. We identified student factors, teacher factors, and administrative-logistic factors to be responsible for the achievement disparities in BECE performance between students in public and private basic schools; explaining respectively 36%, 31% and 20% of the total 87% variability. A two-way Neutrosophic ANOVA test additional interaction factors such as Teacher-Pupil and Teacher-Administrative Logistic factors are part of the remaining 13% variability that was attributed to random effects by the Neutrosophic PCA approach. We recommend for researchers to use Neutrosophic PCA where the opinions of respondents are to be studied; in order to reduce the biasness that may arise from indeterminacy in responses. Further research may formulate Generalised Linear Mixed Effect model with structural and random parameters as a follow-up

UNDER PEER REVIEW

approach to our Neutrosophic-PCA approach to explore the exact variability attributable to interactive factors and random effects that were unobserved.

Reference

[1] Dangara U.Y and Madudili C.G (2019).Evaluation of the Effect of Learning Treatment on Student’s Academic Performance in Nigeria, https://files.eric.ed.gov/fulltext/ED602097.pdf.

[2] Hussain R.M and Rizvi A.A (2018). Effect of Classroom Learning Treatment on Students' Academic Achievement in Mathematics at Secondary Level, Bulletin of Education and Research Vol. 40, No. 2 pp. 207-218.

[3]Bosson-Amedenu, S., (2017a).Predictive Validity of Mathematics Mock Examination Results of Senior and Junior High School Students’ Performance in WASSCE and BECE in Ghana.Asian Research Journal of Mathematics 3(4): 1-8,2017;Article no. ARJOM. 32328 ISSN: 2456-477X.

[4] Adamu,N.,(2015). Impact of Learning Treatment on the Performance of Students in Social Studies in Junior Secondary Schools in Taraba State, Nigeria. Master Degree Thesis,School of Postgraduate Studies, Ahmadu Bello University, Zaria,Nigeria.

[5] Adediwura, A. A. &Tayo, B. (2007). Perception of teacher's knowledge, attitude and teaching skills as predictor of academic performance in Nigerian secondary schools. Educational Research and Review, 2(7),165-171. Retrieved form http://www.academic journals.org/ERR.291

[6] Bosson-Amedenu, S., (2017b).Effect of Physical and Virtual Manipulative on the Mathematical Achievement of Junior High School Students in the Topic of Transformation in Ghana. . Asian Research Journal of Mathematics.4 (4): 1-14, 2017; Article no. ARJOM.34238, ISSN: 2456-477X.

[7] Bosson-Amedenu, S., (2017c).Archbishop Porter Girls’ Senior High School Students’ Perception of Difficult Concepts in the Senior High School Further Mathematics Curriculum in Ghana. Asian Research Journal of Mathematics.4 (3): 1-14, 2017; Article no. ARJOM.32331, ISSN:2456-477X.

[8] Bosson-Amedenu, S., (2017d).Pre-SHS Students’ Perception of Difficult Concepts in Junior High School Mathematics Curriculum in Ghana.Asian Research Journal of Mathematics 3(2): 1-11, 2017; Article no.ARJOM.32329 ISSN: 2456-477X

[9] Mupa P., Tendeukai I.C (2015) . Factors contributing to ineffective teaching and learning in primary schools: Why are schools in decadence?, Journal of Education and Practice www.iiste.org ISSN 2222-1735 (Paper) ISSN 2222-288X (Online) Vol.6, No.19, 2015

[10] Fernandez, C. (2014). Knowledge base for teaching and pedagogical content knowledge: Some useful models and implications for teachers’ training. Problems of education in the 21st century 60 79-100.

UNDER PEER REVIEW

[11] Delvin, M., Kift, S. & Nelson, K. (2012). Effective teaching and support of students from low socioeconomic status backgrounds: Practical advice for teaching staff. Resources for Australian higher education.Australian Government Office for Learning and Teaching.

[12] Sawchuck, S. (2011). "EWA research brief studies say about teacher effectiveness." London: Sage books.

[13]Bonsu H. D. (2016). A Comparative Analysis of Academic Performance of Public and Private Junior High Schools in The Basic Education Certificate In Sekondi/Takoradi, European Journal of Basic and Applied Sciences, Vol. 3 No. 1, 2016, ISSN 2059-305.

[14] Ampadu H.A (2020). Ministray of Eduacation, Ghana Pre-Tertiary School Performance Inspection Report for the 2019–2020 Academic Year, INFOGRAPHIC_-AGG.-INSP.-REPORT_TERM-1_1920_09APR2020_V1.0.pdf (nasia.gov.gh)

[15] Iddi A. (2016). A comparative assessment of the academic performance among public and private junior high schools in the Tamale Metropolis of Ghana. A thesis submitted to the Department of Planning, Kwame Nkrumah University of Science and Technology, Ghana.

[16] Fiaveh, D. Y., &Lamptey, S. N. L. (2011). Factors prompting pupils' academic performance in privately owned junior high schools in Accra, Ghana. International Journal of Sociology and Anthropology, 3 (8), 280-289.

[17] Bosson-Amedenu S., Osei-Asibey E. and Otoo.H. (2020a). Factor Analysis Model for Diagnostic Assessmentand Instructional Approaches in Ghana's NewCurriculum: The Case of Differentiation, Scaffolding,*Archives of Current Research International,20(4): 1-16, 2020; Article no.ACRI.57232, ISSN: 2454-7077*.

[18] Bosson-Amedenu S., Osei-Asibey E. and Acquah J.(2020b). Multivariate Study of the Use of Assessment Strategies and Cross Cutting Issues by Basic School Teachers: The Case of Ghana's New Curriculum, *Advances in Research, 21(5): 14-27, 2020; Article no.AIR.56593 ISSN: 2348-0394, NLM ID: 101666096.*

[19] Agboola A.A.A., On Refined Neutrosophic Algebraic Structures, in Neutrosophic Sets and Systems, Vol. 9, 2015.

[20] Broumi S., Smarandache F., (2014).Neutrosophic Refined Similarity Measure Based on Cosine Function, in Neutrosophic Sets and Systems, 42-48, Vol. 6, 2014.

[21] Broumi S., Smarandache F., Dhar M., Rough Neutrosophic Set, in Neutrosophic Sets and Systems, Vol. 3, 60-65, 2014.

[22] Broumi S., Smarandache F., (2014). On Neutrosophic Implications, in Neutrosophic Sets and Systems, 9-17, Vol. 2, 2014.

[23] Kandasamy W. B. Vasantha, Smarandache F., (2012). Dual Numbers, Zip Publ., Ohio.

UNDER PEER REVIEW

[24] Mumtaz Ali, Smarandache F., Shabir M., Soft NeutrosophicGroupoids and Their Generalization, in Neutrosophic Sets and Systems, Vol. 6, 61-81, 2014.

[25]Mumtaz Ali, Smarandache F., Shabir M., Vladareanu L., (2014). Generalization of Neutrosophic Rings and Neutrosophic Fields, in Neutrosophic Sets and Systems, Vol. 5, 9-14.

[26] Abdel-Basset M, Atef A, SmarandacheF (2018).A hybrid Neutrosophic multiple criteria group decision making approac.h for project selection.CognSyst Res. https://doi.org/10.1016/j.cogsys.2018.10.023.

[27] Aslam M. (2018). A new sampling plan using neutrosophicprocessloss consideration. Symmetry 10:132

[28] Abdel-Basset M, Mohamed M, Smarandache F (2018).An extension of neutrosophic AHP–SWOT analysis for strategic planning and decision-making. Symmetry 10:116.

[29] Borgonovo E, Morris MD, Plischke E (2018) Functional ANOVA with multiple distributions: implications for the sensitivity analysis of computer experiments. SIAM/ASA J Uncertain Quantif 6:397–427.

[30] QuansahF ,Nugba R.M,(2021). A Trend Analysis of Junior High School Pupils' Performance in the Basic Education Certificate Examination (BECE) in Ghana, International Journal of Elementary Education. Volume 10, Issue 3, September 2021 , pp. 79-86. doi: 10.11648/j.ijeedu.20211003.15

[31] Jolliffe I.T, Cadima J.(2016). Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016 Apr 13;374(2065):20150202. doi: 10.1098/rsta.2015.0202. PMID: 26953178; PMCID: PMC4792409.

[32] Okyerefo, M. P. K., (2011). Factors prompting pupils' academic performance in privately owned Junior High Schools in Accra, Ghana International Journal of Sociology and Anthropology Vol. 3(8), pp. 280-289, August 2011 Available online http://www.academicjournals.org/IJSA ISSN 2006-988x ©2011 Academic Journals.

[33] M.Parimala, Rishwanth& Muthusamy, Karthika&Smarandache, Florentin. (2020). A Review of Fuzzy Soft Topological Spaces, Intuitionistic Fuzzy Soft Topological Spaces and Neutrosophic Soft Topological Spaces. 10. 96-104. 10.5281/zenodo.4008341.

[34] Abd El-Latif, A. M., Characterizations of fuzzy soft pre separation axioms, Journal of New Theory, 7, pp.47-63, 2015.

[35] Abd El-latif, A. M., Fuzzy soft α-connectedness in fuzzy soft topological spaces, Math.Sci. Lett., 5 (1),pp. 85-91,2016.

[36] Abd El-latif, A. M., (2016). Fuzzy soft separation axioms based on fuzzy β-open soft sets, Ann. Fuzzy Math.Inform., 11 (2), pp. 223-239,2016.

UNDER PEER REVIEW