INFLUENCE OF SIZE EFFECT ON PREDICTIVE VALIDITY OF UTME SCORES IN NIGERIAN PUBLIC UNIVERSITIES: A META ANALYSIS

By

Kareem Azeez Olawale,
Department of Curriculum & Instruction, Sikiru Adetona College of Education Science & Technology, (SACOETEC), Omu-Ajose
07064207652, kazeez2014@gmail.com
&
Prof. Janet O. Adetayo
Department of Science and Technology Education, Faculty of Education,
Olabisi Onabanjo University, Ago-Iwoye.
08163552814, adetayo.janet@oouagoiwoye.edu.ng
&
Prof. Opoola A. Oyedeji
Department of Science and Technology Education, Faculty of Education,
Olabisi Onabanjo University, Ago-Iwoye.
08034483431, opoolaoyedeji@gmail.com

Abstract
The Unified Tertiary Matriculation Examination (UTME) conducted in Nigeria by the Joint Admissions and Matriculation Board (JAMB) is essentially an aptitude test designed to predict candidates' intellectual ability to successfully undertake undergraduate programmes at the tertiary education level. Many studies have been conducted on the predictive validity of the UTME and their findings varied considerably suggesting the need for a meta-analytical study to further assess and put the various findings in proper perspective. Hence, this work was on influence of size effect on the predictive validity of UTME scores in Nigeria public universities using a meta-analytical approach. The ex-post facto form of descriptive research design was used in the study. A total of ninety six (96) studies on validity of UTME consisting of some published and unpublished articles were gathered for the study. Out of the entire population, thirty (30) studies were purposively selected on the basis of empirical status and relevance. A computer search through the internet and manual – search through the visitation to relevant Departments in the Universities and scholarly Journal papers were consulted in order to obtain the selected studies. The instruments used for this study were research results from published and unpublished journal articles. A self-made Profoma known as Coding Sheet designed by the researchers was used to document the characteristics of the sampled publications. Findings revealed that size contributed a reasonable difference in the magnitude of the selected study. It was also revealed that there was an indication of influence of linear trend in terms of size effect across the set of studies on predictive validity of UTME. Therefore, it was recommended that sample size on predictive validity of UTME should not be less than 120 in other to reduce the influence of size effect.

Keywords: Size Effect, UTME, Meta-analysis, Predictive Validity.

Introduction

In Nigeria, those who successfully complete a course of study with good academic records are awarded certification at any level of education. As a result, at the end of secondary school, students are expected to sit for public examinations such as the West African Senior School

Certificate Examination (WASSCE) administered by the West African Examinations Council (WAEC), the Senior School Certificate Examination (SSCE) administered by the National Examinations Council (NECO), and the National Technical and Business Certificate Examinations (NTCE/NBCE) administered by the National Business and Technical Examinations Council (NBCEC).

Public examinations are defined as external school examinations that are open to the general public and are administered by examining bodies using relevant psychometric testing According to Adeyegbe (2004), the examinations utilized by various public examination boards are frequently better designed than those provided by teachers in the classroom. A minimum of five credit passes in any of these public examinations is required for admission to universities after taking the Joint Admissions and Matriculation Board's University Tertiary Matriculation Examination (UTME).

Candidates' admission or placement at Nigerian universities, whether federal, state, or privately owned, is reliant on attaining the prescribed cut-off mark in the UTME. Prospective applicants have been required to take university Post UTME tests or screening examinations as a condition of admission for more than a decade. It is believed that these entry qualifications and entrance examinations will positively predict candidates' performance in the university. The system seems not to have -totally evolve due to the recent confusion created by the Federal Government's desire to scrap the Post UTME and extend the validity of the UTME scores to three or more years.

A crucial datum of meta-analysis is an estimate of the magnitude of relationship between two variables, referred to as size *effect estimates.* The meta-analysis literature contains several proposals for how best to calculate size effect estimates but no one method is universally adopted. Little is known about the consequences of calculating size effect estimates one way or another and a judgment call must be made by the meta-analyst concerning how to calculate this crucial datum. Ambiguity concerning how best to calculate effect size estimates exists for several reasons. The formula proposed by Glass et al. (1981) and presented earlier is an adaptation of Cohen's (1969). A point of disagreement prominent in the literature concerns the proper denominator for this formula.

Glass et al. (1981) argue that the standard deviation of the control group is the proper term, whereas Schmidt et al. (1984) argue that the denominator should be a variance estimate based on pooling the variance estimates of treatment and control groups. Other estimates of the magnitude of a relationship, such as Rosenthal's (1984), exist that, theoretically, can be used interchangeably with the former. Further, estimates of the magnitude of relationship can be inferred in the absence of direct reports of means and standard deviations or correlations. Glass et al. (1981) provide formulas for estimating size effects on the basis of reported t-test values, analyses of variance results, time-series data and on the basis of data expressed as proportions. These formulas give the Meta analyst considerable power to make use of many forms of quantitative data reported in a literature. Although conceptually similar, little is known about the actual empirical differences that might appear in size effect estimates calculated through the various procedures.

Complexity in research designs also poses difficulties for the calculation of effect size estimates. For example, consider a study containing elements of both true experiments and time series designs, such as one in which groups, randomly assigned to treatment and control conditions, measured on a dependent variable on repeated occasions, some occurring before and some after the treatment. How should effect sizes be calculated in this instance? Should several effect size estimates be calculated, one for each time of measurement of the dependent variable, or should only one effect size be calculated, reflecting an average of the multiple measures of the dependent variable? Is between-group comparison more relevant than the comparison between pre- and post-intervention measures? On what basis should an estimate of the variance in the dependent variable be calculated? Although answering these questions is beyond the scope of this research, their answers do have implications for the value of the effect size estimate, the primary datum of meta-analysis. The point of raising such questions is to illustrate how judgment calls permeate the process of data analysis in meta-analysis.

Detecting Moderators: The final stage of a meta-analytic review involves analyses to detect important moderator variables that might explain variance in observed effect sizes. Depending upon which meta-analytic procedures one is following, the decision of *whether* to examine the data for possible moderator effects is either a judgment call made by the investigator (when using the Glass et al., 1981 techniques) or a decision based on the results of statistical analyses (when using the Hunter et al., 1982 techniques).

According to Glass et al. (1981), any meta-analysis should end with a search for moderator effects. To date, Glassian meta-analyses have typically focused on examining methodological differences across studies to determine whether these methodological differences can account for differences in effect sizes. Thus, Glass et al. encourage researchers to code studies for variables such as whether the study included a true control group, the study's sample size, year of study, and the amount of time that elapsed between the time of the intervention and assessment of outcomes.

Predictive validity as a form of validity that seeks to measure the extent to which a test predicts students' future performance. Prediction in the broad sense of the term consists essentially of estimating the values of some function of variables over time, on the basis of certain present attributes, which may or may not contain random errors. It is the ability to estimate future achievement based on the past or present achievement.

Predictive validity is most commonly used when exploring data in the field of psychological study and analysis. It is used to collect information about various populations, and to create generalizations which may be useful when assessing individuals. For example, it is often used by big companies that administer a test to prospective employees, comparing test data from current employees to determine whether or not someone will be qualified for the job, it is equally used in institutions for comparing performances of students. Since JAMB started its operation, individuals, corporate bodies and different levels of government have accused JAMB of massive corrupt practices (Adebayo, 2011 and Bernardine, 2019).

In Nigeria, the UTME is an aptitude test trying to predict students' achievement in universities and its effectiveness depends largely on the extent to which it could do this, hence the need to always assess its predictive validity. Predictive test is a measurement of how well a test predicts future performance. It is a form of criterion validity in which how well the test works is established by measuring it against a known criterion. In order for a test to have predictive validity, there must be a statistically significant correlation between test scores and the criterion being used to measure the validity. One of the classical examples of this is the UTME. When students apply to Colleges, Polytechnics and Universities, they are usually required to submit test scores – from examinations such as the WAEC, JAMB/

UTME. These scores are used as bases for comparison, in which evaluators look at the performance of students who have had similar scores in the past. The belief is that the test scores can predict how well a student will perform in the university (college). High test scores tend to be correlated with good performance in the university, making students with high scores appealing for admission.

Bernardine (2019) stated that there is no relationship between the chemistry UME scores and their first - year scores in State Universities. For medicine and surgery, in both the Federal and State Universities, there is weak positive relationship between UME scores in chemistry and the students' first year scores in the discipline. (Bala, 2019). The prediction of students CGPA from their performance in UTME and PUTME in Kaduna State University reveals that UTME and PUTME are good predictors of students' final class of degree. Biman (2019) stated that there is no significant contribution of UTME on students' CGPA. UTME has no prediction on the overall students' CGPA. The result shows that UTME results had no predictive strength with β = -0. 013 with R2 = 0% to the overall students' CGPA.

Popoola (2016) stated that University Matriculation Examination (UME) served as a good predictor of the students' performance in their first - year results. Egberha, (2019) revealed that University Tertiary Matriculation Examination (UTME) scores have a low predictive power of 0.009 that is UTME score accounted for only 0.9% of the total variance. This suggests that 99.1% of the variance of student's first year academic performance is accounted for by other factors other than UTME scores. Imasuen (2020) revealed that UTME scores do not significantly predict undergraduate final grades in Nigerian University. UTME score only accounted for about 0.01% in students' final grade in Nigeria universities.

Ituma, Ugwuanyi, and Uzochukwu (2023) stated that there is a moderate positive relationship between the students' UTME scores in Mathematics and the first - year mean achievement scores of the students in undergraduate Mathematics courses. Also, the students' UTME scores in Mathematics accounted for 30.7% of the total variations in the students' achievement in undergraduate Mathematics programme.

Meta-analysis is bringing together of data from a large collection of past research on a particular topic for the purpose of integrating the findings. In meta-analysis, primary research reports constitute the data for statistical integration. Glass (1976) argues that both the primary studies and their findings are quantified so that the statistical integration can be performed. Meta-analysis is to correct the weaknesses in individual research by integrating the findings of past research studies.

Statement of the Problem

There are often a lot of complaints about students' poor academic performance in the Nigeria Universities over the years. One of the fundamental issues associated with the university education system in Nigeria is mandatory use of UTME scores for the purpose of admission for all public and private higher institutions in Nigeria.

However, there have been inconsistencies in the results of past research studies on predictive validity of university matriculation examinations. Some studies reported high predictive validity while some claimed that UTMEs lacked predictive validity. To put the situation into proper perspective, meta-analysis was needed in order to determine the strength of the predictive validity of UTMEs.

Research Questions

i. Will the studies differ significantly among themselves as regards the size effects of the predictive validity of UTME?
ii. What is the significant difference in the size effect of the published and unpublished journals on predictive validity of UTME?

Research Design

An ex-post facto descriptive research design was employed to carry out this study. This study is the determinant of the predictive validity of Unified Tertiary Matriculation Examination (UTME) in Nigeria public Universities a meta analytical approach. The ex-post facto research design was used because the researcher was not in a position to manipulate the past research works that were analysed.

Population of the study

The population of this study included all available published articles or journals, unpublished masters' dissertations, and PhD theses that focused on predictive validity of UTME in Nigeria from 1998 - 2019. A total of ninety - six (96) studies on validity of UTME both published and unpublished articles were gathered for this study.

Sample and Sampling Techniques

Out of the entire population, thirty (30) studies were purposively selected on the basis of empirical status and relevance. A computer search through the internet and hand – search through the visitation to relevant department in the Universities and Scholarly Journal papers were consulted in order to access the selected studies.

Research Instruments

A self-made Profoma known as Coding Sheet designed by researchers was used to document the characteristics of the sampled research results from published and unpublished journal articles.

Results and findings

Research Question 1: Will the studies differ significantly among themselves as regards the size effect of the predictive validity of UTME?

To answer this research question, a diffused test given by Snedeco and Cochran (1967; 1980) was used to calculate the size effects on the thirty (30) empirical studies.

$$\chi 2 = \sum (\mathrm{Nj} - 3)(\mathrm{Z} - \bar{\mathrm{Z}})^2 \text{ with } \mathrm{K} - 1 \text{ df}$$

Table 1: Computation of Chi-squared Using Correlation Coefficient Effect Size 'r'

Study	Sample Size	N-3	r	Zr	Zr-$\bar{Z}$r	(Zr-$\bar{Z}$r)^2	(N-3)(Zr-$\bar{Z}$r)^2
1	100	97	0.880	1.3758	.912943	.833465	80.8460974
2	101	98	0.500	0.5493	.086443	.007472	.7322944
3	500	497	0.241	0.2448	-.218057	.047549	23.6317811
4	300	297	0.820	1.1568	.693943	.481557	143.0223955
5	1500	1497	0.202	0.2027	-.260157	.067682	101.3194520
6	250	247	0.380	0.4001	-.062757	.003938	.9727949
7	2518	2515	0.153	0.1511	-.311757	.097192	244.4389540
8	558	555	0.320	0.3316	-.131257	.017228	9.5617620
9	800	797	0.280	0.2877	-.175157	.030680	24.4519398
10	750	747	0.175	0.1769	-.285957	.081771	61.0832402

11	2400	2397	0.266	0.2715	-.191357	.036618	87.7721510
12	4904	4901	0.080	0.0802	-.382657	.146426	717.6356867
13	1370	1367	0.002	0.0050	-.457857	.209633	286.5683554
14	336	333	0.100	0.1003	-.362557	.131448	43.7720436
15	720	717	0.328	0.3372	-.125657	.015790	11.3212017
16	253	250	0.900	1.4722	1.009343	1.018773	254.6933229
17	103	100	0.177	0.1769	-.285957	.081771	8.1771406
18	143	140	0.437	0.4661	.003243	.000011	.0014724
19	408	405	0.195	0.1975	-.265357	.070414	28.5178067
20	8111	8108	0.009	0.0100	-.452857	.205079	1662.7842815
21	74	71	0.930	1.6584	1.195543	1.429323	101.4819376
22	720	717	0.231	0.2342	-.228657	.052284	37.4876450
23	1134	1131	0.003	0.0030	-.459857	.211468	239.1708288
24	220	217	0.920	1.5890	1.126143	1.268198	275.1989782
25	240	237	0.770	1.0203	.557443	.310743	73.6460195
26	1610	1607	0.001	0.0010	-.461857	.213312	342.7922047
27	135	132	0.006	0.0060	-.456857	.208718	27.5508180
28	471	468	0.018	0.0182	-.280857	.078881	36.9161463
29	1500	1497	0.009	0.0100	-.452857	.205079	307.0039553
30	943	940	0.831	1.1881	.725243	.525977	494.4187645
		33082	Mean Fisher	0.462857			5726.971472
			Weighted Fisher	0.586914			

Results presented in Table 1 showed that the observed Chi-square (χ^2) was 5726.97 while the critical (table) value at 29 degrees of freedom was 42.557. Since Chi-square (χ^2) calculated value was greater than Chi-square (χ^2) critical value, it means that the selected studies are significantly different in terms of their size effects 'r'. This result implied that the selected studies differ significantly among themselves as regards the size effects of the predictive validity of UTME. This outcome is an indication that there is linear trend in terms of effect size across this set of studies. 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 empirical studies. The heterogeneity of the effect sizes was indicative of moderator variables operating. This could be a function of the study characteristics, sample size, publication or methodological features.

The finding on the first research question revealed a reasonable difference in the magnitude of the mean effect sizes of published and unpublished articles. It revealed that the mean effect size of unpublished articles was above the overall weighted mean effect size of the 30 selected articles while the mean effect size of the published articles fell below the overall weighted mean effect of the 30 selected articles. This result tends to be consistent with the findings from a previous study by Sterne, Gavaghan, Egger and Epidemoil (2000) which also reported variance in the mean effect size of published and unpublished articles.

Also, the results indicated that the effect size of the unpublished articles was low compared to the published articles. Moreover, this finding lends credence to Antonakis (2017) conclusion from a study that this trend had been reported in previous studies on publication and related bias in meta- analysis that publication bias is more likely to affect small sample studies which also tend to be of lower methodological quality and this may lead to small study effects where the smaller studies in meta-analysis show larger treatment which may also arise because of between trial heterogeneity.

The selected studies differ significantly among themselves as regards the effect sizes of the predictive validity of UTME. This outcome is an indication that there is linear trend in terms of effect size across this set of studies. 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 empirical studies. This according to Sterne, Gavagha, & Egger (2000) may likely affect small studies which also tend to be of lower methodological quality. This may lead to small study effects where smaller studies in a meta-analysis show larger treatment.

Research Question 2: What is the significant difference in the size effect of the published and unpublished journals on predictive validity of UTME?

To answer this research question, a Chi-square (χ^2) formula given by Rosenthal and Rotibin (1979) was used to determine the significant difference on the size effect of published and unpublished journals on predictive validity of UTME.

$$\chi 2 = \sum_{j=1}^{k} (Nj - 3)(Zrj - \bar{Z})^2 \text{ is distributed for } \chi 2 \text{ with } K - 1 \text{ df}$$

Table 2: Effect Sizes of Unpublished Articles

Study	N	r	Zr
1	100	0.880	1.3758
2	101	0.500	0.5493
3	500	0.241	0.2448
4	300	0.820	1.1568
5	1500	0.202	0.2027
6	250	0.380	0.4001
7	2518	0.153	0.1511
8	558	0.320	0.3316
9	800	0.280	0.2877
10	750	0.175	0.1769
n=10	737.7	0.3951	0.4177

Table 3: Effect Sizes of Published Articles

Study	N	r	Zr
11	2400	0.266	0.2715
12	4904	0.080	0.0802
13	1370	0.002	0.0050
14	336	0.100	0.1003
15	720	0.328	0.3372
16	253	0.900	1.4722
17	103	0.177	0.1769
18	143	0.437	0.4661
19	408	0.195	0.1975
20	8111	0.009	0.0100
21	74	0.930	1.6584
22	720	0.231	0.2342
23	1134	0.003	0.0030
24	220	0.920	1.5890

25	240	0.770	1.0203
26	1610	0.001	0.0010
27	135	0.006	0.0060
28	471	0.018	0.0182
29	1500	0.009	0.0100
30	943	0.831	1.1881
n=20	1289.75	0.311	0.3205

Substituting the values in the above expression

$$Zr = \frac{(737.7 \times 0.4177) + (1289.75 \times 0.3205)}{1105.73}$$

$$Zr = 0.653$$

$$\chi 2 = 737.7(0.4177 - 0.653)^2 + 1289.75(0.3205 - 0.653)^2$$

$$\chi 2 = 40.8436 + 142.5899$$

$$\chi 2 = 183.433$$

Table 4: Computation of Difference in the Effect Size of the Published and Unpublished Articles

Publications	Statistics				
	n	Mean of N	r	Zr	χ^2
Unpublished	10	737.7	0.3951	0.4177	183.433
Published	20	1289.75	0.3110	0.3205	

P < 0.05 is significant

From the results in Table 2, the Chi-square (χ^2) calculated value was 183.433 at 1 degree of freedom at p=0.05. The observed (that is, calculated) Chi-square (χ^2) value of 183.433 was greater than the critical or table value of 3.841. Thus, dismissing the notion that there was no observable significant difference in the effect size of the published and unpublished articles on predictive validity of UTME. It means that the effect size of published articles on predictive validity of UTME was significantly different from unpublished studies or articles. Although, there were fewer unpublished articles, yet there was a sizeable increase in magnitude of correlation coefficient r. The published studies had correlation coefficient r of 0.311 while the unpublished studies had correlation coefficient r of 0.3951. The difference of 0.0811 in their effect sizes was significant and could be attributed to the fact that higher number of studies had significant findings in the unpublished articles. The implication of this finding is that, the study revealed no evidence of publication bias.

The finding on the second research question revealed that the effect size of published articles on predictive validity of UTME was significantly different from unpublished studies or articles. This study was in agreement with the work of Ale (2015) who collected correlation coefficients from published and unpublished literature, he produced evidence of a selective publication effect in his meta-analysis of the relations between social economic status and achievement are weaker in dissertations than in journals.

Conclusion

The study has been able to provide more scientific facts as regards the predictive validity of UTME in public University in Nigeria using a meta analytical approach. The meta analysis used has provided relevant facts on the psychometric worth and usefulness of the UTME examination in Nigeria. With the application of this statistical method, this study has been able to provide reliable database for which other researchers can draw from.

Based on the findings of this research, it was concluded that effect size contributed a reasonable difference in the magnitude of the selected study. It was also revealed that there is an indication that there is influence of linear trend in terms of effect size across this set of studies on predictive validity of UTME.

Consequent upon the outcome of this study, it was recommended that sample size on predictive validity of UTME should not use less than 120 in other to reduce the influence of effect size.

References

Adebayo, L. (2011). "A Note on Post-JAMB Tests," *The Guardian*, p.19, Sunday, November 27, 2011.

Afolabi, A. O. Mabayoje, V. O. Togu, V. A. Oyadeyi, A. S & Raji,Y. (2007). The effect of Mode of Entry into Medical School on Performance in the First Two Years. *Journal of medical science.* 7(6): 1021-1026

Ale V. M. (2015). Resolving Contradictions of Predictive Validity of University Matriculation Examinations in Nigeria: A Meta – Analysis Approach. *Journal of Education and Practice* www.iiste.org ISSN 2222-1735 (Paper) ISSN 2222-288X (Online) Vol.6, No.29, 2015

Bala D. & et al (2019). Are UTME and PUTME Good Predictors of Students' Academic Performance in the University? The Case of Kaduna State University, Kaduna, Nigeria. *Science World Journal Vol 14(No 2) 2019* www.scienceworldjournal.org *ISSN 1597-6343 Published by Faculty of Science, Kaduna State University*

Bernardine, N. N. (2019). The Prediction of Student Performance in Chemistry-based Courses in Public Universities Using University Matriculation Entrance Scores in Chemistry, Enugu State University of Science and Technology, Nigeria.

Biman, K. A., John S., Kawu W. and Fatima M. J. (2019). Entry Qualifications as Predictor of Undergraduate Biology Students' Academic Achievement in Nigerian Universities. Moddibo Adama University of Technology, Yola.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences.* New York: Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. New York, NY: Psychology Press.

Cumming, G. & Calin-Jageman, R. (2017) *Introduction to the New Statistics: Estimation, Open Science, and Beyond.* New York, NY: Routledge.

Denson, N. & Seltzer, M. H. (2011). Meta-analysis in higher education: An illustrative example using hierarchical linear modeling. Research in Higher Education , 52(3), 215-244.

Egger, M., Davey Smith, G. & Altman, D. G. (2000). Systematic Reviews in Health Care: Meta-analysis in Context. London: BMJ Books

Glass, G.V., McGaw, B., & Smith, M.L. (1981). *Meta-Analysis in social research.* Beverly Hills, CA: Sage.

Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5, 3-8.*

Imasuen K. & Stanley O. E. (2020). Unified Tertiary Matriculation Examination (UTME) and the Post Unified Tertiary Matriculation Examination (PUTME) as Predictors of Undergraduate Students' Final Grades.

Ituma, S. C., Ugwuanyi, C. C. & Uzochukwu S. U. (2023). Validity of UTME and Post-UTME in Predicting Students' Academic Achievement in Undergraduate Mathematics Department of Science Education, Faculty of Education, University of Nigeria Nsukka. *International Journal of Research and Innovation in Social Science (IJRISS)*

Kelley, K. & Preacher, K. J. (2012). On effect size. *Psychol. Methods* 17:137. doi: 10.1037/a0028086 PubMed Abstract | CrossRef Full Text | Google Scholar

Rosenthal, R. (1984). *Meta-analytic procedures for social research.* Beverly Hills, CA: Sage.

Russo, M. W. (2007). *How to Review A Meta-Analysis*, Gastroenterology & Hepatology , 3(8): 637-642. Retrieved: *http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3099299/*

Schmidt, F.L. (1984). *Meta-analysis: Implications for Cumulative Knowledge in the Behavioral and Social Sciences.* Paper presented at the annual meetings of the American Psychological Association, August, Toronto, Ontario.