International Scholars Journals

African Journal of Political Science ISSN 3461-2165 Vol. 10 (2), pp. 001-007, February, 2016. Available online at www.internationalscholarsjournals.org © International Scholars Journals

Full Length Research Paper

Estimation of effect size in a meta-analysis of series of validity studies on matriculation examinations in Nigeria

E. O. Adeyemo* and E. R. I. Afolabi

Department of Educational Foundations and Counseling, Faculty of Education, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.

Accepted 12 November, 2015

The study formulated an algebraic path from the series of studies meta- analyzed on the validity of the University Matriculation Examination (UME) in Nigeria; in order to obtain a unique and common metrics with a view of making the results to convey the same interpretation. The study determined both the individual and overall effect sizes of 30 empirical studies. It also established the significant difference, in the probability levels and the effect sizes of the selected empirical studies. The study design is descriptive and involved the integration of correlation co-efficient between two variables. The measure of effect size estimate of each of the studies was based on Pearson's product moment indicator using the conversion process extracted from the works of Rosenthal (1984), Glass and Stanley (1970). The results revealed a high effect size of 0.78 and low effect size of 0.03. The empirical studies were not significantly different in terms of their probability levels (x^2 = 2.680, $p > 0.05$) but were significantly different in terms of their effect sizes (x^2 = 1444.97, $p < 0.05$). This study concluded that there was no statistical linear trend in terms of effect sizes across this set of studies and that the heterogeneity of the effect sizes referred to fluctuation from the average of the group. The calculated effect size did not represent adequately the outcome of all independent study, hence an indication of moderator variables operating within the studies.

Key words: Meta-analysis, empirical, effect size, validity, matriculation, examinations.

INTRODUCTION

Meta-analysis refers to the analysis of analyses. It is the statistical analysis of a large collection of analysis from individual studies for the purpose of integrating the findings. Glass (1976) referred to meta-analysis as a set of statistical procedure designed to accumulate experimental and correlational results across independent studies that address a related set of research questions. Unlike traditional research methods, meta-analysis uses the summary statistics from individual studies as the data points. A key assumption of this analysis is that each study provides a differing estimate of the underlying relation within the population. By accumulating results across studies, one can get a more accurate representation of the population relationship that is provided by the individual study estimators. Bangert-Drowns (1991) defined meta-analysis as a collection of systematic techniques of resolving apparent contradiction in research findings. Meta-analysis translates results from different studies to a common metric and statistically explores relations between study characteristics and findings. Educational research often produces contradicting results. Differences among studies in treatment, settings, measurement instruments and research methods make research findings difficult to compare. Frequent replications can prove inconclusive and literature on a topic may be so extensive as to obscure trends with an overwhelming amount of information. The goal of meta-analysis involves the provision of accurate, impartial and quantitative description of the findings in a population of studies on a particular topic. Meta-analysis seeks a full meaningful statistical description of the finding of a collection of studies and this goal typically entails not only

*Corresponding author. E-mail: seyiadeyemo2007@yahoo.com.

a description of the findings in general but also a description of how the findings vary from one type of study to the other.

The most common criticism of meta-analysis is that it is illogical because it tries to make different studies answer the same questions by forcing incommensurable studies together. Implicit in this concern is the belief that only studies that are the same in certain respects can be aggregated. Glass (1982) clarified this criticism by saying that "the claim that only studies that are the same in all respect can be compared is self contradictory; there is no need to compare them since they would obviously have the same findings within statistical error. The studies which need to be compared or integrated are different studies. The smaller data based on only good studies are likely to have too few instances to address many specific questions. Moreover, even when the results of good and bad studies differ, even the bad or not-so-bad studies can be informative". The basic idea in quantitative research integration is to apply statistical methods to the published statistics in previous studies as the data. In the simplest case, all studies that permit a given comparison of interest of the investigator are collected. From each of these studies, a single value is taken expressing the results of the comparison for the study. These values (one from each study) are statistically combined to yield an interpretable summary. The ways of extracting single values from studies and ways of combining these values are closely related. One of this is the effect size.

CONCEPT OF EFFECT SIZE

This is one of the basic methods for quantitative research integration. It is the computed value obtained with respect to a given statistical data of studies expressed in terms of test statistics and sample size of the studies.

STATEMENT OF THE PROBLEM

For the studies considered (Table 1) for meta- analysis, different statistical methods were used. This various statistics included the use of Pearson product moment correlation formula, t-test, chi-squares and analysis of variance (F-ratio). There is the need for the formulation of algebraic path from the reported statistics to a unique metric, in order for the results to convey the same interpretation. The objectives of this study therefore are to: (i) determine the individual and overall effect sizes of the studies and; (ii) establish the significant differences in the probability level of the selected studies.

RESEARCH HYPOTHESES

i) The selected empirical studies are not significantly different in terms of their probability level.
ii) The selected studies are not significantly different in terms of their effect sizes.

METHODOLOGY

This study involved meta-analysis of correlational studies. The 30 studies on validity of UME in Nigeria were purposively selected for meta-analysis on the basis of empirical status and relevance. The reason for choosing Pearson product moment, "r", as the effect size indicator (among other effect size indicators) for this study is as follows;

1) Most of the research studies to be meta-analyzed are correlational studies.
2) r is preferred over d (derived from Glass's or Cohen d or Hedge's g) as an effect size estimate in that d or g may not be accurately computed from the information provided by the author of the original article.
3) Another reason for preferring r to d as an effect size estimate is the simplicity of interpretation of 'r'; since researchers sometimes report their t's and df's but not their sample sizes.

Sampling

Among the 30 studies (Table 1), 16 used Pearson's Product Moment Correlation formula to establish the relationship between UME and the students' academic scores, (CGPA, SSCE or other equivalent examinations) . The Pearson Product Moment Correlation coefficients "r" reported by primary researchers was recorded as effect size estimates. There was no need for transformation of any kind since Pearson Product Moment Correlation coefficients were reported. A thorough computational checking of previously reported calculations was done for the purpose of verification. When "r"s were calculated for the scores of students in different faculties or universities, the mean "r" was calculated and reported for the purpose of meta-analysis. Four studies used t-test as means of reporting the differences between students' scores in UME and university academic performance or CGPA. The t -ratio obtained was converted to effect size "r". Four of the studies used ANOVA. The F- ratio was transformed to t = r that is, $\sqrt{F} = /t/$ since df = 1. Degree of freedom was based on degree of freedom for between means.

Sidney Siegel (1956) and Kendel and Stuart (1967) have presented a way of calculating contingency co-efficient "r" from an R × C table. The contingency coefficient "r" like r_{xy} is said to measure the extent of association between two sets of attributes. Given chi-square χ^2 with (r – 1) (c – 1) degree of freedom, Kendall and Stuart, and Sidney Siegel have derived a formula for transformation of chi-square χ^2 into r_{xy}, and their formula is $r_{xy} \simeq C = \frac{x^2}{x^2 + n}$ where n is equal to the total sample used. Two studies reported z-score, $r = \frac{Z}{\sqrt{n}}$ was used. Using the thus procedures listed, calculation of effect size estimate "r" became easy to compute.

Data analysis

From Table 2, the highest effect size was 0.855 while the lowest was 0.04. Six studies recorded high effect sizes, 12 studies recorded moderately low while 12 studies recorded very low-effect sizes. The computed effect sizes 'r' for each study was converted to the Fisher scale to take care of both low and high effect sizes. The works of Rosenthal (1994), Glass and Stanley (1970), Jankins

Table 1. Characteristics of the 30 empirical studies selected.

Studies	Type/Level of study	Sample size	Study coverage	Statistics used	Level of significant	Statistical results
1.	Masters (unpublished) (2003)	250	Within faculty	Multivariate	0.05	r = 0.3880
2.	Ph. D (unpublished (2003)	558	Across universities	Multivariate	0.05	Beta UME = 0.3298 for Maths 0.3154 for biological sciences
3.	Masters (unpublished) (1985)	300	Across faculties	Bivariate	0.05	X^2= 4.36
4.	Masters (unpublished) (1983)	121	Within faculty	Bivariate	0.05	t = 0.58 (S.Ed) t = 1.12 (S.S) t = 1.02 (PHE)
5.	Masters (unpublished) (1995)	40	Within a faculty	Bivariate	0.05	r = 0.0421
6.	Ph. D. unpublished (1998)	800	Across universities	Multivariate	0.01	r = 0.2834
7.	Undergraduate (unpublished) (1985)	30	Within faculty	Bivariate	-	r $= 0.\overline{0}9$
8.	Journal (published) (1985))	1800	Comparison with subjects	Univariate (frequency counts)	0.01	Coding reliability r 0.855
9.	Ph. D. (unpublished)(2006)	750	Across universities	Multivariate	0.05	r = 0.175
10.	Journal (published) (2002)	802	Across faculties	Bivariate	0.05	r = $\overline{0}.086$
11.	Journal published (1983)	100		Bivariate	0.01	r = 0.72 (chem.) r = 0.59 (phy) +ve r = 0.41 (econs) but r = 0.20 (geo) –ve r = 0.32 (bio) –ve
12.	Masters (unpublished) (1991)	123	Within faculty using only educationists	Bivariate	0.05	t = 0.405
13.	Ph. D (Unpublished) (1991)	1381	Across universities	Multivariate	0.01	F = 2.47 P = 0.0117
14.	Journal (published) (1985)	42	Within a faculty for a period of 4 years	Multivariate	0.05	r = 0.36
15	Journal (published) (1986)	120	Within a faculty	Multivariate (r & multiple reg)	r = 0.01	Corrolation matrix for JAMB = 0.6995 F=55.11 JAMB
16	Journal (1990)	400	Across faculties	Multivariate	0.05	r = 0.04
17	Masters unpublished (1987)	687	Across faculties	Multivariate	0.01	r =0.27 for UME& SSCE t = 0.302. also UME and first year performance r = 0.207 t = 3.007

Table 1. Contd.

18	Journal (2001)	180	Across faculties	Multivariate	0.05	r = 0.41 (arts) r = 0.32 (science) r = 0.19 (social science)
19	Journal (published) (2004)	54	Within faculty	Bivariate	0.05	r = 0.300
20	Special reports (2006)	866	Across faculties	Multivariate	0.01	r = 0.479
21	Undergraduate (unpublished) (1994)	180	Across faculties	Bivariates	0.05	r = 0.41 (arts) r = 0.21 (science)
22	Journal (published) (2001)	227	Within faculties (medicine)	Multivariate	0.05	r = 0.42
Studies	**Type/Level of study**	**Sample size**	**Study coverage**	**Statistics used**	**Level of significant**	**Statistical results**
23	Special reports (2005)	6462	Across faculties	Multivariate	0.05	t = $0.\bar{7}\ 5$, P value = 0.48 T = -2.92 P value = 0.004
24	Masters unpublished (1988)	107	Across universities	Bivariates	0.05	r = 0.12
25	Journal (published) (1984)	78	Within faculty	Multivariate	0.05	Z =3.15
26.	Journal (published) (1983)	60	Within faculty	Bivariate	0.05	r = 0.6248
27.	Journal (published) (2003)	159	Across universities	Multivariate	0.05	F = 10.414 at 0.05 P = 3.15
28.	Journal (published) (2003)	212	Within faculty	Bivariate	0.05	r = 0.0251
29.	Journal (published) (2001)	30	Within faculty	Fisher's correlation (Bivariate)	0.05	Z = 1.31
30.	Journal (published) (1997)	222	Within faculty	Bivariate	0.05	r = 0.78

(1955), Rosenthal and Rosnow (1984), and Hayes (1973) were used for the analysis to convert various summary statistics (t-values, r-values, chi-square) into product-moment correlation. Before conducting the statistical analyses, Fisher's Z transformation was applied on all correlation co-efficient based on procedures suggested by Glass and Stanley (1970). After performing the appropriate analysis, Fisher Z scores were transferred back into the more interpretable correlation – coefficient. Each probability level was converted to the standard normal deviate and the computed individual and overall effect sizes were determined.

RESULTS AND DISCUSSION

Probability levels of selected studies

Hypothesis 1*: *The selected studies are not significantly different in terms of their probability levels.

This is used to establish the significant differences in the probability level of the selected studies.

From Table 3, observed χ^2 = 2.6804, the critical (table value) of χ^2 at P – level of 0.05, df = 29 is greater than the observed (calculated value). The null hypothesis is accepted. This implies that the selected empirical studies are not significantly different in terms of their probability levels.

Before determining whether the 30 studies differed significantly among themselves with respect to their effect sizes, (r's) it is important to resolve the issue of very low and high effect size

Table 2. Probability level and estimated effect sizes for the 30 studies.

Study	Sample size	P-levels	Effect size 'r'
1.	250	0.05	0.3880
2.	558	0.05	0.3226
3.	300	0.05	0.2106
4.	121	0.05	0.4747
5.	40	0.05	0.0421
6.	800	0.01	0.2834
7.	30	0.01	0.09
8.	1800	0.01	0.855
9.	750	0.05	0.175
10.	802	0.05	0.086
11.	100	0.01	0.573
12.	123	0.05	0.2882
13.	1381	0.001	0.6115
14.	159	0.05	0.3567
15.	120	0.01	0.6995
16.	40	0.01	0.04
17.	687	0.01	0.209
18.	180	0.05	0.365
19.	54	0.05	0.300
20.	866	0.01	0.479
21.	180	0.05	0.31
22.	227	0.05	0.42
23.	6462	0.02	0.4286
24.	107	0.05	0.12
25.	78	0.05	0.3567
26.	60	0.05	0.6248
27.	42	0.05	0.36
28.	212	0.05	0.0251
29.	30	0.05	0.2397
30.	222	0.05	0.783

r's. A glance at the value of r's recorded in Table 2 revealed an extreme high correlation coefficient of r = 0.78 (study 30) and extreme low correlation coefficient of r = 0.03 (study 28). These facts about the effects of extreme low and extreme high correlation coefficients (r') complicate the comparison and combination of r's. This complication has been addressed by Fisher (1928) when he devised a transformation Zr that was distributed nearly normally. Therefore all effect sizes 'r' were transformed to Fisher Zr before any computation could be carried out. Thus the use of Fisher's Zr helped to resolve the problem of the effect and contribution of extreme low and high correlation coefficients. Although the median value could be used as alternatives to solving the problem. Yet the Fisher's "Zr" gave heavier weights to 'r' that were further from zero in either direction.

Hypothesis 2: The selected studies are not significantly different in terms of their effect sizes.

To test the null hypothesis, a diffused test given by Snedeco and Cochram (1967; 1980) was used to asses the statistical heterogeneity of the 30 effect sizes …..

$$\chi^2 = \Sigma\,(Nj - 3)(Z - \bar{Z})^2 \text{ with } k - 1 \text{ df.}$$

From Table 4, the observed chi-square (χ^2) = 1444.97 while the critical (table) value at 29 df = 42.557. Since χ^2 observed is greater than χ^2 critical, the null hypothesis is rejected. This implies that the selected studies are significantly different in terms of their effect sizes 'r'. This is an indication that there is no statistically significant linear trend in terms of effect size across this set of stu-

Table 3. Computation of chi-square using P-value.

Study	P (one-tailed)	Z (Standard normal deviate)	Z - $\overline{Z}$	$(Z - \overline{Z})^2$
1.	0.05	1.64	-0.18	0.0324
2	0.05	1.64	-0.18	0.0324
3	0.05	1.64	-0.18	0.0324
4	0.05	1.64	-0.18	0.0324
5	0.05	1.64	-0.18	0.0324
6	0.01	2.32	0.5	0.25
7	0.01	2.32	0.5	0.25
8	0.01	2.32	0.5	0.25
9	0.05	1.64	-0.18	0.0324
10	0.05	1.64	-0.18	0.0324
11	0.01	2.32	0.5	0.25
12	0.05	1.64	-0.18	0.0324
13	0.01	2.32	0.5	0.25
14	0.05	1.64	-0.18	0.0324
15	0.01	2.32	0.5	0.25
16	0.05	1.64	-0.18	0.0324
17	0.01	2.32	0.5	0.25
18	0.05	1.64	-0.18	0.0324
19	0.05	1.64	-0.18	0.0324
20	0.01	2.32	0.5	0.25
21	0.05	1.64	-0.18	0.0324
22	0.05	1.64	-0.18	0.0324
23	0.05	1.64	-0.18	0.0324
24	0.05	1.64	-0.18	0.0324
25	0.05	1.64	-0.18	0.0324
26	0.05	1.64	-0.18	0.0324
27	0.05	1.64	-0.18	0.0324
28	0.05	1.64	-0.18	0.0324
29	0.05	1.64	-0.18	0.0324
30	0.05	1.64	-0.18	0.0324
		1.82		2.6804

*p>0.05 (not significant)

Table 4. Computation of chi-squared using correlation coefficient effect size ‘r’.

Study	Sample size	N-3	r	Zr	$Zr - \overline{Zr}$	$(Zr - \overline{Zr})^2$	(N-3) $(Zr - \overline{Zr})^2$
1	250	247	0.39	0.4118	0.018763	0.000352	0.086956392
2	558	555	0.32	0.3310.6	0 .061437	0.003775	2.094850258
3	300	297	0.21	0.232	0.0179837	0.32341	9.605379931
4	121	118	0.47	0.5101	0.117063	0.013704	1.617042024
5	40	37	0.04	0.04	0 .353037	0.124635	4.611499565
6	800	797	0.28	0.2877	0 105337	0.011096	8.843419204
7	30	27	0.09	0.0902	0 302837	0.09171	2.476176711
8	1800	1797	0.86	1.1155	0.722463	0.521953	939.9491571
9	750	747	0.18	0.182	0 .211037	0.044537	33.26885168
10	802	799	0.09	0.0902	0 .302837	0.09171	73.27648861
11	100	97	0.57	0.6475	0.254463	0.064751	6.280887582
12	123	120	0.29	0.2986	0 .094437	0.008918	1.070201636

Table 4. Contd.

13	1379	1376	0.61	0.7089	0.315863	0.099769	137.2827422
14	30	27	0.24	0.2448	$\overline{0}$ .148237	0.021974	0.593303621
15	40	37	0.04	0.04	$\overline{0}$ .353037	0.124635	4.611499565
16	120	117	0.70	0.8673	0.474263	0.224925	26.316271
17	687	684	0.21	0.2132	$\overline{0}$ .179837	0.032341	22.12148105
18	180	177	0.37	0.3884	$\overline{0}$ .004637	2.1505	0.003805813
19	54	51	0.30	0.3095	$\overline{0}$ .083537	0.006978	0.355899949
20	860	857	0.48	0.533	0.139963	0.01959	16.78832265
21	180	177	0.31	0.3205	$\overline{0}$ .072537	0.005262	0.9313406097
22	227	224	0.42	0.4477	0.054663	0.002988	0.669321759
23	6462	6459	0.43	0.4477	0.054663	0.002988	19.29977341
24	107	104	0.12	0.1206	$\overline{0}$ .272437	0.74222	7.719079573
25	78	75	0.36	0.3769	$\overline{0}$ .016137	0.00026	0.019530208
26	60	57	0.62	0.725	0.331963	0.110199	6.281367702
27	159	156	0.36	0.3769	$\overline{0}$ .016137	0.00026	0.040622832
28	212	209	0.03	0.03	$\overline{0}$ .363037	0.131796	27.54533544
29	42	39	0.36	0.3769	$\overline{0}$ .016137	0.00026	0.010155708
30	222	219	0.78	1.0454	0.652365	0.425577	93.20146895
		16683	Mean. Fisher	0.393037			1444.972198
			W. Fisher	0.434713			

*p<0.05 (significant).

dies. The heterogeneity of the set of effect sizes referred to fluctuations from the average of the group. The implication of this is that the calculated average effect size did not represent adequately the outcome of all independent studies. The heterogeneity of the effect sizes was indicative of moderator variables operating. This could be a function of the sample size, publication or methodological features.

Conclusion

The combined effect size of series of studies for meta-analysis on validity of UME in Nigeria is statistically significant. This is an indication that there is no linear trend in terms of effect size across this set of studies.

REFERENCES

Abdullahi A (1983). A study of the predictive value of Joint Matriculation Examination in selected school subjects. J. Niger. Educ. Res. Asso. 3(3): 29-34.

Aina JA (1985). "The Cognitive demands of the Joint Matriculation Examination (JME) Mathematics" J. Teachers' Educ. 14: 43-57.

Alonge MF (1986). Cognitive entry characteristics and formative Evaluation as measures of academic performance among university undergraduate. Afr. J. Res. Educ. (1) No 1.

Bangert-Down RL (1991). Meta-Analysis in educational research. Practical assessment research and Evaluation.

Bangert-Drown RL (1986). Review of development in meta-analytic methods. Psychological Bulletin pp. 388-399.

Fisher RA (1928). "Statistical Methods for research work (2nd ed) London. Oliver and Boyd.

Glass GV (1976). "Primary, Secondary and meta-analysis of research "A paper presented at the annual meeting of the American Educational Research Association San-Francisco

Glass GV (1982). School class size: Research and Policy, Beverly Hills C.A. SAGE Publication.

Glass GV, Stanley JC (1970). "Statistical methods in education and psychology" Englewood Cliffs, New Jersey: Prentices-Hall.

Hayes WL (1973). "Statistics for the Social Sciences" New York: Holt Rinehart and Winston.

Jerkins WL (1955). "An improved method for tetrachoric 'r' Psychometrical 20(3): 253-258.

Kendall MG, Stuart A (1967). "The Advance theory of statistics" London Griffin p. 557.

Rosenthal R (1984). "Meta-analytic procedures for social research "Beverly Hills California" Sage Publications Inc.

Rosenthal R, Rosnow RL (1984). "Essential of behavioural research (methods and data analysis)" McGraw-Hill, Inc.

Sidney S (1956). "Non parametric Statistics for the behavioural sciences" Mc Graw-Hill Book Company; New York pp. 196-202