Measuring Change in UTME Use of English through Pre and PostTest Analysis

Sub Theme: Innovations in educational assessment

BY

DibuOjerinde, KunmiPopoola, Patrick Onyeneho and Chidinma Ifewulu

Email: dibu65@yahoo.com,*kunmipopoola@yahoo.compatrickonyeneho@yahoo.co.uk*,

chyfewulu@yahoo.com

Joint Admissions and Matriculation Board (JAMB) Abuja, Nigeria.

Abstract

Pre-test and post-test studies are used primarily for the purpose of comparing groups and/or measuring change resulting from experimental treatments. The rationale for this study is to explore the use of pre-and-posttest analysis method to measure change in UTME Use of English as well as to establish relationships. Experimental treatment was used with the nonrandomized control group pre-test-post-test design. This is because the participants are not randomly assigned to groups. Statistical methods often used include: Analysis of Variance (ANOVA) on the gain scores; Analysis of Covariance (ANCOVA), ANOVA on Residual scores and Repeated Measures ANOVA. This study will explore the use of ANCOV A method which tends to reduce error variances and systematic bias. The purpose of the study is to find out the power of this regression-based method (ANCOVA) in detecting difference in mean change between the Pre and Post-test scores in UTME Use of English. The idea of pre and post testing of students is often accepted as a feasible method of assessing the extent to which an educational intervention has had an impact on student 'learning'. In this analysis, comparison was carried out on groups to establish the differences between pretest (treatment) with respect to their posttest measurements sometimes called change scores. The methodology employed is the one-way experimental design. A random sample of 1,990responsesfrom 2012 Use of English pretest and 2013 posttest on same subject was used. This experiment was carried out using ANCOVA method. The result showed no significant differences in the adjusted means of the pair-wise comparisons across the five groups of candidates’ discipline of study.

Keywords: ANCOVA, systematic bias, error variance, pretest, posttest

Introduction

Pre-test-post-test designs are widely used in behavioral research for the purpose of comparing groups or measuring change that results from experimental treatments. The following statistical methods are traditionally used in comparing groups with pre-test and post-test data. These include: (a) Analysis of variance (ANOVA) on the gain score, (b) Analysis of covariance (ANCOVA), (c) ANOVA on residual scores and (d) Repeated measures ANOVA. In this study, ANCOVA design method which tends to reduce error variances and systematic bias was used.

The analysis of covariance (ANCOVA) can be thought of as an extension of the one-way ANOVA that incorporates a "covariate". As in the one-way ANOVA, the ANCOVA is used to determine whether there are any significant differences between the means of two or more independent groups. However, the ANCOVA has the additional benefit of allowing one to statistically control for a third variable (known as a "confounding variable"), which may negatively affect results. This confounding variable is the "covariate" that one may include in an ANCOVA analysis.

The main purpose of ANCOVA is to adjust the post-test means for differences among groups on the pre-test, because according to Dimiter and Phillip (2003) such difference are likely to occur with intact groups. Experimental design is used in this study because the researcher is interested in determining cause and effect. The causal variable is the independent variable – which is made up of five disciplines of study, and the outcome or effect is the dependent variable which in this case are scores obtained from the UTME Use of English. The purpose of using the pre-test scores as a covariate in ANCOVA with a pre-test-post-test design is to reduce the **error variance and to eliminate systematic bias**. When pre-test scores used are not reliable, the treatment effects can be seriously biased in nonrandomized designs especially if measurement error is present on any other covariate in case ANCOVA uses more than one (the pre-test) covariate. Another problem with the use of ANCOVA lies on the pre-test difference (systematic bias) between groups which can affect the interpretation of post-test differences.

Another analytic strategy used in experiments of this nature involves subjecting the data provided to a repeated measures analysis of variance (ANOVA). However, Huck and McLean (1975) were of the opinion that the statistical results yielded by this type of analysis can easily be misinterpreted, since the score model underlying the analysis is not correct. To overcome this, it is advisable to directly use ANCOVA with the pre-test scores as a covariate. In recently published articles, it was demonstrated that this statistical procedure has led to (i) incorrect statements regarding treatment effects, (ii) completely redundant re-analyses of the same data and (iii) problems with respect to post hoc investigations.

As with all experiments, the starting point must be to define clearly why the experiment is being performed. It is a common knowledge that students with different skills and backgrounds come to study a particular subject, therefore there is need to establish a base measure of the students' knowledge and understanding of a topic in order to be able to quantify the extent of any changes in their learning out comes by the end of a particular period. Ideally, the paper intends to find out not only that the educational intervention has had an impact on the student, hopefully a positive one, but also want to be able to quantify the level of the impact.

Statement of the Problem

A trial-testing of the items developed and moderated by the Test Development Department of JAMB carried out in 2012. The intention was to ascertain the psychometric characteristics, validity, reliability and behavior of the items in terms of the item difficulty, ability to discriminate between high achievers and low performing examinees as well as the level of trait exhibited by the candidates. In calibrating the items, the 3-parameter IRT model was used. In valid items were removed and sent for remediation while valid items were selected and used for the 2013 UTME.

The rationale for this study is to assess the extent the trial-tested items (pretest) has been able to bring about a change in the performance of candidates in the 2013 UTME using ANCOVA method. The issue of assessing the extent the pre-test items has been able to bring about the desired change in the performance of candidates in the 2013 UTME. In this study, attempt was made to determine the impact of "discipline" applied by the candidates on their performances?

Purpose of the study

The idea of pre and post testing of students is often accepted as a feasible method of assessing the extent to which an educational intervention has had an impact on student 'learning'. The Analysis of Covariance is a technique that sits between analysis of variance and regression analysis. Test items pre-tested in 2012 were used in the 2013 UTME on the three modes of examinations conducted namely: Paper-and-Pencil (PPT), Dual-Based Test (DBT) and the Computer-Based Test (CBT). The purpose of this study is to analyze the pre and post-tests of the UTME by comparing the **treatments(pretest)**with respect to the **post-test measurements**. The difference between pre-test and post-test is sometimes called **change scores** or **gain scores**.

Hypothesis

1 **H_0**: There is no significant relationship between pre and post test scores on the UTME Use of English using ANCOVA?

2 **H_0**: There is no significant difference in the adjusted means of the pair-wise comparisons between the five groups of discipline?

Related Literature

When an ANCOVA is conducted, what is looked for is the overall relationship between the outcomes (dependable variable) and the covariate; a fitting a regression line to the entire data set and ignoring which group a person belongs. In fitting this overall model, it is therefore assumed that the overall relationship is true for all groups of participants. For example, if there's a positive relationship between the covariate and the outcome in one group, the assumption is that there is a positive relationship in all other groups too. If, however, the relationship between the outcome (dependable variable) and the covariate differs across the groups, then the overall regression model is inaccurate (it does not represent all the groups). This assumption is very important and it is called the assumption of homogeneity of regression slopes (Norusis, 1990).

A study on pre-test/post-test by Bennett (1983) examined the relative effectiveness of two means of analyzing the pre-test/post-test control group experimental design. Samples were randomly drawn from a standardized normal population and assigned to one of the four cells of the design. A set of experimental differences were induced in the post-test experimental cell. Each case was subjected to a t-test of pre-test and post-test difference scores, and an analysis of covariance (ANCOVA) using the pre-test scores as the covariate. Analysis of 200 cases in each of four sample sizes indicated that: (1) the ANCOVA is a more sensitive test having greater power to detect the induced differences but also being more susceptible to a Type I error; and (2) the findings were essentially the same when cases involving heterogeneity of regression were included and when they were excluded.

In a similar study carried out on Pre-test/Post-test analysis, Dallal (2012) opined that one way around the problem is to compare the groups on differences between post-test and pretest, sometimes called **change scores** or **gain scores**. The test can be carried out in a number of equivalent ways: (a) t-test of the differences; (b) 2-group ANOVA of the differences, (c) repeated measures analysis of variance. However, there is another approach that could be used--**analysis of covariance**, in which the post-test measurement is the response, treatment is the design factor, and the pre-test is a covariate .It is possible for the analysis of covariance to produce a significant treatment effect while the t-test based on differences does not, and vice-versa. The question, then, is which analysis method is appropriate to use.

Data Analysis

Procedure

In calibrating the items to generate item parameters and other item statistics, XCALIBRE 4.3 was used. This involved running the software in five stages: The first stage involves defining input and output data files. *Xcalibre 4.3*has a friendly graphical user interface (GUI) that makes it easy to run the program, even when the researcher is not highly knowledgeable on IRT. The GUI is organized into six tabs: *Files, Input Format, IRT Model, Calibration, Estimation, and Output Options.*

In conducting the One-Way analysis of covariance (ANCOVA) using SPSS, the following steps were implemented.

1. **General Linear Model**, involving u**nivariate analysis with full factorial.**
2. The dependent variable used was Use of English scores (post score)
3. The independent variable used was the discipline applied
4. The Covariate was the 2012 trial-test item scores (pre-score) and
5. The analysis was conducted using the univariate post hoc multiple comparisons for the observed means

Results

SPSS was used for this analysis to produce several outputs as part of the ANCOVA statistical process. Since the sample size is large (N = 1,995), it is suggested that *a prior* alpha level be set at .01 for the Use of English score as the dependent variable.

Result of the One-Way ANCOVA

Table 1.0: ***Descriptive Statistics***

Dependent Variable:Postscore

Discipline	Mean	Std. Deviation	N
Arts/Humanities	44.41	7.600	232
Sciences	46.79	7.564	223
Engineering/Tech	46.47	7.809	283
Social Sciences	43.07	8.493	138
Education	43.71	7.172	119
Total	45.26	7.849	995

Table 1.0 describes the unadjusted means generated by the SPSS as output from the ANCOVA for the Use of English on the five discipline of study applied. Before starting the ANCOVA analysis, the experimenter must test the assumptions underlying the ANCOVA. For this, tests of local independence, equality of error variance and homogeneity-of- regression were carried out. Local independence was carried out using tetrachloric correlation of Vista-Tetrachor software, while tests of equality of error variance and homogeneity-of- regression were derived from the ANCOVA analysis using SPSS.

The test involving the homogeneity-of-regression evaluates the interaction between the pre-score (covariate) and the factor discipline (independent variable) in the prediction of the post-score (dependent variable). A significant interaction between the covariate and the factor suggests that the differences on the dependent variable among groups vary as a function of the covariate. If the interaction is significant – the result from the ANCOVA, are not meaningful and ANCOVA should not be conducted. In this study, the interaction source was labeled as Discipline*Pre-score. The result of the homogeneity-of-regression test suggests that the interaction was not significant. $F(4,985) = .994$. $p = .410$. That is, p (.410) > σ (.01). Based on this finding, the researchers can now proceed with the ANCOVA analysis.

Hypothesis 1: There is no significant relationship between pre and post-test scores in the UTME Use of English scores

Univariate Analysis of Variance

From the analysis conducted, it was observed that the underlying assumption of homogeneity of ariance for the one-way ANCOVA was met. This is evidenced from $F(4,990) = .763$. That is, p .549) > σ (.01). The covariate included in the analysis was to control for the differences on the ndependent variable. While it is not the central focus of the analysis, the results are available as art of the output. The primary purpose of the test of the covariate is that it evaluates the relationship etween the pretest and the posttest variable, controlling for the factor (i.e., for any particular group). he result of this relationship is significant, $F(1, 989) = .007$, $p < .01$ as can be seen in Table 2.1. Vhat this shows is that there is relationship (effect) between the pretest and the posttest variables. f this has not been significant, the question then would be asked on the appropriateness of the election of the covariate. Remember that the covariate must be linearly related to the dependent ariable.

Tale 2.1: ***Tests of Between-Subjects Effects***

Dependent Variable:Postscore

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	2483.895[a]	5	496.779	8.361	.000
Intercept	178685.611	1	178685.611	3007.481	.000
Prescore	430.455	1	430.455	7.245	.007
Discipline	**2030.155**	**4**	**507.539**	**8.542**	**.000**
Error	58760.165	989	59.414		
Total	2099587.000	995			
Corrected Total	61244.060	994			

a. R Squared = .041 (Adjusted R Squared = .036)

The results shown in Table 2.1shows that the group source (highlighted as Discipline) evaluates the null hypothesis that the population adjusted means are equal. The results of the analysis indicate that this hypothesis should be rejected, $F(4, 989) = 8.542$, $p < .01$. The test assesses the differences among the adjusted means for the five groups of discipline, which are reported in the Estimated Marginal Means in Table 2.2as 44.699 (Arts/Humanities), 46.776 (Sciences), 46.382 (Engineering/Tech), 42.975 (Social Sciences) and 43.503 (Education). Usually, in ANCOVA analysis, the results for the covariate are not reported in a Results section.

Table 2.2:*Estimated Marginal Means by Discipline*

Dependent Variable:Postscore

Discipline	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
Arts/Humanities	44.699[a]	.517	43.685	45.714
Sciences	46.776[a]	.516	45.764	47.789
Engineering/Tech	46.382[a]	.459	45.481	47.284
Social Sciences	42.975[a]	.657	41.686	44.264
Education	43.503[a]	.711	42.108	44.898

a. Covariates appearing in the model are evaluated at the following values: Prescore = 37.07.

Calculating the Measure of Association (ω2)

Calculating the measure of association (omega squared) for the ANCOVA is very similar to that for the One-Way ANOVA. We only need to make a few minor adjustments to the formula – to account for the adjusted values of interest. Using the formula:

$$\omega^2 = \frac{SS'_B - (K - 1)MS'_W}{SS'_T + MS'_W} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(1)$$

The **ω** index was computed as $\boldsymbol{\omega^2 = 0.029239354}$ **or 0.03 (≈3%)**

This means that the five levels of discipline (the independent variable) accounted for approximate 3% of the total variance in the students' UOE scores (the dependent variable), controlling for the effect of their **pre-score** (the covariate).At this point in the analysis, we know that the main effect for the independent variable (**discipline**) is significant in controlling for the effect of the covariate (**post-score**).

Using APA guidelines, the report of the statistical information is as follows:

$F(4, 989) = 8.542$, $p < .01$, $\omega^2 = .03$

From the foregoing, it is evident that there is significant relationship between pre and post test scores on the UTME UOE using ANCOVA and so the null hypothesis which stated that there is no significant difference between the pre-score and the post-score in UOE is rejected.

Hypothesis 2: There is no significant difference in the adjusted means of the pair wise comparisons between the five groups of discipline applied.

Conducting pair wise comparisons following a significant main effect

There are two ways of conducting pair wise comparisons. One way is to use the Tukey HSD (equal *n*s) or Tukey/Kramer (unequal *n*s) to calculate the *Q* statistic. Another way is to conduct pair wise comparisons by performing the/matrix command in SPSS (maintaining the appropriate error term) to evaluate whether there are significant differences in the adjusted means between the groups. If the **main effect** is significant, then there is need to conduct a post hoc follow-up. To conduct the pair wise comparisons among the four groups, the following steps were carried out:

1. Click **Analyze**, click **General Linear Model**, and then click **Univariate.**

2. If you have not exited SPSS, the appropriate options should already be chosen, if not, conduct Steps 2 through 5 of conducting the one-way ANCOVA.

3. Click **Paste** to exit the General Linear Model dialog box and go to theSyntax1: SPSS Syntax Editor Window.

4. Remove all but the first three lines:
UNIANOVA
Post-score BY discipline WITH pre-score
/METHOD = SSTYPE(3)

5. Type six separate /matrix commands to compare the adjusted means for Group 1 and Group 2, Group 1 and Group 3, Group 1 and Group 4, Group2 and Group 3, Group 2 and Group 4, and Group 3 and Group 4. The completed syntax is not shown for want of space.

Explaining the /matrix commands for pair wise comparisons

The results of the first pair wise comparison are shown in Table 3.1. This comparison valuates the difference in adjusted means between the first and second groups. For each pair wise comparison (labeled Custom Hypothesis Tests), the Contrast Estimate in the Contrast Results box reflects the difference between the two adjusted means for that contrast. The Test Results box reports the *F* test, degrees of freedom, and the *p*-value (Sig.) for a comparison.

Custom Hypothesis Tests #1

Table 3.1: Contrast Results (K Matrix)[a]

Contrast			Dependent Variable Post-score
L1	Contrast Estimate		-2.077
	Hypothesized Value		0
	Difference (Estimate - Hypothesized)		-2.077
	Std. Error		.731
	Sig.		.005
	95% Confidence Interval for Difference	Lower Bound	-3.512

	Upper Bound	-.642

a. Based on the user-specified contrast coefficients (L') matrix: discipline 1 versus discipline 2

Table 3.2: *Test Results*

Dependent Variable:Postscore

Source	Sum of Squares	Df	Mean Square	F	Sig.
Contrast	479.256	1	479.256	8.066	.005
Error	58760.165	989	59.414		

The difference in adjusted means (labeled Contrast Estimates) is -2.077 (which vary from 44.41 – 46.79, due to rounding). The Sig. (*p*) value is .005; however, it cannot be compared to the *a priori* alpha level (for this example, .01) without making an adjustment to protect against inflating the Type I error rate. While there are different options available – one suggestion is using the Bonferroni adjustment. For this example, since we had eight unique pair wise comparisons – the adjusted alpha (σ') would be .0013 (.01/8 for the eight comparisons). Comparing our obtained *p* value of .005 to the adjusted alpha (.013) we would conclude that there was a significant difference between Groups 1 and 2.

Since we found a significant pair wise difference – we will need to follow this up with determining the Effect Size, measuring the magnitude of this effect using the formula:

$$ES = \frac{X'_i - X'_K}{\sqrt{MS'_W}}$$

where adjusted values are used…

Note: the *MS*'w comes from the ANCOVA summary table (and SPSS provides this in the "Test Results" Table 2.1) and the means will come from the absolute values of the estimated marginal means Table 2.2.

For this example:

$$ES = \frac{2.077}{\sqrt{59.414}} = \frac{2.077}{7.708} = 0.27$$

(Effect Size) ES = .27

Now, let's take a look once more on the comparison of groups 1 and 3.The difference in adjusted means is 1,197. The Sig. (*p*) value is .179, however, remember it cannot be compared to the *a priori* alpha level (for this example, .01) without making an adjustment to protect against inflating the Type I error rate. Again using the Bonferroni adjustment – the adjusted alpha (σ') would be .0013 (.01/8). Comparing our obtained *p* value of .179 to the adjusted alpha (.013) we would conclude that there was a significant difference between Groups 1 and 3.

Interpretation of Results

A one-way analysis of covariance (ANCOVA) was conducted for this study. The independent variable, discipline, included five levels: Arts, Sciences, Engineering/Technology, Social sciences and Education. The dependent variable was the students' UOE scores and the covariate was the students' score on the pre-test. A preliminary analysis evaluating the homogeneity-of-regression (slopes) assumption indicated that the relationship between the covariate and the dependent variable did not differ significantly as a function of the independent variable, $F(4, 989) = .994$, $p = .410$. The ANCOVA was significant, $F(4, 989) = 8.542$, $p < .001$ (See Table 3.1). However, only 3% ($\omega^2 = .03$) of the total variance in UOE scores was accounted for by the five levels of discipline of study controlling for the effect of the students' pretest scores.

Follow-up tests were conducted to evaluate pair wise differences among the adjusted means for independent variable -discipline. The Bonferroni procedure was used to control for Type I error across the eight pair wise comparisons ($\dot{\alpha} = .01/8 = .0013$). The results showed that students' whose discipline of study was Sciences (*M* = 46.776) had significantly higher UOE scores, controlling for the effect of their post-score test, than students whose discipline was Engineering/Technology (*M* = 46.382). In the same way, students in the Social Science discipline has the least (*M* = 42.975). The effects sizes for these significant adjusted mean differences were .27, .22, .493, .37, and .22, respectively. (See Table 4).

Table 4:*Pairwise Comparisons and Effect Sizes of UOE by Discipline*

			Adjusted Mean Differences ($X'i - X'k$) (Effect Sizes are indicated in parentheses)				
Group	Mean	Adjusted Mean	1.	2.	3.	4.	5.
1. Arts/Humanities	44.41	44.70	---	--	---		
2. Sciences	46.79	46.78	2.08	--	---		
3. Engineering/Tech	46.47	46.38	1.68	-0.4	---		
4. Social Sciences	43.03	42.98	-1.72	-3.8	-3.4		
5. Education	45.26	43.50	-1.20	-3.3	0.52		
			(.27)	(.22)	(0.49)	(.37)	

* $p < .013$

Note: The Bonferroni adjustment is considered to be a rather conservative adjustment – and as such should be used with that in mind. An alternative (somewhat less conservative method) is to use the Holm's Sequential Bonferroni adjustment method to control for Type I error or to hand calculate the pair wise comparisons using the Tukey or Tukey-Kramer method.

Discussion of findings

The result of the analysis of variance (ANCOVA) using experimental design showed that the scores obtained during the pretest (covariate) and that of the post-test (treatment) in the UTME Use of English had positive relationship. The primary purpose of the test of the covariate is that it evaluates the relationship between the covariate and the dependent variable (treatment), controlling for the factor (i.e., for any particular group).

Conclusion

The paper reaffirmed that the use of ANCOVA which allows the researchers to make comparison between groups that are not comparable with respect to some important variables such as the covariate (pre-test) is a welcome development in the sense that the method removes variations due to the covariate and therefore provides a more precise analysis. Thus the results from this study showed that there is significant relationship between the pre and post test scores in the UTME UOE with the five levels of discipline applied for by the candidates accounting for approximately 3% of the total variance in the students' UOE scores (the dependent variable) as well as controlling for the effect of their pre-score (the covariant).This means that the main effect for the independent variable (discipline) was significant in controlling for the effect of the covariate (post-score).

Recommendation

In measuring change, the Analysis of covariance (ANCOVA) method is recommended to be used in in pre and post-test studies because of the advantage it provides in reducing error variances and systematic bias which may occur in the process of conducting experiments of this nature.

Future studies should examine and compare the repeated measures analysis with the power of this regression-based method (ANCOVA) in detecting differences in mean change between the Pre and Post-test scores of UTME. The idea of pre and post-test studies of students' performances is often accepted as a feasible method of assessing the extent to which an educational intervention has had an impact on student 'learning. It is recommended that other UTME subjects should also be subjected to this form of experiment in other to find out the relationships between the covariate and the post-test treatments.

References

Bennett, R, P. (1983). Comparison of Two Analysis Techniques for the Pre-Test/Post-Test Control Group Experimental Design. ERIC Number: ED226069.

Dimitrov, D, M. and Rumrill, D. (2003). Pretest-posttest designs and measurement of change 507 White Hall, College of Education, Kent State University, Kent, OH 44242-0001, USA ddimitro@kent.edu*Work 20 (2003) 159–165 159*IOS Press.

Gerard, E.D.(2012). The analysis of Pre-test/ Post test experiments, www.jerrydallal.com LHSP/prepost.htm.

Huck, S. W. and McLean, R. A. Using a repeated measures ANOVA to analyze data from a pretest-posttest design: A potentially confusing task, *Psychological Bulletin* **82** (1975),511–518.

Norusis, M. J. (1990). SPSS/PC + Statstics 4.0 Base Manual. 444 N, Michigan Avenue, Chicago, IL, 60611. ISBN 0-923967-11-7.