

Appendix I

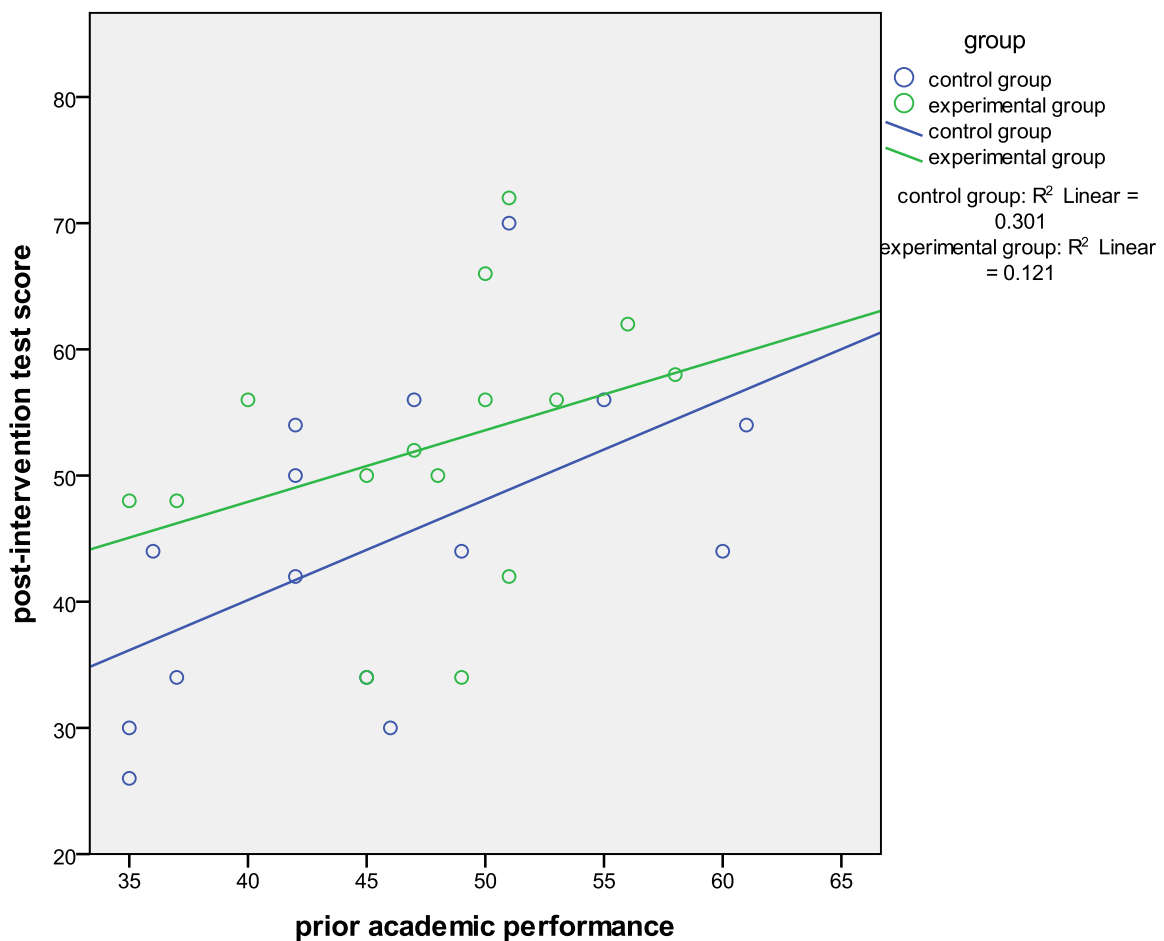
ANCOVA testing of assumptions

The initial methodological checks involving the levels of measurements were all met. No relationship between the observations within or between groups was identified. A linear relationship between prior academic performance and post-intervention test scores, measuring academic achievement, was established, as confirmed by an ANOVA, $F(1,28) = 0.600$, $p = 0.445$ and assessed by visual inspection of a scatterplot.

ANOVA

prior academic performance

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	34.133	1	34.133	.600	.445
Within Groups	1593.067	28	56.895		
Total	1627.200	29			



This implies that prior academic performance was roughly equal across and independent of the 2 groups, thus suitable to be included as a covariate. Homogeneity of regression slopes was met as the interaction term (group*prior_academic_performance) was not statistically significant, $F(1,26) = 0.175$ $p = 0.679$.

Tests of Between-Subjects Effects

Dependent Variable: post-intervention test score

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1273.640 ^a	3	424.547	3.863	.021	.308
Intercept	185.536	1	185.536	1.688	.205	.061
group	47.036	1	47.036	.428	.519	.016
prior_academic_performance	688.559	1	688.559	6.265	.019	.194
group * prior_academic_performance	19.214	1	19.214	.175	.679	.007
Error	2857.560	26	109.906			
Total	74408.000	30				
Corrected Total	4131.200	29				

a. R Squared = .308 (Adjusted R Squared = .228)

This implies that the relationship between the post-intervention test score and prior academic performance was the same for each group. Standardised residuals for the 2 groups and for the overall model were normally distributed, as verified due to the small sample size (<50) by the Shapiro-Wilk test ($p > 0.05$) and visual inspection of the Normal Q-Q plots.

Tests of Normality

group	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual for control group	.104	15	.200*	.962	15	.729
post_intervention_test_score experimental group	.201	15	.106	.943	15	.417

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Tests of Normality

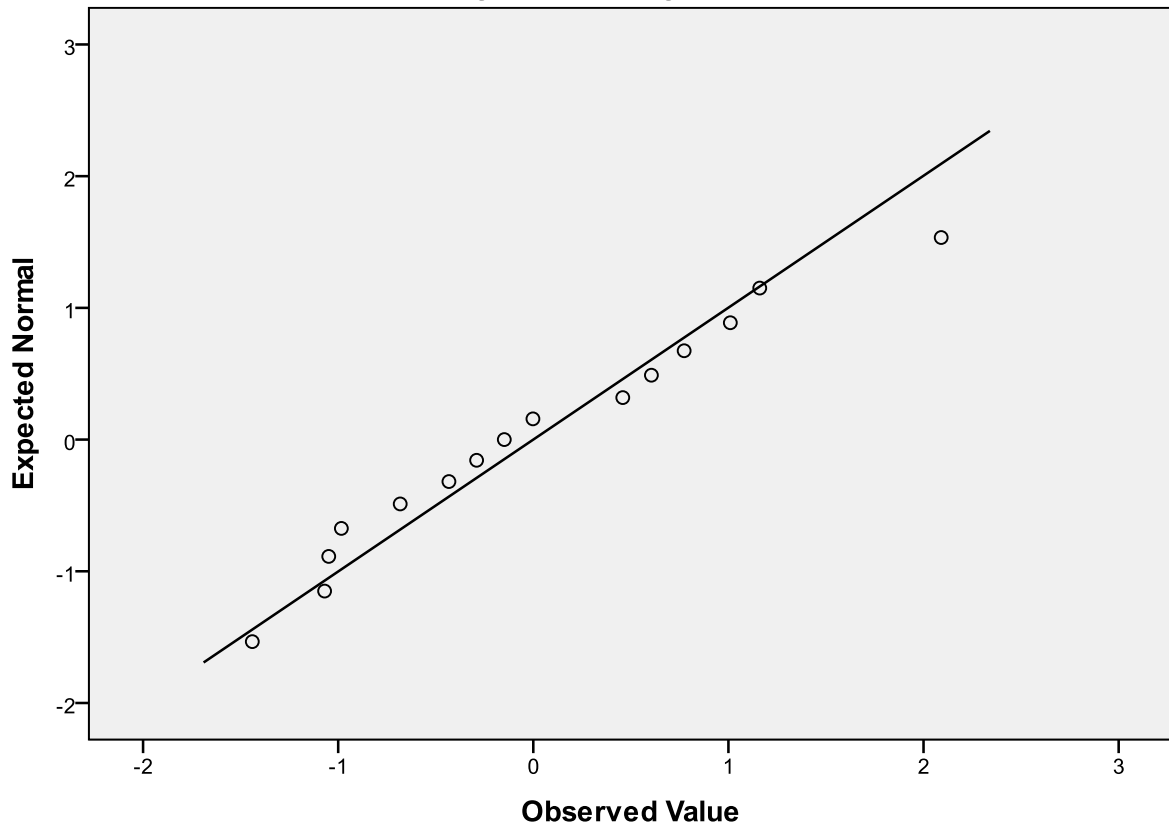
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Standardized Residual for post_intervention_test_score	.082	30	.200*	.986	30	.957

a. Lilliefors Significance Correction

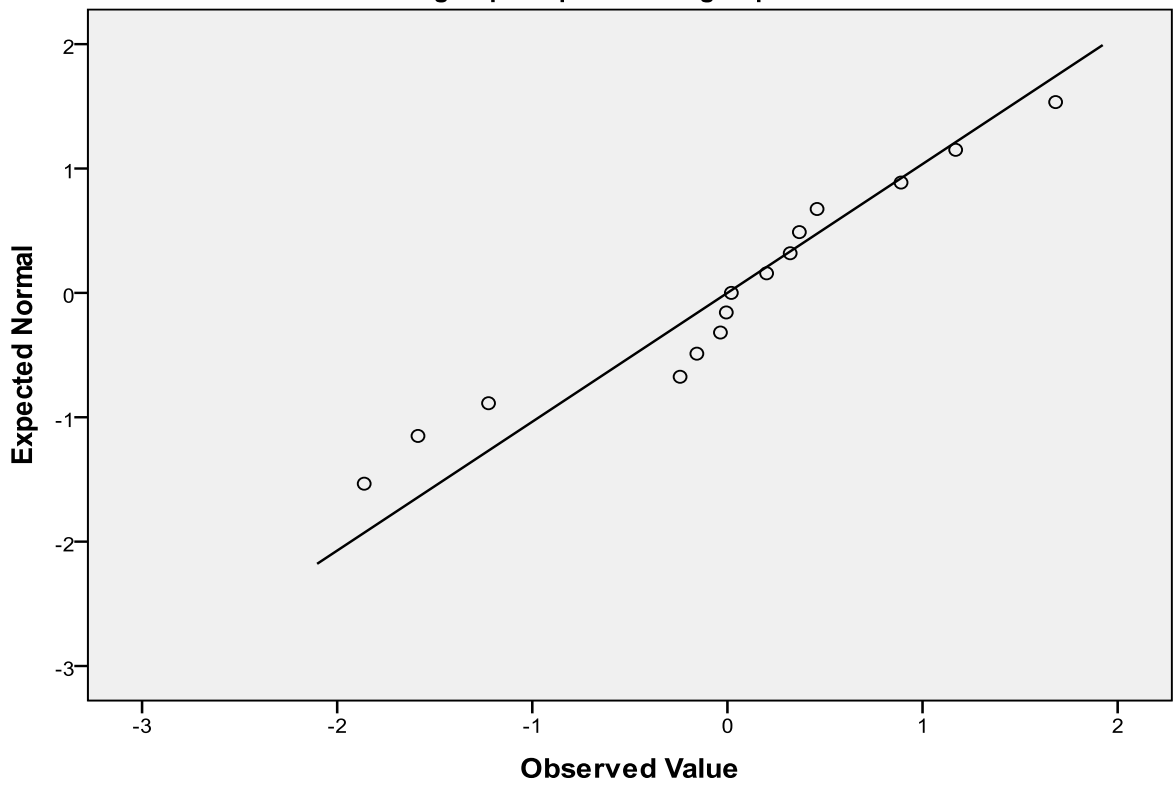
*. This is a lower bound of the true significance.

Normal Q-Q Plot of Standardized Residual for post_intervention_test_score

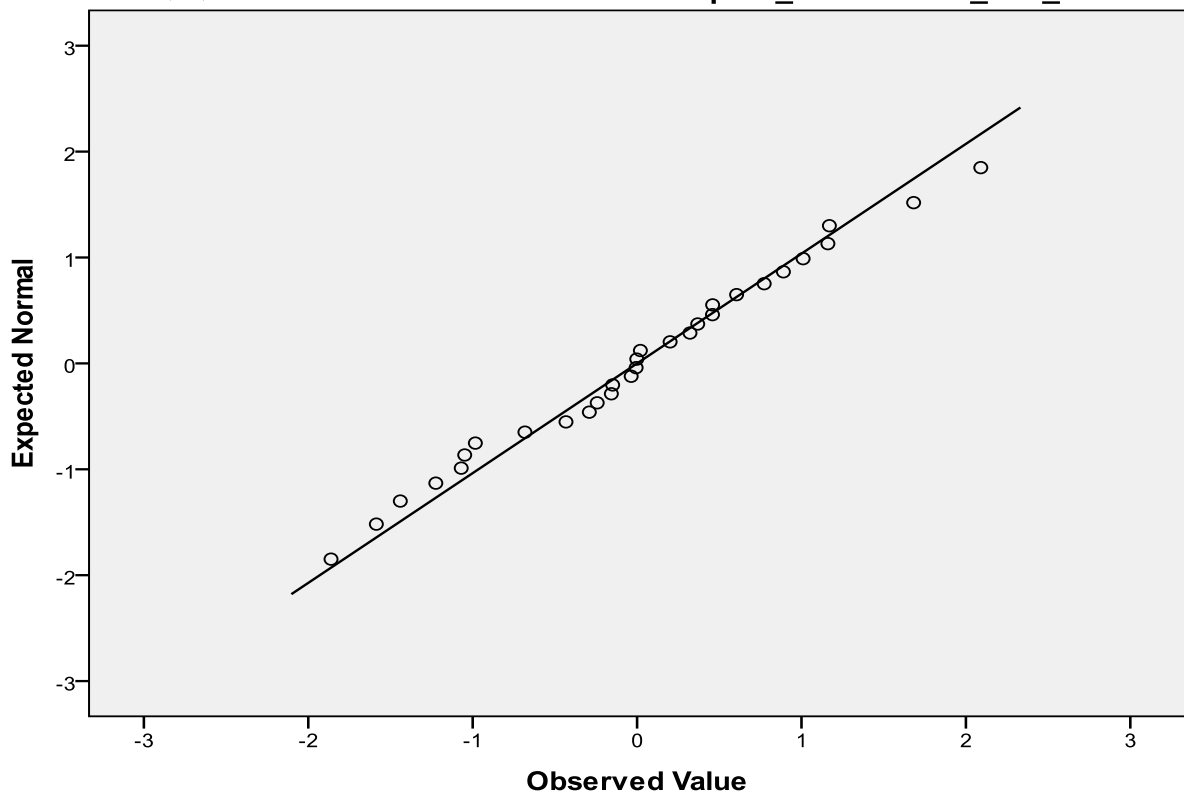
for group= control group



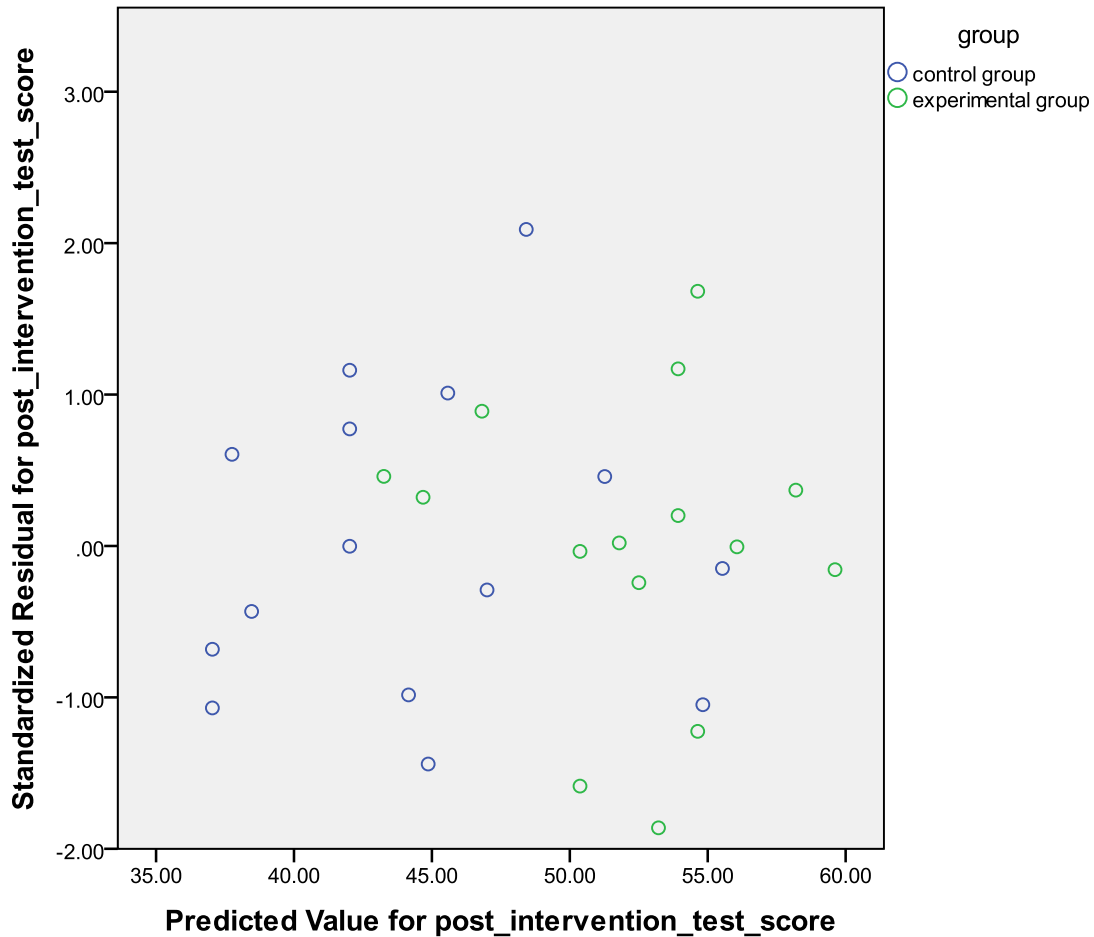
Normal Q-Q Plot of Standardized Residual for post_intervention_test_score
for group= experimental group



Normal Q-Q Plot of Standardized Residual for post_intervention_test_score



Homoscedasticity and homogeneity of variances were confirmed, as assessed by visual inspection of the scatterplot and Levene's test of homogeneity of variance, $p = 0.554$.



Levene's Test of Equality of Error Variances^a

Dependent Variable: post-intervention test score

F	df1	df2	Sig.
.358	1	28	.554

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + prior_academic_performance + group

Furthermore, no standardised residuals greater than ± 3 standard deviations were observed.