International Journal of Educational Research and Technology
P-ISSN 0976-4089; E-ISSN 2277-1557
IJERT: Volume 9 [3] September 2018: 54-61
© All Rights Reserved Society of Education, India
Website: www.soeagra.com/ijert.html
ICDS : 5.8 [University of Barcelona, Barcelona]
Global Impact Factor: 0.765 Scientific Journal Impact Factor: 3.72
Journal Impact Factor (JIF) : 1.54, NAAS Rating: 2.96

IJERT

OPEN ACCESS

ORIGINAL ARTICLE

Application of minimum Bayes' factor as an alternative to probability-cum-alpha values for significance testing of hypothesis: Impact of learning activity package instruction on achievement in Physics

[1] Boniface G. Nworgu [2] Fidelis O. Nnadi, [3]Rose C.Anamezie
[1]Department of Science Education, University of Nigeria, Nsukka
[2&3]Department of Science & Computer Education, Enugu State University of Science and Technology, Enugu-Nigeria.
Email: [1]boniface.nworgu@unn.edu.ng, [2]obi.nnadi@esut.edu.ng,.[3]rose.anamezie@esut.edu.ng

ARTICLE HISTORY
Received:
12.03.2018
Revised
17.04.2018
Accepted
19.07.2018

ABSTRACT
This study sought to use minimum Bayes' factor (MBF) to determine the main and interaction effects of method and gender on Physics achievement, instead of the traditional probability-cum-alpha value approach to significance testing. Pretest-posttest, non-equivalent control group design was adopted. The population for the study was four thousand, two hundred and forty six senior secondary three Physics students nested within 208 public secondary schools in Nsukka Education zone of Enugu state. A sample of 163 Senior Secondary two Physics students was used for the study. Data were collected using Electricity Achievement Test. The internal consistency of the final EAT ranged from .66 to .71, whereas the Cronbach's alpha reliability for the whole items was .76. Bayesian Analysis of Covariance was used to analyze the data. Whereas methods showed significantly different means in Physics achievement, in favour of LAP, gender and interaction of gender with method showed no significant difference in their means.). It was recommended that Physics teachers should adopt the learning activity package to enhance increased achievement of their students in addition to its robustness in bridging the gender divide in Physics achievement. Moreover, researchers should adopt Bayesian inference because it is a hybrid inference system.
Keywords: Minimum Bayes' factor, significance testing, Learning Activity Package and Physics achievement.

CITATION OF THIS ARTICLE
Boniface G. Nworgu, Fidelis O. Nnadi, Rose C.Anamezie. Application of minimum Bayes' factor as an alternative to probability-cum-alpha values for significance testing of hypothesis: Impact of learning activity package instruction on achievement in Physics. Inter. J. Edu. Res. Technol. 9 [3] 2018; 54-61.
DOI: 10.15515/ijert.0976 4089.9.3.5461

INTRODUCTION

The Federal Government of Nigeria, FGN through its various statements on statutory documents including that of the Federal Ministry of Science and Technology (FMS & T, [2012]) has emphasized the need for scientific literacy of the citizens of Nigeria, in keeping with global best practices. Scientific literacy, which is conceptualized as the ability of the citizens of any country to understand science and technological concepts, so as to use them in a world of ever-changing culture, has been reported to be low in Nigeria. Conversely, technology and globalization have brought about rapid economic transformation of everyday life in the past decades and consequently, countries, school systems and scholars across the globe have laid emphasis on the fundamental role of Science, technology, engineering and Mathematics education in technological and global economic competition and have therefore called for reforms in education (Sun, Bradley & Akers, 2012; Mbajiorgu, 2014). There is a strong need for reforms in the statistics used in decision making process by scientists and researchers across the globe, irrespective of discipline. The need for the reformation of statistics for testing the significance of hypothesis stems from the inability of the traditional probability (p) value (which was always compared to alpha value) to withstand the test of modernity, as a result of the advent of new computer software. Statisticians including Page and Sakate (2017) reported that minimum Bayes factor (MBF) was a better alternative to the frequentist significance testing approach. The researchers also reported that the MBF was hinged on Bayesian inference, and as such, provided better, credible and reliable results compared to the frequentist-based inference system. Bayesian inference overcomes sampling errors through the use of Markov Chain Monte Carlo (MCMC) algorithm. It gives posterior results and determines its benchmark of significance from the posterior results, unlike the frequentist method of inference where the alpha value of results was set before the experiment. This action by the frequenist researchers increases the subjectivity level of research results. Also in support for a shift from a mechanical inference to an automated or Bayesian inference in data analysis, Wasserstein and Lazar (2016) reported that

Practices that reduce data analysis or scientific inference to mechanical rules (such as $p < .05$) for justifying scientific claims or conclusions can be erroneous beliefs and poor decision making… A conclusion does not immediately become true on one side of the divide and false on the other (p.10).

From the above quotation, it can be deduced that there is the need to bridge the gap of a two-sided mutually inclusive and boarderline decision-making systems by incorporating a common interpretable boundary between them. It also implies that the p-value is a mechanical rule for significance testing. It is mechanical because as a sample-dependent statistic, its value is compared to a hypothetical confidence interval, called alpha value which is a target population dependent parameter, as if there was no sampling. The p-value rule cannot account for a common feature at the boundary of its dichotomous rule. Furthermore, Wasserstein and Lazar also noted that data analysis should not end with the calculation of p-value when other approaches including Bayes factor was appropriate and feasible.

It is good to take a look at the brief history of p-value. The traditional p value was introduced into the realm of significance testing by Fisher in the 1920s (Goodman, 2001). Goodman further noted that there was no definitive tool to estimate the strength of evidence in the hypothesis test, although many scientists and researchers used Fisher's p-value as a tool to actualize statistical inference. In addition, Fisher's p value is a conditional probability of finding a sample with data equal to or more extreme than the actual observed sample data, given that the null hypothesis is true. The implication of the last sentence is that if the null hypothesis is rejected, the Fisher's p-value is no longer tenable to give any information concerning the alternative hypothesis. The p-value was compared to alpha, the probability of a set of future outcomes which by definition lies on the boarder of the tail region of a normal curve. It was from the result of the comparison of p-value and alpha that gave rise to two traditional hypothesis significance testing decisions: reject the null hypothesis, if the estimated p-value is less than alpha and do not reject the null hypothesis, if the estimated p-value is greater than or equal to alpha (Page & Sakate, 2017)). It is worthy of note that although probability and alpha values have some commonalities within the probability space of the sample (sub-population) but outside of the sample space, they are two dissimilar concepts. The p-value requires sampled data to estimate the conditional probability used to describe the characteristics of a target population, when the null hypothesis is true. In contrast, alpha is a parameter which pertains to the target population only. It represents the amount of the disturbances which the experimenter cannot guarantee within the population.

Research evidences including (Nnadi, 2014) abound in Physics Education, for example where the result of p-values in conjunction with alpha value was used to interpret results where the null hypothesis was rejected. What has been the convention in educational research studies is to assign a probability value that is less than or equal to .05 to null hypothesis. Researchers have mistakenly used the alpha value in place of p-value, which only pertains to the sample. Statisticians have insisted that since population parameters remain truly unknown, population-based hypothesis should technically not be accepted. It is either rejected or failed to be rejected. The alpha digit of *.05* is therefore hypothetical and does not evolve from the sample. So, p-value and alpha comparison seems to be faulty. This statement is in line with an earlier observation made by Page and Sakate (2017) that Fisher's p-value is limited to either reject the null hypothesis under a give type I error or fail to reject the null hypothesis under a given type 1 error. Type 1 error is the error committed by the experimenter by rejecting a null hypothesis he/she should accept and vice versa. Page and Sakate also reported that the practice of using the 95% confidence interval as alpha value lacked statistical objectivity and that the level of significance should depend on the context of the investigation. This implies that a sound inference system should be directly dependent on the nature of the observed data.

Further evidences according to Page and Sakate (2017) against the use of the conventional p-value vis-à-vis alpha value in testing the significance of hypothesis have been hinged on the ground that it was: **(a)** asymmetric: p-value does not favour alternative hypothesis. It also overstates the amount of evidence against the null hypothesis. **(b)** sample dependent (p-value fluctuates with varying sample size. Sample size is usually a limitation in Educational studies. **(c)** a dichotomous scale (the yardstick of measurement is .05): A marginal probability value of .049 is rejected as having significantly different parameter values under the null hypothesis framework, while a marginal probability value above the yardstick like .051 is accepted as having no significantly different parameters values. The two probability values are the same up to 1 significant figure. **(d)** a significance that has no extent of evidence. It cannot answer the question, is the evidence small or large enough to be accepted or rejected (failed to be accepted)? (E) determined *a priori* (before the experiment commences).This is a way of increasing the subjectivity level of results.

The alternative to p-value was introduced into the realm of decision making process by Jeffery (1935). Jeffery developed a methodology for testing a scientific theory in a book titled: Theory of probability. The hallmark of the theory was the discovery of a number now called Bayes-factor, which stands for the posterior odds of the null hypothesis, when the prior odds of the null hypothesis is one-half. According to Wagenmakers, Lee, Rouder and Morey (n.d), Bayes factor equals the probability of the observed data under null versus alternative hypothesis. Defazio (2016) also noted that Bayes factor integrates out any nuisance parameter. In other words, the product of prior odds of a hypothesis being true before seeing the data and Bayes-factor gives the posterior/ final odds of null hypothesis being true. The adjective "minimum" that qualifies the phrase "Bayes factor" came into being as a result of averaging the distinct values of the alternative hypothesis (Page & Sakate, 2017). This is because the alternative hypothesis consists of composites (1%, 2%, etc) of the alternative hypothesis differences. Thus, Bayes factor determination involves the selection of a specific composite hypothesis to be compared with the null hypothesis. In a Bayesian framework, objective prior like uniform probability is blended with data to give objective posterior estimates of the parameters of a target population (Nnadi, 2017). Under the minimum Bayes factor (MBF), the identified drawbacks inherent in p-value and alpha significance testing are overcome through the following ways **(a)** MBF takes its value from what is simplified as a ratio of null to alternative hypothesis or vice versa. **(b)** It uses Markov Chain Monte Carlo (MCMC) sample in generating its value. The MCMC is an artificial sample that is randomly generated using software, as a result of recent advances in technology. The MCMC can be large enough to represent the target population and it would appear as if the data were collected from the entire target population. **(c)** MBF uses the fuzzy logic inference to overcome the dichotomous p-value scale, by having a shared region between mutually inclusive outcomes. **(d)** The degree of evidence in a minimum Bayes factor ranges from decisive evidence for alternative hypothesis, H_1 to decisive evidence for null hypothesis, H_0 on an eleven-point scale. (E) MBF is determined *a posteriori* (at the end of the experiment).

What appears to be wrong with the p-value is its comparison with the alpha probability value in decision making. The alpha value of an experiment is set a *priori* and that the experimenter is not sure of its true value at the end of the experiment. Some of the alpha probability features are mutually exclusive to the sample characteristics. The paradigm shift in the decision making process involves the use of sample dependent p-value and prior odds to compute the minimum Bayes factor (without alpha probability value), which can simply be described as a ratio of null to alternative hypothesis and vice versa. This explanation is in line with earlier submission of Azizi, Ali and Ping (2012) that the posterior (Bayes factor) is the product of observation probability and previous information (prior). This technique favors both the null and alternative hypothesis. So, the error in the traditional comparison of the p-value and alpha probability value for a true null hypothesis in terms of judgment for an alternative hypothesis is overcome. So, the minimum Bayes factor does not utilize the hypothetical alpha probability value in its decision, as does the traditional p-value, which uses alpha probability value as its decision benchmark. However, it combines the p-value and prior distribution to estimate its posterior value. The MCMC sample size utilized to estimate minimum Bayes' factor is usually large enough to represent the population set for the study. Thus, the result (minimum Bayes factor) can be applied on the target population. Minimum Bayes' factor is an extension of only the p-value in decision making and it seems to have no business with the hypothetical alpha probability value which used to be compared with p-value in the traditional hypothesis testing.The essence of integrating minimum Bayes' factor in testing the significance of an instructional method vis-à-vis a control is to ensure that the sample parameters are very close to the true and unknown population parameters. Bayesian method is robust in minimizing the standard error of parameter estimation (Nnadi, 2017). Higher standard error in the results of traditional and mechanistic data analyses in the sciences of which Physics is an integral part of undermines the credibility of the research results.

There has been reported poor achievement of students in Physics, especially in external examinations in Nigeria over the past years. This is an indication of pedagogical failure (Nworgu, 2016). The most widely used science pedagogy in Nigerian schools is the lecture. Judging from the students' consistent poor achievement in Physics, as evidenced by the West African Examinations Council (WAEC) chief examiners' (2016) report in Physics, it means there is the need for a paradigm shift to other more innovative pedagogies in Physics including the learning activity package (LAP). LAP was invented by Cardarelli (1972) in Florida. It is student-centered and activity-oriented strategy, with the teacher acting as a facilitator of the learning process (Neboh, 2012). Neboh further reported that LAP strategy consists of the following steps: Topic and sub-topics, rationale, behavioural objectives, pretest, learning activity, quizzes/unit activities and posttest. LAP is an individualized instructional strategy where the students learn at diverse rates. The effectiveness of LAP over lecture has been established in Biology (Neboh, 2012) and Agricultural science (Njoku & Akamobi, n.d).

The influence of gender on Physics achievement has remained controversial. Whereas some researchers including Agbaje and Alake (2011) were of the view that male Physics students achieved higher than their female counterparts in Physics, others including Adeoye (2012) were of the opposite view. The third group including Nworgu, Ugwuanyi and Nworgu (2013) believed that there was equality between male and female students in achievement in Physics. Based on the foregoing, the researchers hypothesized that main and interaction effects of method and gender would not be significant under MBF significance testing.

Purpose of the Study

The purpose of the study was to determine the effect of Learning Activity Package (LAP) on secondary school Physics students' achievement in Electricity using MBF significance testing. Specifically, the study sought to determine:

1. if any significant difference existed between the mean achievement scores of Physics students taught Electricity using LAP and their counterparts taught the same concept using lecture.
2. if any significant difference existed between the mean achievement scores of male and female Physics students taught Electricity using LAP.
3. if any significant interaction existed between methods and gender in Physics achievement.

Hypotheses

The following null hypotheses were tested using minimum Bayes factor,MBF ($100<MBF<.01$):

H_{01}: There is no significant difference between the mean achievement scores of Physics students taught Electricity using LAP and their counterparts taught the same concept using lecture.

H_{02}: There is no significant difference between the mean achievement scores of male and female Physics students taught Electricity using LAP.

H_{03}: There is no significant interaction between methods and gender.

MATERIAL AND METHODS

The design of the study was pretest-posttest non-equivalent control group. The population for the study consisted of four thousand, two hundred and forty six senior secondary three Physics students nested within 208 (Igbo-Etiti,54; Nsukka, 98 & Uzo-Uwani,56) public secondary schools in Nsukka Education zone of Enugu state (Ministry of Education Enugu, 2010). The sample for the study consisted of 163 SS2 Physics students from Nsukka education zone. The sample was made up of one hundred and twenty one students (121) in the experimental group and forty two students (42) in the control group. It also consists of a total of ninety-seven (97) female students and sixty six (66) male students. Two intact classes subjected to treatment condition were sampled from only Boys' and Girls' schools in Nsukka and Igbo Etiti local goverments respectively while one intact class was sampled from Uzo-Uwani as the control. The sampling technique adopted was multi-stage. Stage one involved using simple random sampling, specifically balloting with replacement to sample Igbo-Etiti and Nsukka Local Government Areas in Nsukka education zone of Enugu state. Stage two involved the use of purposive sampling to sample five single-sex schools out of 54 public secondary schools in the area. Stage three involved the use of purposive sampling to sample five intact classes (one in each school) in the sampled schools. In schools that had two or more classes, simple random sampling was used to select one intact class. Finally, in school that had more than one stream of science class, simple random sampling, specifically balloting with replacement was used to sample only one intact class. The instrument used to collect data was electricity achievement test (EAT). It was developed by the researchers. EAT originally consisted of 53 dichotomously scored items measuring three content areas: Ohm's law, capacitive circuit and resonance. EAT was subjected to face validation by Physics Education experts in the Department of Science Education, University of Nigeria, Nsukka in terms of clarity and adequacy of the items. Three items of

EAT failed the face validity test. The 50 surviving items were further subjected to item analysis, using four-parameter (difficult-'a'parameter, discriminating-'b'parameter, pseudo-guessing-'c'parameter and carelessness/inattention-'d'parameter) logistic model. The item analysis was implemented using cognitive diagnosis modeling (CDM) package in r software, version 3.51. Six items were dropped on the basis of having poor psychometric qualities. The differential item functioning (DIF) was also conducted on 44 surviving items of EAT, using Mantel Hanszel's method (difMH), which was implemented in r software, version 3.51 using difR package. Five items were dropped as having significant DIF items. 39 items were also subjected to differential discrimination functioning (DDF), using ddf package, specifically ddf multiple linear regression (ddfMLR) in r. Five items were also dropped on the basis of having p-value less than .05. So, the final item number of EAT was 34. The instrument consisted of three clusters: 14 items measured Ohm's Law, 10 each measured capacitive circuit and resonance. The internal consistency of the final EAT ranged between .66 to .71, whereas the Cronbach's alpha reliability for the whole items was .76.
The research subjects in four schools were randomly assigned to the treatment and one school was the control. The experiment lasted for two weeks. The regular Physics teachers in experimental groups were trained on the use of LAP. Pretest was administered before the experiment began. At the end of the experiment, a reorganized pretest called the posttest was administered to either group. The posttest scores of the students in either group were not homogenous and as such the researchers used the scale function in r software to normalize them. With normalization, the means of the dataset were set to zero and their standard deviations were compared. The reason was to ensure that the scales of the variables in the experiment attained equilibrium before comparison. The data collected were analyzed using mean, standard deviation and Bayesian Analysis of Covariance (BANCOVA), available in open-source Jeffrey's Amazing Statistics Program, JASP version 9.0.0. The initial group differences in the abilities of the research subjects arising from non-randomization were controlled by BANCOVA. The null hypotheses were: (a) accepted when MBF value falls within the range: .10-.33 (*substantial evidence for Ho*), .03-.01 (*strong evidence for Ho*), .01-.03 (very *strong evidence for Ho*) and MBF < .01(decisive *evidence for Ho*), (b) neither accepted nor rejected when MBF value falls within the range: .33-1 (anecdotal *evidence for Ho*), 1 (*no evidence*), 1-3 (*anecdotal evidence for H_1*), and (c) rejected when MBF value falls within the range: 3-10 (*substantial evidence for H_1*), 10-30 (*strong evidence for H_1*), 30-100 (very *strong evidence for H_1*) and MBF>100 (decisive *evidence for H_1*) (Williams, 2016).

RESULTS

Table 1:Descriptive statistics for methods of instruction

Method	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	N
LAP	20.37	4.92	32.41	3.45	121
Lecture	18.55	5.67	28.16	4.43	42

The pretest mean achievement with standard deviation score of the Physics students exposed to LAP was 20.37 and 4.92 respectively, while the their counterparts exposed to lecture had a mean of 18.55 with standard deviation of 5.67 respectively. In the posttest, the students exposed to LAP had a mean of 32.41 with standard deviation of 3.45 respectively. The lecture group had a posttest mean of 28.16 with standard deviation of 4.43 respectively. Furthermore, the pretest mean for the LAP group appeared to be more stable relative to the lecture group which had higher standard deviation. During the posttest, the mean score for LAP was higher and also more stable than the lecture group, whose outliers were more than that of the LAP group.

Table 2: Descriptive statistics due to gender

Gender	Method	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	N
Female	**LAP**	20.53	6.36	24.56	4.02	59
	Lecture	25.46	6.48	32.22	5.84	23
Male	**LAP**	20.21	3.48	40.26	2.88	62
	Lecture	11.64	4.86	24.10	3.02	19

The pretest mean achievement scores of female Physics students in the LAP group was 20.53 with a standard deviation of 6.36 while that of their male counterparts were 20.21 and 3.48 for the mean and standard deviation respectively. The pretest mean for male students appeared to be more stable relative to their female counterparts. In the posttest, the female Physics students in the LAP group had a posttest mean of 24.56 with a standard deviation of 4.02. Similarly, the male Physics students exposed to LAP had a mean of 40.26 with a standard deviation of 2.88. The male Physics students exposed to LAP had higher mean relative to their female counterparts. For the pretest scoring, the female students exposed to lecture had a mean of 25.46 with a standard deviation of 6.48 while their male counterparts had average score of 11.64 with a standard deviation of 4.86. This means that the female students had initial higher ability relative to their male counterparts. In the posttest for the lecture sub-group, the female Physics students had a mean score of 32.22 with a standard deviation of 5.84, while their male counterparts had a relatively lower mean of 24.10 with standard deviation of 3.02.

Hypothesis 1

H_{01}: There is no significant difference between the mean achievement scores of Physics students taught Electricity using LAP and their counterparts taught the same concept using lecture.

Table 3: Bayesian ANCOVA analysis of Physics students' achievement scores.

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.100	0.002	0.020	1.000	
GENDER	0.100	3.728e-4	0.003	0.172	1.136e-5
PRETESTLAP	0.100	5.490e-4	0.005	0.254	0.003
GENDER + PRETESTLAP	0.100	1.026e-4	9.234e-4	0.047	1.085
METHOD	0.100	0.304	3.928	140.310	3.161e-8
GENDER + METHOD	0.100	0.349	4.827	81.208	0.097
PRETESTLAP + METHOD	0.100	0.055	0.520	25.240	1.586
GENDER + PRETESTLAP + METHOD	0.100	0.075	0.725	34.429	1.145
GENDER + METHOD + GENDER ✻ METHOD	0.100	0.180	1.970	82.935	0.073
GENDER + PRETESTLAP + METHOD + GENDER ✻ METHOD	0.100	0.035	0.327	16.211	2.752

From Table 3, the result shows that the minimum Bayes factor (BF_M) value for method was 3.928. This means that the data were 3.928 times more likely to have occurred under the alternative hypothesis than under the null hypothesis. The value of the minimum Bayes factor fell under the category "Substantive evidence for alternative hypothesis". This means that the alternative hypothesis had a substantive evidence for acceptance. Therefore, there was sufficiently substantive evidence from the data to reject the null hypothesis. Therefore, the alternative hypothesis was accepted. So, there was a significant difference in the mean achievement scores of Physics students taught Electricity using Learning Activity Package (LAP) and their counterparts taught the same concept using lecture, in favour of LAP which had a higher mean achievement score of 32.41 with lower standard deviation value of 3.45.

Hypothesis 2

H_{02}: There is no significant difference between the mean achievement scores of male and female Physics students taught Electricity using LAP.

The result in Table 3 shows that the minimum Bayes factor value for gender was .003. The value of the Bayes factor fell under the category "Decisive evidence for null hypothesis". This means that the data were 333.33 (1/.003) times more likely to have occurred under the null hypothesis than under the alternative hypothesis. It also means that the null hypothesis had a decisive evidence for acceptance. There was sufficiently decisive evidence from the data to accept the null hypothesis. So, there was no significant difference in the mean achievement scores of male and female Physics students taught Electricity using Learning Activity Package.

Hypothesis 3

H_{03}: There is no significant interaction of methods and gender.

Table 3 shows the result of minimum Bayes factor for the sum of gender and method plus the interaction of method and gender (GENDER + METHOD + [GENDER ✻ METHOD]) as 1.970, whereas the minimum Bayes factor for the sum of gender and method (GENDER + METHOD) was 4.827. Therefore, the minimum Bayes factor for the interaction of method and gender (GENDER ✻ METHOD) was computed to be -2.857 (1.970-4.827). The value of the interaction fell under the category: "Decisive evidence for null hypothesis". It means that the data were .35 times more likely to have occurred under the null hypothesis than under the alternative hypothesis. This indicates that the alternative hypothesis was slightly favored. So, there was decisively strong evidence from the data to accept the null hypothesis. So, there was no significant interaction of method and gender in Physics achievement.

DISCUSSION

The result of the test of hypotheses in Table 3 shows that there was a significant difference in the mean achievement scores of Physics students taught Electricity using Learning Activity Package (LAP) and their counterparts taught the same concept using lecture, in favour of LAP. This present finding is in tandem with earlier finding of Neboh (2012) who reported that.LAP was more effective relative to lecture in Biology achievement. The result is also in consonance with the finding of Njoku and Akamobi (n.d) who reported a significantly different achievement means between students exposed to LAP and lecture instructions in Agricultural Science. However, the methods of inference adopted by the earlier researchers (Neboh, Njoku & Akamobi) were at variance to the inference system adopted in the present study. Though, similar findings were observed.The result of the null hypothesis in respect of gender showed that there was no significant difference between male and female senior secondary school two Physics students on achievement in Physics. This finding agrees with Nworgu, Ugwuanyi and Nworgu (2013) who reported that male and female Physics students do not differ significantly in their achievement in force and motion concepts test. However, this finding contradicts the positions of some researchers including Agbaje and Alake (2011) who reported that male Physics students achieved better than their female counterparts in Physics and the third group of researchers including Adeoye (2012) who reported that female Physics students achieved better than their male counterparts in Physics.
In terms of interaction effect of method and gender on Physics achievement, the result of this study showed that there was no significant interaction of method and gender on Physics achievement. This means that the combination of method and gender does not significantly produce differing mean achievement scores. This result however contradicts earlier finding of Njoku and Akamobi (n.d), who reported a significantly different interaction for method and gender. In addition, whereas the earlier researchers used traditional analysis of covariance (which was strictly sample dependent and a mechanical technique of estimation of parameter values) the present researchers adopted a cutting edge Bayesian analysis of covariance, which overcomes the identified drawbacks in the traditional ANCOVA by blending the data with prior distribution to give posterior estimates. Also, the yardstick for inference used under the Bayesian perspective (minimum Bayes factor) was a posterior-dependent quantity, unlike alpha value which was set prior to experimentation.

CONCLUSION

The result of the study showed that learning activity package was more effective than lecture in Electricity concept attainment. Gender and the interaction of method and gender were not significantly different in Physics achievement ($100<MBF< .01$). It was recommended that Physics teachers should adopt the learning activity package to enhance increased achievement of their students in addition to bridging the gender divide in Physics achievement. Moreover, researchers should adopt Bayesian inference in data analysis because of its lower estimation error relative to traditional techniques of estimation.

REFERENCES

1. Adeoye, F.A. (2010). Impact of systematic assessment of instruction on secondary school students' Physics achievement at cognitive level of knowledge. *Eurasian Journal of Physics and Chemistry*, 2(1), 44-45.
2. Agbaje, R.O. & Alake, E.M. (2014). Students' variables as predictor of Secondary School Students' academic achievement in Science subjects. *International Journal of Scientific and Research Publications*,4(9),1-9. http://www.ijsrp.org.
3. American Statistical Association, ASA (2016). *American Statistical Association releases statement on statistical significance and p-values. https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf*
4. Amunga, J. K., Amadalo, M. M., & Musera, G. (2011). Disparities in the Physics academic achievement and enrolment in Secondary Schools in Western Province : Implications for strategy renewal. *Problems of Education in the 21st Century,* 31, 18-32. http://www.scientiasocialis.lt/pec/files/pdf/vol31/18-32.Amunga_Vol.31.pdf
5. Cardarelli,S.(1972). The LAP: A feasible vehicle for individualization. *Journal of Educational Technolody*,12(3), 23-29.
6. Defazio, A. (2016). A complete guide to the Bayes factor test.
7. Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, 72, 78-89.www.elsevier.com/locate/jmp
8. Federal Ministry of Science and Technology, FMS & T (2012). *Science, Technology and innovation Policy 2012*. https://www.scienceandtech.gov.ng›docs›ST
9. Goodman, S. H. (2001). Of p-values and Bayes: A modest proposal. *Epidemiology*, 12(3), 295-298.
10. Held, L. (2011). Introducing Bayes Factors.
11. Kass,R.E., & Adrian, R. F. (1993). Bayes factors and model uncertainty. *Technical Report,* No 254. Department of Mathematics, GN-22, University of Washington, USA

 © 2018 Society of Education, India

12. Ly, A., Verhagen, J., & Wagenmakers,E. (n.d). Harold Jeffery's Default Bayes Factor Hypothesis Tests: Explanations, extension and application in Psychology. *Online Manuscript.*
13. Neboh, O. I. (2012). Effect of Learning Activity Package (LAP) on male and female students' achievement in secondary school Biology. *Journal of Science and Computer Education (JOSCED)*, 2(1), 1-14
14. Njoku, C.O., Akamobi, i.(n.d). Effect of learning activity package (LAP) on students' academic achievement in Agricultural Science.
15. Nnadi, F.O. (2017). Application of Bayesian causal modeling estimation technique in analysis of calibrated non-cognitive variables promoting achievement in Physics. *Unpublished Ph.D Thesis*, Department of Science Education, Faculty of Education, University of Nigeria, Nsukka.
16. Nworgu, B .G. (2016). *Averting pedagogical failure in science: Insights from educational measurement and evaluation*. *103rd* Inaugural Lecture of the University of Nigeria, Nsukka delivered on Thursday, February 25.
17. Nworgu, B.G., Ugwuanyi, C.S., & Nworgu, L.N. (2013). School location and gender as conceptual understanding of Force and motion. *International Journal of educational research and technology*, 4(4), 71-76. http://www.soeagra.com/ijert/ijert.htm
18. Page, R.., Satake, E. (2017). Beyond P values and hypothesis testing: Using the minimum Bayes factor to teach statistical inference in undergraduate introductory Statistics course. *Journal of Education and learning*, 6(4), 254-267.doi:10.5539/jel.v6n4p254
19. Rolfe, M. (2010). Bayesian models for longitudinal data. *Unpublished Ph.D Thesis*, Discipline of Mathematical Sciences, Faculty of Science and Technology, Queensland University of Technology, Australia. http://eprints.qut.edu.au/34435/1/Margaret_Rolfe_Thesis.pdf
20. Wagenmakers, E., Lee, M., Rouder, J.& Morey,R (n.d).*Online manuscript*. Another Statistical paradox.
21. Wasserstein,R.L., & Lazar, N.A. (2016). The American Statistical Association's (ASA) statement on p-values: context, process and purpose. doi:10.1080/00031305.2016.1154108
22. West African Examinations Council (2016). *Chief Examiner's Report (*May/June). Lagos: WAEC
23. Williams, M.N.(2016). Bayes factor Null hypothesis tests are still null hypothesis tests. Presentation at modern modeling methods conference, Storr CT, 24-25th May, 2016 at the school of Psychology, Massey University, New Zealand

 © 2018 Society of Education, India