



Original Research Article

Zero-augmented models for exploring the factors affecting the pass rate of 2016 grade 10 learners in Khomas region, Namibia

Rapikama Mumbuu^{1*}, Lilian Pazvakawambwa² and Opeoluwa Oyedele²

¹Department of Mathematics & Statistics, Namibia University of Science and Technology, Windhoek

²Department of Statistics and Population Studies, University of Namibia, Windhoek

*Corresponding author: rmumbuu@nust.na; Tel: 061-2072543

ARTICLE INFO

Received: March 2021
Accepted: April 2022

Keywords:

Hurdle model, zero augmented models, Generalized Linear Models, Zero-Inflated models

ABSTRACT

The poor performance of grade 10 learners has been a big concern over the last few years and in the effort to understand this phenomenon there has been efforts to present models that explain it. This study aimed at exploring the factors which influence Khomas Region grade 10 learners' pass rate using Generalized Linear Models (GLMs). The data used for this study was obtained from the Directorate of National Examination and Assessment for the year 2016, with permission from the Permanent Secretary of the Ministry of Education (DNEA). With the presence of excess zeros in the study data, six GLMs were explored (Poisson, Negative Binomial, Hurdle Poisson, Hurdle Negative Binomial, Zero Inflated Poisson and Zero- Inflated Negative Binomial) to assess their goodness of fit on modelling the zero-inflated DNEA count data. Afterwards, the better performing GLM was used in achieving the study aim. The Zero- Inflated Negative Binomial performed better based on its lowest Akaike Information Criterion (AIC) values among the six fitted GLMs. Results from the fitted Zero- Inflated Negative Binomial model revealed that the age of the learner, school location and the type of school (private/state) had significant differential in the pass rate of grade 10 learners, with p-values < 0.05 in the Zero- Inflated Negative Binomial model. Thus, it is recommended that for densely populated areas, emphasis should be put on building more schools in these areas so that classrooms are not overcrowded per subject. In addition, overaged learners should also be given extra assistance such as extra classes and extra motivation.

1. Introduction

In the era of globalization and technological revolution, education is considered as the first step for every human activity. Education plays a vital role in the development of human capital and is linked to an individual's well-being and opportunities for a better living (Saxton, 2008). It is one of the most powerful instruments known for reducing poverty and for laying the basis for a sustainable economic growth. It ensures the acquisition of knowledge and skills that enable individuals to increase their productivity and improve their quality of life (Battle & Lewis, 2002). This increase in productivity also results in more employment opportunities which enhance the economic growth of a country. In addition, education can be viewed as a process through which the intellectual, moral capacities, proper

conduct, and technical competency of individuals are developed to make them cultural members of their respective societies (Tuan, 2009).

Studies by Miller-Grandvaux & Yoder(2012) on secondary schools education revealed that the main challenges in secondary school education seem to be the academic performance of learners. Generally, the academic performance of learners varies from learner to learner, school to school, location to location and country to country. Anecdotal evidence indicates that the school location, environment, inadequate facilities and infrastructure are some of the factors that account for the differences in academic performance of learners across different subjects. Although this study focused on the grade 10 learners' performance in the Khomas region, the

problem of poor academic performance is a national debacle.

In Namibia, the education system is divided into three stages, namely primary level, secondary level and tertiary level. The primary level, Grade 1 to 7, prepares children for secondary education. In other words, primary education is the basic education provided at primary school level. On the other hand, the secondary level stretches over a period of 5 years from Grade 8 to Grade 12 (Namibia Government, 2001). Learners are presented with a Junior Secondary School Certificate (JSC) after successfully completing Grade 10, and they get a Senior Secondary Certificate at the end of Grade 12 (NSSC).

Since 1993, grade 10 learners in Namibia, regardless of the type of school attended have written the National Junior Secondary Examinations (NJSE) administered by the Directorate of National Examinations and Assessment (DNEA) in Namibia. The NJSE is compulsory for all registered grade 10 learners in Namibia and is used to assess the achievement of learners in a curriculum in order to provide an estimate of the learners' achievement level in the education system. In addition, it is used to make performance comparisons among the 14 regions of Namibia, to further identify schools/regions in need of interventions.

Factors leading to poor performance in secondary schools

Poor academic performance is most commonly determined by combining demographic, socioeconomic and environmental factors such as the parents' educational level, occupational status and income level. It is believed that a low socio-economic status negatively affects the academic achievement of learners in secondary schools (David, 2014). David (2014) further elaborated that learner performance is dependent on the socio-economic background (SEB). They observe that, "High school learners' level of performance is with statistically significant differences, linked to their sex, grade level, school location, school type, learner type and socio-economic background (SEB)." Considering the physical geographical location of most secondary schools in the Sumbawanga District is rural, and the physical infrastructure is poor and limited, the communities might be affected by low socio-economic which influence academic performance (David, 2014).

Several studies have been carried out to identify and analyze the numerous factors that affect academic performance in various centres of learning. Their findings identified factors such as the learners' efforts, literacy level of parents' education, parental

involvement (Jeyness, 2012); self-motivation, the age of learners, learning preferences (Obiero, Mwebi, & Nyang'ara, 2017); class attendance and entry qualifications as factors that have a significant effect on the learners' academic performance in various settings

The influence of age and sex on academic performance has been investigated in a number of studies with widely differing conclusions. Research has also shown that men perform better than women in certain settings while women outperform men in other settings (Sommerville & Singaram, 2018). Scholarly observations show that recent changes in educational policies around the world have led to an increase in the number of mature-age admissions in educational institutions (Sommerville & Singaram, 2018).

The relationship between sex and the academic achievement of learners has been contested. However, a gap between the achievement of boys and girls has been found, with girls showing better performance than boys in certain instances (David, 2014). According to Considine and Zappala (2002), the educational performance in school has also been influenced by the learner's sex. Boys suffer an educational disadvantage relative to girls, especially in terms of performance in literacy. Several explanations for this increasing sex gap which include: biological differences; sex biases (such as reading the fact is seen as not being masculine); teaching, curricula and assessment (for instance less structured approaches to teaching grammar) may have weakened boys (Considine & Zappala, 2002).

According to Jeyness (2012) poor academic performance is most commonly determined by combining demographic, socioeconomic and environmental factors such as the parents' educational level, occupational status and income level. In addition, the low Socio-Economic Status (SES) negatively affects academic achievement of learners in secondary schools (Hansen & Mastekassa, 2013). While a positive relationship between self-motivation and academic performance has been established, the effect of the family's income and parents' level of education on academic performance is far from being unraveled without equivocation. The Socioeconomic status of learners and their families show moderate to strong relationships with academic performance (Jeyness, 2012). However, these relationships are contingent upon a number of factors such that it is nearly impossible to predict academic performance using SES.

A study conducted by Orlu (2013) among six hundred teachers and learners aimed at establishing the environmental influence on the academic performance of secondary school learners. It was found that the school environment has a significant

influence on academic performance, and that its location can affect learners' performance. For example, when a school is situated in a noisy area like an airport or in the heart of a city where activities disrupt the teaching and learning of the learner, one would not expect the learners to do well academically. In fact, noise in any learning environment interferes with the teaching and learning process.

Overcrowding is another factor that affects the teaching and learning environment. Chuma (2012) observes that overcrowding in classrooms makes it difficult for pupils to write. The teacher is also unable to move around the class freely to assist needy pupils and this affects the teaching-learning process. This means that crowded classroom conditions not only make it difficult for learners to concentrate but inevitably limit the amount of time teachers can spend on innovative teaching methods such as cooperative learning and group work (Chuma, 2012). Parents' occupations also influence the learners' achievement in academic work. This is a result of the parents' levels of investment in their children's education which determine their level of purchasing capacity. Learners' academic achievement is negatively correlated with the low level of the parent's income which hinders the individual from gaining access to sources and resources of learning (Jeyness, 2012).

Those studies carried out did not identify the determinants of poor academic performance of the learners, such as, age, sex, location of the school and types of school. Most of the studies done focused on factors influencing the performance of learners in secondary school not specifically in grade 10. Nevertheless, the literature review was used to provide general information on factors influencing learners' performance in general.

Statistical models for count data

The models that are developed to handle count data are normally the Poisson regression and the negative binomial. However due to excess zeros, the hurdle and zero inflated models become very important models in studies on count data. General linear models although very useful, have limitations, such as when the response is restricted to binary and count and when the variance of the response depends on the mean. However, the Generalized linear models (GLM's) extend the general linear framework to address both of the above issues (Zeileis, Kleiber, & Jackman, 2008). GLM involves probability that can be expressed in exponential form. Such distributions are members of the exponential family of distributions written as:

$$f(y; \theta, \phi) = \exp \left[\frac{(y\theta - b(\theta))}{a(\phi)} + c(y, \phi) \right] \quad (1)$$

where $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are some functions, with θ being a function of the location parameter of the distribution (e.g. the mean). This exponential family of distributions include well-known distributions such as the normal distribution, the Poisson distribution and binomial distribution (Zeileis, Kleiber, & Jackman, 2008).

Zero-inflated count models

An alternative approach for modelling zero-inflated data is the zero-inflated count model proposed by Lambert (1992). This model assumes that data are from a mixture of a regular count distribution, such as the Poisson distribution, and a degenerate distribution at zero. The EM algorithm or the Newton-Raphson method can be used to obtain the maximum likelihood estimates. Compared to the hurdle model, this model is more complex to fit, as the model components must be fitted simultaneously.

A zero-inflated negative binomial regression model (With hidden Markov chain)

Wang & Alba (2006) consider a random variable Y of event counts with a piece of data set of k subjects, $\{(y_{ij}, x_{ij}); i = 1, \dots, k, j = 1, \dots, n_i\}$, where y_{ij} is observed event counts for subject i during the j th period, associated with a vector of covariates x_{ij} , and the total sample size $n = \sum_{i=1}^k n_i$. The proposed model assumes that:

- (1) for observed event counts y_{ij} for subject i during period j , there corresponds a partially observed binary random variable, S_{ij} , representing the state of a two-state discrete time Markov chain with $S_{ij} = 1$ when $y_{ij} > 0$ and $S_{ij} = 0$ or 1 when $y_{ij} = 0$;
- (2) The partially observed binary random vector $(S_{i1}, S_{i2}, \dots, S_{in})$ for subject i follows the two-state discrete time Markov chain with transition probabilities defined by

$$\Pr(S_{ij} = 0 | S_{i(j-1)} = 0) = p_{00}, \Pr(S_{ij} = 1 | S_{i(j-1)} = 0) = p_{01} \\ = 1 - p_{00} \quad (2)$$

$$\Pr(S_{ij} = 1 | S_{i(j-1)} = 1) = p_{11}, \Pr(S_{ij} = 0 | S_{i(j-1)} = 1) = p_{10} \\ = 1 - p_{11} \quad (3)$$

where p_{00}, p_{01}, p_{10} and p_{11} are unknown parameters. For observed count y_{ij} , (3) conditional on $S_{ij} = 1$ follows a Negative Binomial distribution with

$$f_1(y_{ij} | x_{ij}, \alpha, \beta, S_{ij} = 1) =$$

$$\frac{\Gamma(y_{ij} + \alpha^{-1}) \left(\frac{\alpha \lambda_{ij}}{1 + \alpha \lambda_{ij}}\right)^{y_{ij}} \left(\frac{1}{1 + \alpha \lambda_{ij}}\right)^{\alpha^{-1}}}{\Gamma(\alpha^{-1})} \quad (4)$$

where $\lambda_{ij} = \exp(\beta'x_{ij})$, $\beta = (\beta_1, \dots, \beta_d)$ is an unknown parameter vector and $\alpha \geq 0$ is the dispersion parameter; conditional on $S_{ij} = 0, y_{ij} = 0$, i.e.

$$f_0 = (y_{ij}|S_{ij} = 0) = \begin{cases} 1, & \text{if } y_{ij} = 0 \\ 0, & \text{if } y_{ij} > 0 \end{cases} \quad (5)$$

The Zero-inflated models are useful when the data contains excess zeros that are both structural and non-structural (sampling zeros) (Hall, 2000). These models have been extensively used in fields such as econometrics and medical fields (Hall, 2000).

This paper adopts the conceptual framework of David (2014) who stated that the academic performance of learners in secondary schools over a given period of time may be influenced by socio-economic factors originating from their families, school environment and the learners themselves. In addition, the article notes that socio-economic, socio cultural and socio-political variables influence learners’ performance directly or indirectly, either by the increase or decrease of the number of learners’ average scores of grades A, B and C for Pass and grade D and F for Fail. The main goal of the study was to investigate factors affecting the pass rate of grade 10 learners across schools in the Khomas region, using the Junior Secondary Certificate (JSC) examination results for the year 2016 obtained from the DNEA. The specific objectives of the study were to explore the various models that can be potentially applied to analyse the relationships between the pass rate and the demographic and socio-economic variables; apply the best model to analyse the relationship between the pass rate and demographic and socio-economic variables; estimate the effects of demographic and socio-economic variables on the pass rate based on the best model and to suggest measures and strategies that can be used to improve the pass rate of grade 10 learners in the Khomas region.

2. Methods

The quantitative cross-sectional study was based on secondary data obtained from the Directorate of

National Examinations and Assessment (DNEA) in the Ministry of Education’s database for the year 2016. The population of this study were the 45 schools in the Khomas region that offer grade 10 education whereby all the 45 schools were used for analysis thus, there was no sampling performed. The dependent variable for this study was the number of subjects passed and the independent variables used were learner’s sex, age, school location (urban, semi-urban, or rural), and type of school (government or private). Descriptive statistics was done to graphically explore the data and provide some basic summary statistics. Six models were explored, namely the Poisson regression, Negative Binomial, Hurdle Poisson, Hurdle Negative Binomial, Zero Inflated Poisson and Zero- Inflated Negative Binomial models. The hurdle Poisson and the hurdle Negative binomial and Zero-Inflated models are used to account for variables with many zeros, particularly in our case to analyze the variable “number of subjects passed. The Poisson regression is susceptible to over dispersion and the Quasi Poisson as well as the Negative Binomial are useful when there is over dispersion, which means that the variance is higher than the mean. The analysis of this study did not explore the Quasi Poisson due to some limitations experienced during the R programming. Models with the lowest AIC’s were more preferable. The data was analysed using both the R programming software version 3.3.1 and the Statistical Packages for Social Sciences (SPSS) software. Several in-built R packages (such as MASS, pscI and AER packages) were used to handle cases involving the hurdle models and the Zero-Inflated models. Data for this research was obtained from the DNEA broad sheets provided to each region that contains distribution of subject’s rankings, results per school, school name etc.

3. Results

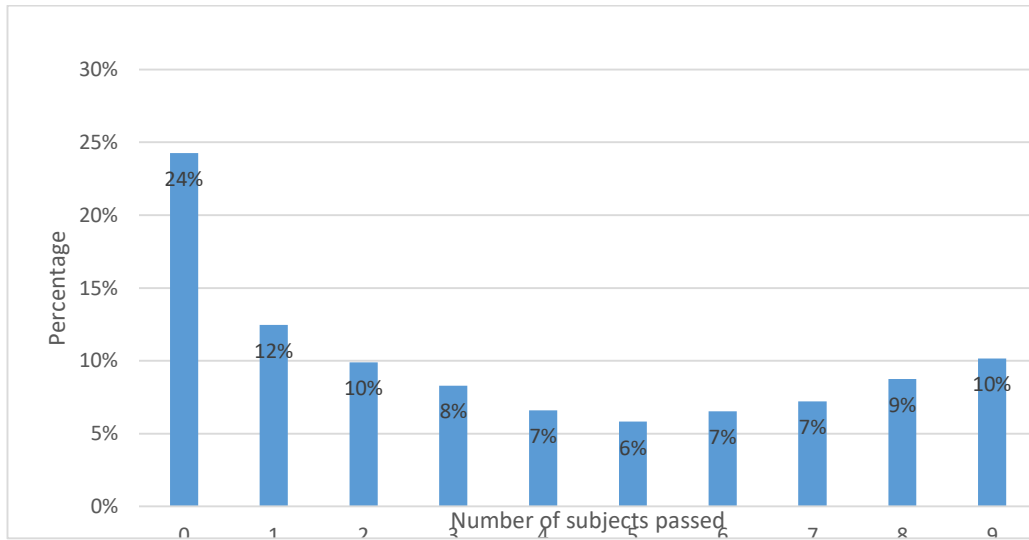
Descriptive Profile of the Pass Rate across Sex

Table 1 shows that the failure rate in 2016 from the Khomas Grade 10 results was higher (55%) than the pass rate (45%). Hence, the study’s intention is to explore factors that affect the pass rate in the Khomas region.

Table 1: Pass rate vs Sex of grade 10 learners

	Sex					
	Female		Male		Total	
	Count	%	Count	%	Count	%
Fail	1527	53.4%	1305	57.0%	2832	55.0%
Pass	1333	46.6%	985	43.0%	2318	45.0%
Total	2860	100.0%	2290	100.0%	5150	100.0%

Figure 1: Bar chart showing the distribution of the number of subjects passed.



The bar chart, Figure 1, reveals that zeros are in excess, compared to the remaining groups. Thus, a hurdle model or zero inflated model was more

appropriate to use in the inferential data analysis section to achieve the objectives of this study.

Table 2: Distribution of the types of school across school location

		Location							
		Low Density		High Density		Rural		Total	
		Count	%	Count	%	Count	%	Count	%
Type	Private	51	2.2%	314	11.8%	69	45.4%	434	8.4%
	State	2296	97.8%	2337	88.2%	83	54.6%	4716	91.6%
	Total	2347	100.0%	2651	100.0%	152	100.0%	5150	100.0%

Table 2 shows that of the total number of schools in the high density location, 434 (8.4%) were privately owned while 4716 (91.6%) were state owned.

Models Comparison

Six different GLMs (Poisson regression, Negative Binomial, Hurdle Poisson, Hurdle Negative Binomial, Zero-Inflated Poisson and Zero-Inflated Negative Binomial) were fitted to analyse the effect of the independent variables on the pass rate of grade 10

learners in the Khomas Region. Table 3 shows the obtained AIC and log-likelihood values for these fitted models. The purpose here was to make a comparison that would yield the best model for the analysis of the variability in the pass rate of grade 10 learners in the Khomas Region. The Poisson regression and Negative Binomial models in Table 4 were fitted using the *glm* package in R, while Hurdle Poisson model, Hurdle Negative Binomial, Zero-Inflated Poisson and Zero-Inflated Negative Binomial models were fitted using the *pscl* and *AER* packages in R.

Table 3: The AIC and log-likelihood values of the 6 fitted GLMs for the DNEA data

	Model	AIC	Log-likelihood
1	Poisson regression	26496	-13242.2
2	Negative Binomial	23771	-11878.35
3	Hurdle Poisson model	23448.58	$-1.71 \times e^4$
4	Hurdle Negative Binomial	22995.1	$-1.148 \times e^4$
5	Zero Inflated Poisson	23448.55	$-1.71 \times e^4$
6	Zero- Inflated Negative Binomial*	22988.84	$-1.148 \times e^4$

*best model

From Table 3 the lowest AIC value (22988.84) and the highest log likelihood occurred when using the Zero-Inflated Negative Binomial (model 5), which means that the Zero- Inflated Negative Binomial model was

the best option for explaining the variability in the grade 10 results of the Khomas region of 2016. Results from the Zero-Inflated Negative Binomial model are presented in Table 4.

Table 4: Zero – Inflated Negative Binomial model coefficients (Binomial with logit link)

Variable	Estimate	Std. Error	P-value	OR(Odd Ratio)
Intercept	-15.122	0.879	<0.001***	0.000
Sex				
Male	-0.015	0.087	0.861	0.985
Female (ref)				
Age	0.675	0.039	<0.001***	1.964
Location				
High density	0.131	0.255	0.608	1.140
Low density	-0.788	0.261	0.0033***	0.455
Rural (ref)				
School type				
State	1.831	0.337	<0.0001***	6.240
Private (ref)				

***significant at 5 % level of significance.

Table 4 above reveals that the odds of having no subject passed among male learners is 0.985 times lower than the odds of having no subject passed among females. However, the p-value of 0.861 (in Table 4) is larger than 0.05, hence one can conclude that the male children do not necessarily fail more

subjects than female learners. A one unit increase in the number of years (being one year older), increases the odds of not passing a subject by 1.964 times. A learner in a low-density area has a reduced chance of 0.455 (45.5%) of failing a subject compared to a learner in a rural school. Being in a state school,

increases the chance of not passing a subject by 6.240 compared to a learner in a private school.

Table 5: Zero-Inflated Negative Binomial model coefficients (with log link):

Variable	Estimate	Standard error	P-value	Expected Rate (ER)
Intercept	5.425	0.219	0.000***	227.011
Sex				
Male	0.022	0.021	0.295	1.022
Female (ref)				
Age	-0.220	0.012	0.000***	0.803
Location				
High density	-0.454	0.065	0.000***	0.635
Low density	-0.208	0.065	0.001***	0.812
Rural (ref)				
School type				
State	0.300	0.051	0.000***	1.350
Private (ref)				

***significant at 5 % level of significance.

Table 5 reveals that a male learner has a higher chance of 1.022 times of passing than a female learner. However, among learners with a positive number of subjects passed, the p-value of 0.295 is larger than 0.05, hence male children do not necessarily pass more subjects than female learners. The value of 0.803 indicates that if a learner's age increases by one year among learners with a positive number of subjects passed, it will lead to the reduction of the number of subjects passed by a learner by 0.803 times. A learner who is attending school in a highly populated area has a reduced chance of 0.635 of passing subjects compared to a learner in a rural school, given that the learner has a positive number of subjects passed. Moreover, a learner attending school in a low-density area has a low chance of 0.812 times of passing subjects than a learner in a rural area, given that the learner has a positive number of subjects passed. The value of 1.350 indicates that the learner at the state school has a higher chance of 1.350 times of passing more subjects than a learner at the private school, given that the learner has a positive number of subjects passed.

4. Discussion

This study concluded that the poor performance of the grade 10 learners is a challenge in the Khomas region, Windhoek. It was found that the age of the learners, location and school had an effect on the performance of learners. The study also established that the sex of the learners had no impact on their overall performance. As such, the study concurred with [Considine & Zappala's \(2012\)](#) research on factors influencing the educational performance of learners from disadvantaged backgrounds. Although they fitted a Binomial Logistic regression to estimate the extent to which individual, family, behavioural and socio-economic factors contribute to learners' achievement, the results from the Wald test revealed that the coefficients were statistically significant for sex, ethnicity, and parental education. However, in terms of the location, the study revealed that schools in low-density areas performed better than rural schools. They concluded that the geographical location did not significantly predict school performance outcomes. This study however yielded the same results in the case of learners' performance

in highly populated areas, where the variable was insignificant in the Zero Inflated Poisson model.

5. Conclusions

The Zero-Inflated Negative Binomial performed better based on its lowest AIC values among the six fitted GLMs. The results revealed that the age of the learner, school location and the type of school (private/state) had significant differential in pass rate with p-values less than 0.05 in the Zero- Inflated Negative Binomial model.

Based on the findings of this study, the following recommendations are made that could be implemented to alleviate the poor performance of the Grade 10 learners on the national examinations, especially in Khomas region.

1. State owned schools should strive to have the same privileges, infrastructures and teaching methodologies as private schools.
2. Emphasis should be put on building more schools in the area so that classrooms are not overcrowded.
3. Rural schools should be given the same attention as urban schools. The current bush allowance should be improved to attract qualified teachers to rural schools.
4. The state schools should bring the teacher-learner ratio of 1:40 in secondary schools on par with the private schools' ratio of 1:25 to give more attention to the slow learners and ease the marking load which consume teachers' time of preparation for lessons that will promote effective teaching and learning.
5. The Ministry should make sure that the teachers employed in state schools are qualified to teach at the relevant levels unlike the current prevailing situation whereby principals appoint and keep unqualified teachers at the expense of qualified ones.
6. Students should be motivated to focus on studies when they are still young.

References

- Battle, J., & Lewis, M. (2002). The increasing significance of class: The relative effects of race and socioeconomic status on academic achievement. *Journal of Poverty*, 21-35.
- Chuma, P. C. (2012). *Challenges Affecting Teaching-Learning in Primary Schools in Kenya. A case study of Central Division Mandera East District*. Nairobi: Moi University.
- Considine, G., & Zappala, G. (2002). Factors influencing the educational performance of students from disadvantage background. *Competing visions*, 91-107.
- David, N. M. (2014). *Determinants of poor academic performance of secondary school in Sumbwanga district, Tanzania*. Sokoine: Sokoine University of Agriculture.
- Hall, D. (2000). Zero-inflated Poisson and binomial regressions with random effects: a case study. *Biometrics*, 1030-1039.
- Hansen, N. M., & Mastekassa, A. (2013). *Social origins and academic performance at university*. Oxford: Oxford University Press.
- Jeyness, W. (2012). A Meta-Analysis of the Efficacy of Different Types of Parental Involvement Programs for Urban Students. *Urban Education*, 706-742.
- Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*.
- Miller-Grandvaux, Y., & Yoder, K. (2012). *A literature review of community schools in Africa*. Washington DC: Bureau for Africa.
- Namibia Government. (2001). *Education Act No. 16*. Windhoek: Ministry of Education.
- Obiero, K., Mwebi, B., & Nyang'ara, N. (2017). Teacher Absenteeism in Public Secondary Schools in Borabu Sub-County. *International Journal of Education and Research*, 130-133.
- Orlu, C. (2013). Environmental Influence on the Academic Performance of Secondary School Students in Port Harcourt Local Government Area of River State. *Journal of Economic and Sustainable Development*, 23.
- Saxton, J. (2008). Investment in Education. *Private and Public Returns*.
- Tuan, L. (2009). Teaching English Discrete Sounds through Minimal Pairs. *Journal of Language Teaching and Research*, 540-561.
- Wang, P., & Alba, J. (2006). *A zero-inflated negative binomial regression model with hidden Markov chain*. Singapore: Nanyang Technological University.
- Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression Models for Count Data in R. *Journal for Statistical Software*.