academicJournals

Vol. 12(8), pp. 494-501, 23 April, 2017
DOI: 10.5897/ERR2017.3194
Article Number: 9B5158363946
ISSN 1990-3839

http://www.academicjournals.org/ERR

Educational Research and Reviews

Full Length Research Paper

Clusters and factors associated with complementary basic education in Tanzania mainland

Paul Edwin[1]*, Msengwa S. Amina[2] and Naimani M. Godwin[2]

[1]Department.of Statistics, College of Natural and Mathematical Sciences, University of Dodoma, Tanzania.
[2]Department of Statistics, University of Dar es Salaam, Tanzania.

Received 2 March, 2017; Accepted 11 April, 2017

Complimentary Basic Education in Tanzania (COBET) is a community-based programme initiated in 1999 to provide formal education system opportunity to over aged children or children above school age. The COBET program was analyzed using secondary data collected from 21 regions from 2008 to 2012. Cluster analysis was applied to classify the 21 regions in terms of enrolments by cohort, dropouts, gender, and regional per capital Gross Domestic Product (GDP). The cluster analysis classified 21 regions into four (4) distinct clusters. The first cluster constituted nine regions; second cluster had four regions; the third had seven regions and fourth cluster had only one region. There are variations between those clusters with cluster four (Dar es Salaam region), with minimum dropout and cluster two (Kilimanjaro, Mbeya, Arusha and Iringa regions) with minimum enrolment among all clusters. The study concluded that the number of enrolment by cohort, dropout, gender, regional per capital GDP, and time in years can be used to classify regions into four distinct clusters. However, among the factors associated with the number of enrolment and dropout in the COBET centre; time in years, cohort (age) and clusters were statistically significant at 0.05 level of significance. This study recommends that new plans should be initiated based on these classifications in order to make this programme sustainable and set the new tracking system for follow up of COBET students after completing their studies.

Key words: Cluster analysis, poisson regression, and complimentary basic education (COBET).

INTRODUCTION

The Sustainable Development Goals (SDGs) emphasize estimates using national aggregates as well as variations across different population defined by group and individual characteristics aggregates. According to the fundamental policy, non-formal education is generalized as out of school education, distinguished from formal education which is obtained in schools. However, either type may include, at certain stages, some aspects of the other.However, due to the existence of over aged children who cannot be enrolled in formal primary school, Complimentary Basic Education was established.

According to UNICEF (2006), complementary basic education in Tanzania (COBET) or its Kiswahili equivalent MEMKWA) was a program initiated in 1999 to

*Corresponding author. E-mail: edwinpaultuzo@gmail.com. edwinmsseke@yahoo.com.

provide opportunity for the acquisition of basic education to out of school children aged between 8 to 18years. The research of Johnson et al. (2005) show that, this program was initiated with a special focus on girls, orphans and vulnerable children following a specialized three-year course of study.

In Tanzania as it's in most Sub-Saharan African countries, priority in education is favourable to male children compared to female. UNICEF (2015) report revealed that female children continue to have severe disadvantage of being excluded in education systems despite recent year's positive progress. According to Segumba (2015), dropout is also highest to female compared to male children. The UNICEF report of (2006) defined Orphans and Vulnerable Children (OVC) as those children at risk of missing school, from households with poor food security, suffering from anxiety and depression, and who is at higher risk of exposure to human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV/AIDS).

According to Segumba (2015), there were increased number of out of school children caused by dropout due to sickness, pregnancy, lack of food in the household, forced labour, fear of teachers, excessive corporal punishment, overcrowded classrooms, ineffective teaching, persistence poor performance, long distance from school, lack of food provision in school and poor administration.These lead the demand side for COBET tobecome higher than the availability of centers, which are known to suffer from limited resources and as a result most of them are closed. On the other hand, dropping out of school has emerged as a major threat to achieving Education for All (EFA) goals. This is because it threatens the very fabric of education in terms of inputs/outputs of its structure, organization and provision.

That why COBET assessment was vital. Based on the work of Ngodu (2010) dropout rate was highest in Standard/Grade III-IV in the year 2008/9. An average trend of drop out increased from 3.4% in 2005/6 to 3.7% in year 2008/09. Most of the studies conducted in COBET were focused in piloted districts, and were applied using qualitative techniques and selected small sample with no statistical justification examples Levira (2002) and Michael (2008), therefore this study employs quantitative methods. Based on the UNICEF report (2015) there is relationship between education and regions example in sub-Saharan Africa there is lowest gender parity proportion compared to all other regions.

According to Tanzania development report (2014), Kilimanjaro, Arusha and Dar es Salaam regions are more developed compared to all other regions.Likewise according to the Tanzania Development Report (2014) and UNICEF report (2011) dropout and enrolment varies regionally.

This therefore, calls for this study to assess the number of students enrolled across the regions by triangulating COBET data (that is, dropout, age, year, gender and enrollment) and regional per capital Gross Domestic Product (GDP).The study also identifies the factors associated with COBET enrollment and dropout so that policy makers can be aware and hence take necessary measures to reduce dropout and increase enrolment.

Research objectives

The main objective of this study was to triangulating Tanzania regions based on the number of COBET enrolment by cohort or age, dropout, gender, year, and regional per capital GDP. Moreover, the study examined factors associated with the number of enrolment and dropout in the COBET centers from secondary data collected from Ministry of Education and Vocation Training (MoEVT) and regional per capital GDP collected from National Bureau of Statistics (NBS).

MATERIALS AND METHODS

Given the diversity of the student populations' needs, as well as teachers' availability, a country-wide evaluation was performed to classify regions based on total enrolment by cohort or age, dropout, year, gender and regional per capital GDP from 2008 to 2012. Cluster analysis was used to classify regions based on similar characteristics. The clusters formed were then analyzed to identify the variation between them and factors associated with enrolment and dropout were then performed.

Summary statistics

Table 1 reveals that both enrollment and dropout for both male and female were declining with time being highest in the year 2008 and lowest in 2012. Although according to Johson et al. (2005). COBET were introduced to favour girls compared to boys the results shows that in all years male (boys) enrollment were higher compared to girls.

Cluster analysis

The main objective of conducting cluster analysis is to discover natural groupings of the items or variables. Hierarchical cluster analysis is the major statistical method for finding relatively homogeneous clusters of cases based on measured characteristics. It starts with each case as a separate cluster, i.e. there are as many clusters as cases, and then it combines the clusters sequentially by reducing the number of clusters at each step until only one cluster is left.

According to Antonenko et al. (2012), cluster analysis is an important technique used for examining data in educational research. The prepared by Johnson and Wichern (1992) shows that the data are grouped on the basis of similarities. In its most general

Table 1. Complimentary basic education enrollment and dropout in relation to gender.

Variable		Years				
		2008	2009	2010	2011	2012
Enrollment	Male	61.854	46.729	39.251	44.339	41.241
	Female	50.735	35.637	33.118	37.621	35.626
Dropout	Male	17.637	9.441	5.852	4837	4.608
	Female	13.574	7.292	5.303	4.184	4.247

form, a growing interest in applying data mining to evaluate educational systems makes educational data mining a rising and promising research field this is according to finding done by Romero and Ventura (2007).

Cluster analysis and k-means analysis can be used as data mining techniques. The area of application can be education, different from the usual data mining studies. The research conducted by Erdoğan and Timor (2005) which illuminated on clusteranalysis reveals that use of this technique in education may provide us with more varied and significant findings, and may lead to the increase in the quality of education.

For this study, cluster analysis was used to classify regions with similar characteristics in terms of COBET enrolment by cohort, year, dropout, and per capital regional GDP. Therefore, the clusters formed were further analyzed to determine the variation between clusters as well as if a cluster is one among determinant of COBET dropout and enrolment.

Poisson regression model

In Basic Education Research, one often encounters situations where the outcome variable is numeric, but in the form of counts. The Poisson regression models are often used to model count data. The book prepared by Kutner et al. (2005) shows, Poisson regression models are appropriate for count data because they use probability distributions for the dispersion of the dependent variable scores around the expected value for dependent variables which take on only nonnegative integer values.

Also, Daniel (2008) supports Kutner et al. (2005) idea that, Poisson varieties can take any non-negative integer value. The Poisson-regression model is a nonlinear model for the expected response whereby the expected response is a count. The Poisson distribution is characterized by a parameter λ whereby the probability that variable Y equal to variety y is given by;

$$P(Y=y)=\frac{e^{-\lambda}\lambda^{y}}{y!}\quad y=0,1,2\ldots \text{where } \lambda=E(Y). \tag{1}$$

The Poisson mean in GLMs is commonly modelled using a log-link, $\log(\lambda)=\alpha+\beta x$. For this model, the mean satisfies the exponential relationship:

$$\lambda=\exp(\alpha+\beta x)=\exp(\alpha)\exp(\beta)^{x}.$$

In this study, robust standard deviation had being used as being recommended by Cameron and Trivedi (2009).

FINDINGS AND DISCUSSION

The main attempt of this study was to classify regions based on common characteristics. The hierarchical cluster analysis was performed, by using ward's method to partition the data points into disjoint groups. This implies that, data points belonging to same cluster are similar while data points belonging to different clusters are dissimilar.

According to Dymnicki (2011), ward's method is one of the clustering methods which use centroids to represent clusters by optimizing the squared error function. In this analysis, dendrogram is presented to visualize the clusters formed based on regional per capital GDP, enrolment and dropout rate at regional level.

Figure 1 illustrates four distinct clusters formed. The distance represents variation between clusters. At first, cluster 1 and 2 were merged at approximately Euclidian distance of 4 units then the two clusters, (clusters 1 and 2) merged with cluster 3 at approximately Euclidian distance of 8 units and lastly clusters 1, 2 and 3 merged with cluster 4 at approximately Euclidian distance of 25 units. Thus, there was a great variation between the first threeclusters (clusters 1, 2, and 3) in comparison with cluster4 (Dar es Salaam region). There was a difference of 17 units when the first three clusters werejoined with the fourth while Euclidian distance which merged together the first three clusters (1, 2, and 3) was 8 units.

The distribution of 21 regions was categorized into four clusters (Figures 1 and 2). The first largest cluster consists of 9 regions, namely Manyara, Mara, Lindi, Mtwara, Ruvuma, Morogoro, Rukwa, Tanga and Mwanza. The second cluster consists of 4 regions, namely Kilimanjaro, Mbeya, Arusha, and Iringa. The third cluster which was the second largest consists of 7 regions namely Dodoma, Singida, Kagera, Shinyanga, Kigoma, Pwani and Tabora. The last cluster which was the smallest consists of only Dar es Salaam region.

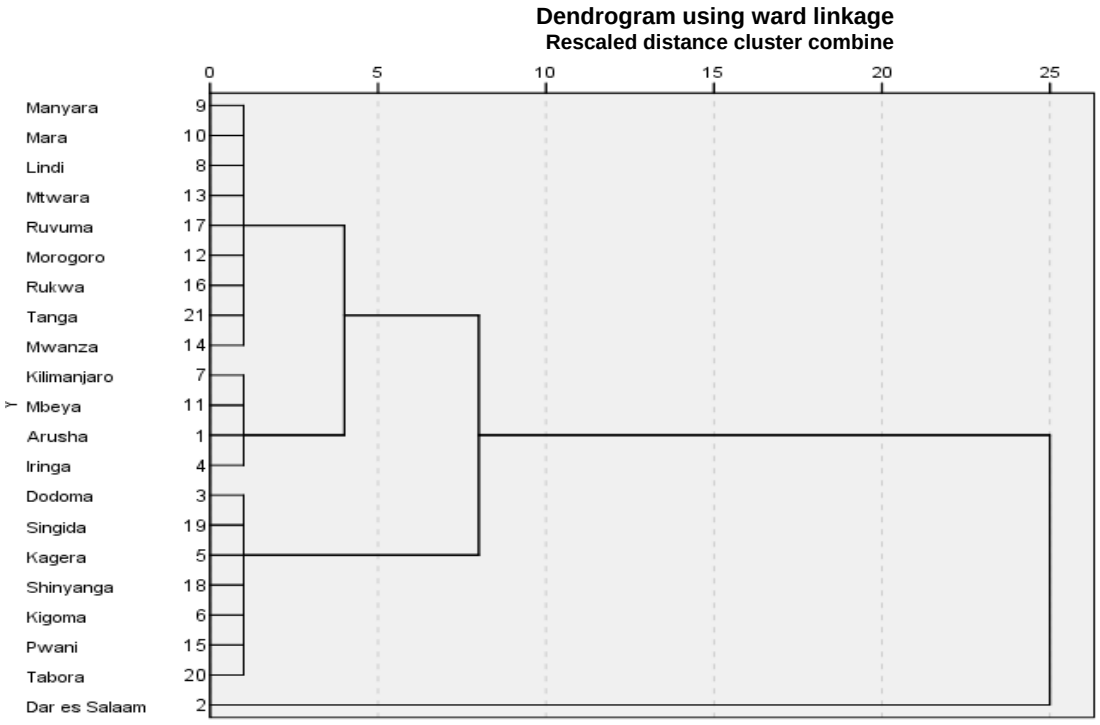

Figure 1. Ward's Linkage Dendrogram showing four clusters of regions.

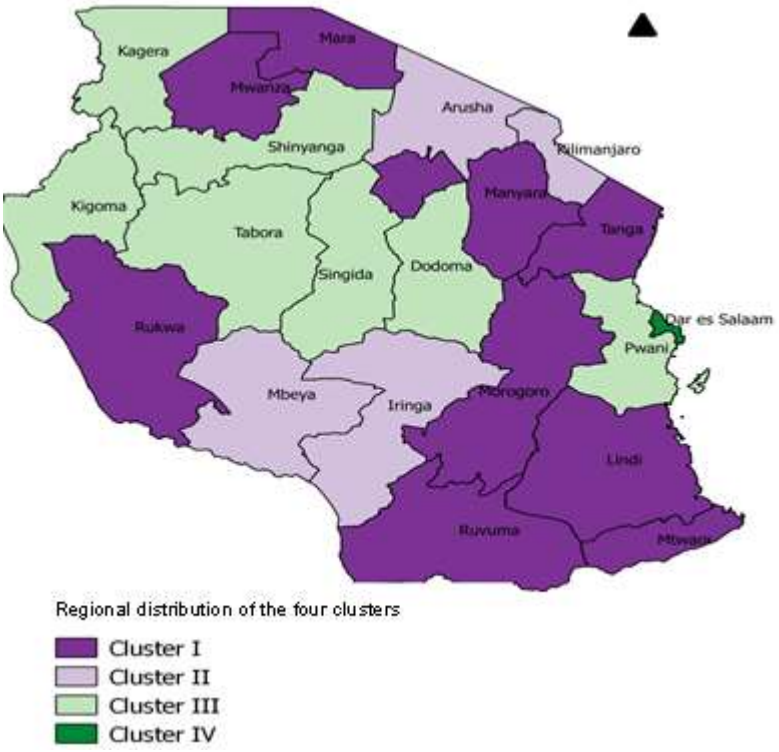

Figure 2. Map showing distribution of four clusters of regions.

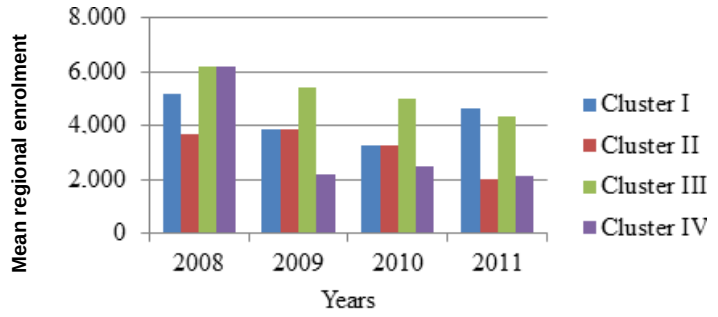

Figure 3. Mean enrolment for the four clusters.

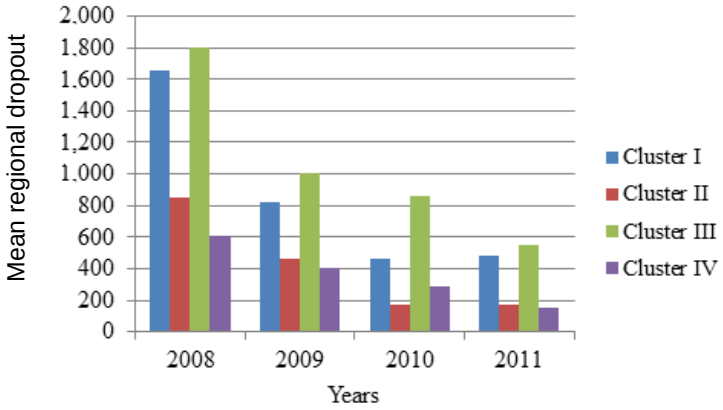

Figure 4. Mean regional dropout for the four clusters.

These classifications put Dar es Salaam region in its own cluster. This is because Dar es Salaam is the region leading in per capital regional GDP and highest literacy rate. Also the first, second, and third clusters show that there is close relationship between regions belonging to the same cluster in terms of regional per capital GDP, literacy rate, females and males literacy gap, and urbanization as also being supported by NBS (2011) Tanzania mainland report. This classification results were also supported by Ministry of Education and Vocational Training report (MoEVT) tests which showed Dar es Salaam to be different in terms of illiteracy rate, female male literacy gap and poverty levels when compared to other regions.

Within–and between-group characteristics

Through Ward's method, the distances between clusters were examined. According to Loureiro, Torgoand Soares (2004) work, cluster analysis is also known as a method for outlier detection, where all the variables from the original data set are used for the description. Clusters summary statistics which shows variations between them are presented in Figure 3 and Figure 4.

Based on the clustering features, the smallest cluster which consists of Dar es Salaam is referred to as an outlier. In explaining the characteristics features related to this outlier, some factors were suggested by UNICEF (2006). These factors include per capital income (deepening poverty), weather condition, food insecurity, migration, lack of enough education facilities such as books, facilitators, classrooms and desks as well as willingness of the guardians to take their children to the Complimentary Basic Education centers before or even after joining the centers. Moreover, urbanization was also suggested as the cause of an outlier by United Republic of Tanzania Vice President's report (2005).

Figure 3 shows that enrolment for clusters 1 and 2 declined from 2008 to 2010 whereas for the year 2011 moved upward. Cluster 3 enrolments were declinin gin all years. Unclear pattern was observed in cluster 4. Figure

Table 2. Parameter estimates and robust standard error of multivariate Poisson model for enrolment.

Variable		Parameter estimate	Robust standard error	P-Value
Gender	Male*	-	-	-
	Female	-0.71688	0.50084	0.152
Clusters	1*	-	-	-
	2	-0.6682	0.1291091	<0.0001
	3	0.37978	0.1011596	<0.0001
	4	-1.2618	0.511284	0.014
Cohort	-	0.0004084	0.0000218	<0.0001
GDP	-	1.33E-06	4.43E-07	0.003
Years	-	-0.163968	0.0367666	<0.0001
_cons	-	335.7398	73.65454	<0.0001

*Indicates reference category.

4 reveals that the dropout rate was declining over years for all four clusters from one year to another.

Multivariate Poisson regression model for enrolment

The preliminary analysis was done to check the relationship between each predictor with the response. All covariates were significant at 5% level $(p < 0.05)$. Therefore all covariates were included in multivariate analysis. Table 1 presents the parameter estimates together with standard error of the final model.

The result shows that clusters, cohort, regional per capital GDP and time in years were significant predictors of enrollment (p<0.05) whereas gender was not (p>0.05). The mean number of COBET enrolment varies from one cluster to another. Controlling the other covariates in the model, the mean number of enrolment for cluster 3 (0.3798) was highercompared to cluster 1. But the mean number of enrolment for cluster 2 (-0.6682) and cluster 4 (-1.2618) were lower than that of cluster 1.

The other significant predictors for the mean number of enrolment were cohort. The model shows that the mean number of enrolment increases with increase in cohort. The result also shows that COBET enrollment increases with increase in regional per Capital GDP. In case of time in years, parameter estimate was -0.164, hence mean COBET enrollment decreases with increase in years. However gender was not statistically significant (p>0.05) the coefficient (β) for females is -0.7169 (negative), and the male students group was taken as control group. This implies that the mean enrolment for males was higher compared to that of females (Table 2).

Multivariate Poisson regression model for dropout

The preliminary univariate analysis was also done to check out the relationship between each predictor with the response. All covariates were significant at 5% level $(p < 0.05)$. All covariates were included in Multivariate analysis. In addition, interactions between gender and cluster and between year and gender were associated to mean number of dropouts. However, some interaction effect between cluster and year $p \geq 0.05$ for individual clusters, were also included in the model because the overall effect was significant and the model converged.

Table 3 presents the parameter estimates together with robust standard error of the final model. The result shows that the mean number of dropouts varies from one cluster to another. Controlling the other covariates in the model, the mean number of dropout for cluster 2(-0.5014), cluster 4 (-1.4856) and were lower as compared to cluster 1, whereas that of cluster 3 (0.2439) were higher compared to cluster 1. The effect of cluster on mean number of dropout depends also on gender. The other significant predictors for mean number of dropout were cohort and years.

The model shows that the mean number of dropout decreases with increase in years while it increases with cohort. However gender was not significantly associated with mean drop out, the coefficient (β) for femaleswere-51.58 (negative) and since males were taken as a control

Table 3. Parameter estimates and standard error of multivariate Poisson model for dropout.

Variable		Parameter estimate	Robust standard error	P-value
Gender	Male*	-	-	-
	Female	-51.5802	154.7265	0.739
Clusters	1*			
	2	-0.5013926	0.175406	0.004
	3	0.2438957	0.1896583	0.198
	4	-1.485565	0.733225	0.043
Male and Cluster 1*	-	-		
Female and Cluster 2	-	-0.4799697	0.1930813	0.013
Female and Cluster 3	-	0.0735148	0.1856015	0.692
Female and Cluster 4	-	-0.2984889	0.3732503	0.424
Cohort	-	0.0004294	0.0000397	<0.0001
GDP	-	9.20E-07	6.88E-07	0.181
Years	-	-0.5348705	0.0727583	<0.0001
Male and Years	-	-0.4653697	0.0724126	<0.0001
Female and Years	-	-0.4397222	0.0853475	<0.0001
_cons	-	939.2599	145.17	<0.0001

*Indicates reference category.

group, this implies that the mean dropout for males is higher compared to that of females. However dropouts for both males and females decrease with increase in years. The results also reveals that mean dropout was not related with regional per capital GDP (p>0.05)

Conclusions

On the basis of the research findings, the following conclusions have been made; that the 21 regions of Tanzania Main land can be grouped into four dissimilar clusters of regions. There are variations between those clusters with cluster four (Dar es Salaam region), having minimum dropout and enrollment compared to all others. Also, this study concluded that, based on the result of Poisson regression model the significant predictors for enrolment and dropout were the same except regional per capital GDP which was significant predictor for enrolment but not dropout. The significant predictors were time in year, cohort (age) and clusters.

RECOMMENDATIONS

This study found out that there are variations between clusters identified. Therefore evaluation for the program as its more than 15 years since its establishment may be vital, in order to identify if objectives of its establishment have been attained. More researches should be done on how COBET can be sustainable programme as there is still dropout in formal school since availability of these schools will make it possible for those dropouts to have another chance for schooling.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

REFERENCES

Antonenko PD, Toy S, Niederhauser DS (2012). Using Cluster Analysis for Data Mining in Educational Technology. Res. Educ. Technol. Res. Dev. 60(3):383-398.

Cameron AC, Trivedi PK (2009). Micro econometrics Using Stata, College Station, TX: Stata Press. Available at: www.stata.com/bookstore/pdf/mus11.pdf

Daniel JW (2008). Poisson Processes and Mixture Distributions. Available at: http://www.actuarialseminars.com/Misc/PPjwd.pdf accessed on 04 April, 2017.

Dymnicki AB, Henry AB (2011).Use of Clustering Methods to Understand More about the Case, Methodological Innovations Online. Method. Innov. Online 6(2):6-26.

Erdogan SZ, Timor M (2005). A Data Mining Application in a Student Database. J. Aeronautics Space Technol. 2(2):53-57.

Johnson H, Lyimo B, Keates D (2005). Inventory on current Education Policies and Practices, Internationally, Nationally and Regionally on School Cost Exemption and good low Cost Educational

Initiatives (Including non-formal education), Moshi, Mkombozi centre for Children Research.
Johnson RA, Wichern DW (1992). Applied Multivariate Statistical Analysis(3rd Edition). New Jersey: Prentice Hall Inc and ISBN: 0130418072. pp. 356-394
Kutner MH, Nachtsheim CJ, Neter J, Li W (2005). Applied linear statistical Models, New York: McGraw.
Levira B (2002). Learner's Evaluation of the Complimentary Basic Education Programme in Kisarawe District, (Master's thesis). retrieved from University of Dar es Salaam Database.Dar es Salaam University Press.
Loureiro A, Torgo L, Soares C (2004). Outlier detection using clustering methods: a data cleaning application.In Proceedings of KDNet Symposium on Knowledge-based Systems for the Public Sector, Bonn, Germany.
Michael L (2008). Implementation of Complimentary Basic Education in Tanzania, (Master's thesis). retrieved from University of Dar es Salaam Database. Dar es Salaam University Press.
NBS (2011). Basic Facts and Figures on Human Settlements, Tanzania Mainland 2009, Ministry of Finance, Dar es Salaam.
Ngodu AS (2010). Alarming Drop-out Rate, A threat of Internal Efficiency in Tanzania Primary Education, United Republic of Tanzania and UNESCO, Ministry of Education and Vocational Training (MOEVT). Available at: http://natcomreport.com/Tanzania/pdf-new/alarming.pdf.
Romero C, Ventura S (2007). Educational data mining: A survey from 1995 to2005. Expert Syst. Appl. 35:135-146.
Segumba SI (2015). Factors Leading To Problems of Drop Out inPrimary School Pupils in TemekeDistrict, Open University of Tanzania. Available at: repository.out.ac.tz/1464/1/SALEHE_IDDI_SEGUMBA.pdf, accessed on o4 April, 2017.
UNICEF (2006).Complementary Basic Education in Tanzania (COBET) Evaluation of the Pilot Project, Dar es Salaam, UNICEF Tanzania, ISBN 9987443079, 9789987443079.
UNICEF (2009). Promoting Quality Education for Orphans and Vulnerable Children, New York, United Nations Children's Fund.
UNICEF (2015). Girls' education and gender equality. Available at: https://www.unicef.org/education/bege_70640.html.
United Republic of Tanzania (2005). National Strategy for Growth and Reduction of Poverty (NSGRP), Vice President's Office June, 2005. Available at: https://www.imf.org/external/pubs/ft/scr/2006/cr06142.pdf
United Republic of Tanzania (2014).Tanzania Human Development Report 2014, Economic Transformation for Human Development, Available at: hdr.undp.org/sites/default/files/thdr2014-main.pdf.