Application of Standard Scores in Assessing Course Performance of Student in College Probation Examination

Queensoap Memory*, Azor Promise A, Oguara Richard, & Owaba Charles O. E
Department of Biostatistics,
School of Foundation Studies,
Bayelsa State College of Health Technology,
Otuogidi, Ogbia-Town
mqueensoap@gmail.com, memory2011.mq@gmail.com

Dogitimiye Memory & Sagbeyefa W.Gbanaiweri
Department of Health Information Management Technology,
School of Allied Medical Studies,
Bayelsa State College of Health Technology,
Otuogidi, Ogbia-Town

Abstract
The study critically applied standard scores in assessing course performance of students in College Probation Examination in Bayelsa State College of Health Technology. The study covered all the various programs of study in the College, namely Junior Community Health Extension Worker (JCHEW), Community Health Extension Workers (CHEW), Pharmacy Technician, Medical Laboratory Technician (MLT), Medical Laboratory Assistants (MLA), Medical Social Work, Dental Health Surgery Technician, Environmental Health Technology (ND), Environmental Health Technician (EVT), Environmental Health Assistants (EHA) and Health Information Management Technician (HIMT). The sample size of 600 students was used which also formed the population of the study. The instrument used was a valid and reliable record of results of all examinees according to performance with raw scores and percent average scores. Two research questions and one null hypothesis were formulated to guide the study. A descriptive-archival research design was adopted. Again, a paired sample t-test was used to test the null hypothesis at .05,alpha level of significance which on analysis indicated sample correlation of .390 (P=.340), t=1.58, degrees of freedom = 7 and a P=.158 showing that the null hypothesis was not rejected. Also, Z-score and T-score were used to convert the raw scores and were observed that it was chemistry that the students perform best instead of citizenship education while they perform worst in mathematics instead of the History and Philosophy of Science as presented by the raw scores. The study concluded that standard scores present student performance better than raw scores or average percent scores. Hence, it was recommended that management should adopt the use of standard scores in presenting students reports especially when students' performances are to be compared for placement, promotion, award etc.

Key words: *Standard Scores, Probation Examination, Raw Scores, T-Scores, Percent-Correct Scores.*

1. Introduction

College probation examination is an examination conducted in most of Colleges of Health Technology in Nigeria, purposely to admit competent students into the Institution. The performance of the students is presented with raw scores and percent – correct scores. According to Tan and Michel (2011), raw score is the total number of score points a test taker

or an examinee obtains by answering questions correctly on a test reexamination while a percent-correct score denotes the percentage of questions a test taker or an examinee answered correctly on a test or examination. For example, an examinee answered correct 30 out of 60 multiple choice questions, the test taker's raw score is 30 marks if each item is graded 1 mark each while the percent – correct score becomes 50%.

Consequently, examiners used such scores to assess performance among students and get them admitted if any meets up with the criterion score. Most worrisome, the best students among those that attempted the examination, which is made up of several courses, is assessed by percent – average score from the raw scores. In most cases such student went home with awards or scholarship. Is this comparison fair? Knowing that without standardized scores, it is difficult to make comparisons. A raw score of 30 on one test and a raw score of 125 on another test do not have much meaning until we know where each score is in relations to the mean (Siegle, n.d). Siegle further stated that it is only when we know how many standard deviations each score is above or below the mean that we can compare the two performances. The author concluded that standard scores allow us to make comparisons of raw scores that come from different sources.

Meanwhile, the use of raw scores and percent-correct scores to assess student's performance is something common among educationists, health workers and examiners of various fields. Parents compare their wards results with others based on raw score interpretation. Government and non-governmental organization likewise make comparison based on raw interpretation and most at times grant prizes for best performance in a particular course or all round performance. The College of Health Technology is not in exception to this practice. Therefore, the researcher is bordered to know who actually done well in the Probation Examination? Though, Hijazi and Naqvi (2005) stated that measuring of academic performance of student are product of socio-economic, psychological and environmental factors, the researcher in a purposeful manner carried out the application of standard scores in assessing course performance of students in College Probation Examination.

According to Cross (1995), a standard score is the number of standard deviations the number-right or percentage score is above or below average. "The Woodcock Johnson 111 Diagnostic Reading Battery" presented that the major purpose of standard scores is to place scores for any individual on any variable having any mean and standard deviation on the same standard scale so that comparison can be made. In a nutshell, a standard score is another way to compare a student's performance to that of the representative sample.

A common question asked by policy maker's teachers, parents and some test-takers is: how well did a student perform in his/her course work compared to others in the class and who did the best in the test or examination or question? Which course is most simple or difficult? Most commonly, answers are giving to such questions based on the use of raw scores or percent-correct scores which are usually unhelpful for the purpose of comparing performance among test takers or examinees ("formal assessment"). Among examiners and policy makers, questions have been asked why is it not appropriate to use raw or percent-correct scores for comparing test takers across different test editions (Tan & Michel, 2011). In the midst of these questions, raw scores still used in the socio-educational sector, health sector etc. and most comparison are made for the purpose of placement, scholarship awards, promotion, admission, honors etc. thus the researcher in this study was to apply standard scores in assessing course performance of probation students in College of Health Technology which will enable the management make valid comparison among examinees who participated in the examination against the background of the study.

2. Literature Review

According to Tan and Michael (2011), a raw score is the total number of score point a test taker obtains by answering questions correctly on a test while a percentage – correct score represent the percentage of questions of a test taker answers correctly on a test. These are scores that are usually adopted in class room situations. Typically, for example, student A, B, C, sat for an examination on the subject biology, mathematics and English. The examiner scores them as given below; student "A" scores 60, 70 and 65 respectively, student B scored 65, 80 and 70 respectively, student scored 70, 90 and 80 respectively. These are percent-correct scores. While 20 correct answers means raw scores of 20.

Consequently, according to Ukwuije (2003) raw or percent- correct scores cannot provide answers to same question such as which student had the best over raw performances, or which test did the different students do best? Ukwuije emphasized that raw score have little or no meaning. This is so because the raw scores for each subject have different meaning and standard deviation. Similarly, Schoen and Asley (2006) presented that a raw score or number correct score, on any test has limited inherent meaning. For this reason the raw scores of most test are converted to a standard scores (SS) scale. They said, standard scores scales provide scores that are more directly interpretable than raw scores. Furthermore, raw scores or percent-correct scores have a grave implication in test that have multiple forms but are of similar difficulty index.

Tan and Michael (2011) argued that, it is hard to use the percent correct score for fair comparisons of test takers performance in different forms of the same test. For example, the authors presented that getting 50% current in a hard form of test may mean the test taker has more knowledge and skill than another test taker getting 65% current on a relatively easier form of test. They pointed out that for this same reason, the raw scores cannot be used to compare test takers performance on different forms. When two test taker get the same raw scores on two different forms, the test taker who took the more difficult forms has demonstrated a higher level of performance than the test taker who took the relatively easier form, this, in most times, is over looked and student who anything whatever not to be considered as near all best is given reward as such. In essence, before comparison should be done scores on different forms of test should indicate the same level of performance no matter which form the test taker received. For Ukwuije (2003), scores should be made equally comparable, that is, all the means and standard deviations have to be reduced to the same level. This is achieved by converting the raw scores or percent-correct scores to standard scores such as z-scores, T- scores, K-scores or other standard scores.

Logston (2014) defined standard scores as scaled scores that are used in norm-referenced assessment on a test to the performance of other students' scores. Standard scores estimate whether students are abused accurate, average or below average, compared to peers. They also enable comparison of students' scores in different types of tests (Logston, 2014). The standard score is a very useful statistics because it allows us to calculate the probability of a score occurring within normal distribution (Lund Research, 2013). The paper argued that in a situation where a tutor set a piece of test in a particular subject, say mathematics, 50 student in his class. Having looked at the performance of the tutor's class, one student, Sarah, has asked the tutor, if by scoring 70 out of 100, she has done well. Bearing in mind that the mean score was 60 out of 100 and that Sarah scored 70, then at first sight it may appear that since Sarah has scored 10 marks above the average mark, she has achieved one of the self-marks. However, this does not take into consideration the (variation in scores amongst the 50 students in other; the standard deviation). After all if the standard deviation is 15, then there is a reasonable amount of variation amongst the scores when compared with the mean whilst Sarah has scored much higher than the mean score, she has not necessarily achieved one of the best marks in her class. The question arises: How well did Sarah perform in her

mathematics test compared to the other 50 students? It is obvious that one may quickly response that the raw scores should be used however Fund Research (2013) advocated that the best approach is to use standard normal distribution and its related Z-scores to answer the question above. Tan and Michael (2011) stated that the utility of standard scores comes from allowing for meaningful scores interpretations and, at the same time, minimization misinterpretation and in appropriate inferences. Meanwhile, Schoen and Ansley, (2006) presented that standard scores provide scores that are more directly interpretable than scores, that many types of standard scores are used in practice.

For equity and clear understanding if students' scores, Flanayam and Cattabiano (2004) stressed that when a student takes either an individual or group-administered test at school, the results are made available to both parents and teachers. It is important that parents and teachers understanding the meaning of scores that comes from tests or examination. They said most educational tests that are based on a scale that has a statistical mean of 100. If a student earns a standard score that is less than 100, then that student is said to have perform below average. Unlike the circular universities and Colleges in Nigeria, Bayelsa State College of Health Technology bases her admission of students in three facets namely WAEC / NECO (SSCE), Entrance Examination and Probation Examination which all students must meet up to the standard set up. These admission requirements are peculiar because of the responsibility of the College to train middle levels man power for the grass root implementation of Primary Health Care delivery system. The middle manpower expected to be trained at this level are professional in successful graduation in areas of Medical Laboratory Science, Community Health Science, Environmental Health Science, Pharmaceutical Technician Studies, Health Information Management Technology , Dental Health Sciences and Medical Social Work as approved by the College Academic Board.

Methods

A descriptive study was conducted by using an archival research design. The choice of this design was informed by the nature of data required for the study. The population of the study was the probation students of 2014/2015 session in Bayelsa State College of Health Technology Otuogidi, Ogbia Town which total 600 students that cuts across the various programmes of study such as Junior community Health Extension Workers, Community Health Extension workers, Pharmacy Technician, Medical Laboratory Technician, Medical Laboratory Assistants, Medical Social works, Dental Health Technicians, Environmental Health Technician, Environmental Health Assistants, and Health Information Management |Technicians. No sampling technique was adopted rather census, that is, counting all students who wrote the examination was done hence the sample size became 600.

The instrument for data collection was record of students' results published by the Management of the College. The record showed individual raw scores and percent-correct per course in the Probation Examination. This result was presented according to the various programmes of choice by the students. The demographic data were obtained from records available to the researcher from the exams and record department of the College Registry.

The instrument used for the study was a valid and reliable record because it was not designed by anybody which, had passed through the College Academic Board before making it public for students to see their performance and used as a criteria for admission. Based on the design of the study, the data for the study were obtained from records/documentary source. To this, the researchers sought for permission from the College Management to release photocopies of results of all probation students per programme of studies. This was granted and the document was released to the researchers and data were organized for further statistical analysis. The research questions were analyzed with mean, standard deviation and standard scores such as Z-score and T-scores while the null hypothesis was tested with paired sample

t-test at .05, level of significance. The decision was $P<.05$, null is rejected hence there was significance. Scores were assumed of a normal distribution after presenting scores with histogram. The test statistics was carried out with Statistical Package for Social Sciences (SPSS) version 20.

1. RESULTS/DISCUSSION

Table 1.Presentation of Demographic Information

S/N	VARIABLE	FREQUENCY	PERCENTAGE
1.	PROGRAMMES OF STUDY		
	CHEW	203	33.83
	JCHEW	29	4.83
	EHT (ND)	83	13.83
	EVT	40	6.67
	EHA	6	1
	MSW	3	0.5
	DST	15	2.5
	MLT	103	17.17
	MLA	2	0.34
	PHARM TECH	47	7.83
	HIMT	69	11.5
TOTAL	600	600	100.00
2	GENDER		
	MALE	143	23.83
	FEMALE	457	76.17
TOTAL	600	600	100.00

Table 1`above showed the demographic information of those who registered duly for the program. Looking at table 1 indicated that the Community Health Extension Workers program had the highest admission (33.83) into the probation program while Medical laboratory Assistant program had the least admission (.34) into the program. Also, table 1 pointed out that more females were screened into the program than male, 76.17:23.83 for female and male ratio respectively.

4.1 Research Question 1. Which student did best in the probation examination?

Table 2 Comparison Based on Ranks Based on Average Raw and T-Scores

S/N	STUDENT	AVERAGE RAW SCORES	RANK	AVERAGE T SCORES	RANK
1	A	93.00	1th	60.75	1th
	B	91.00	2th	60.25	2th
	C	91.00	3th	60.25	3th
	D	90.25	4th	60.06	4th
	E	90.00	5th	59.81	6th
	F	90.00	6th	59.94	5th
	G	88.75	7th	59.75	7th
	H	88.50	8th	59.5	10th
	I	88.4	9th	59.62	8th
	J	88.00	10th	59.62	9th
	K	88.00	11th	59.5	11th
	L	87.75	12th	59.44	12th

M	87.50	13th	59.25	13th
N	87.50	14th	56.06	14th
O	87.00	15th	59.3	15th
P	86.50	16th	58.92	20th
Q	86.50	17th	59.19	16th
R	86.50	18th	59.11	17th
S	86.25	19th	59.11	18th
T	86.25	20th	58.99	19th

Table 2 showed that student A had an average raw score of 93.00 and an average T-score of 60.75 placing him/her in the 1st rank. Table 2 also showed that student E who had 90.00 and 59.81 as average raw scores and T- score respectively had position 5th and 6th on the average raw scores and T-score. Similarly, positions of student H, P, R, S and T changes by average T-scores.

4.2 Research Question 2 which course did students did best or worst?

Table 3 Comparison Based on Course Performance with Average row scores and Average T-score

S/N	COURSE	AVERAGE RAW SCORES	AVERAGE T-SCORES
1	MATHS	45.3	33.89
2	PSYCHOLOGY	56.9	50.13
3	PHYSICS	53.32	50.04
4	HISTORY	44.57	50.01
5	CHEMISTRY	54	50.14
6	ENGLISH	52.5	50.06
7	BIOLOGY	44.95	50.13
8	CITIZEN EDUCATION	62.9	49.71

Table 3 showed that citizenship education (CE) having 62.9 as average raw scores while history and philosophy of science (history) had an average raw scores of 44.57 implying that student perform best in citizenship Education and performs worst in History and Philosophy of science. Table 3 also presented on conversion of raw scores to T-scores, chemistry had 50.14 while students performs best in chemistry and worst in mathematics as against citizenship education and history of the philosophy of science presented in their average raw scores as best and worst performed course respectively.

4.3 Null Hypothesis: there is no significance difference in students' performance of different courses in conversion of student's raw scores to standard scores.

Table 4 Paired Samples Statistics

PAIR	MEAN	N	SD	STD ERROR OF MEAN
VAR OOOO1	51.80	8	6.53	2.31
VAR OOOO2	48.01	8	5.71	2.02

Table 4 above showed Var 00001 (Average raw scores) having a mean of 51.80 standard deviation of 6.53 and standard Error of mean of 2.31 while Var 00002 (Average T-scores) has a mean of 48.01, standard deviation of 5.11 and a standard Error mean of 2.02. This implies that the average raw scores showed higher means than the Average T-scores. And the variability among students in the various course performances was higher with Average raw scores than Average T-scores.

Table 5 Paired Samples Correlation

PAIR	N	Correlation	Sig
Var00001 vs Var00002	8	.390	.340

Table 5 above showed the paired correlation of average scores (var 00001) and average T-scores (var 00002). Average raw scores was having a positive correlation of 390 and a *P* = .340 which is greater than .05 alpha. This implies that the two scores have no statistically significant correlation.

Table 6: Paired Sample Test

Pair	X-	SD	Std Error X^-	T	Df	Sig(2-Tailed)
Var 00001 – var 0002	3.79	6.79	2.40	1.58	7	.158

Table 6 showed a paired samples test which presents a mean of 3.79, standard deviation of 6.79, t of 1.58, degree of freedom 7 and *P* = .158 which is greater than .05. This implies that the null hypothesis of students' performance of different courses in conversion of students raw series to standard scores.

This current study was able to assess the performance of 600 candidates who took a probation examination in Bayelsa State College of Health Technology. The scores were presented with average raw scores and were presented with average raw scores and were admitted based in their performance. The study converted the raw scores of the T-score formula. On the conversion scale from raw score, Via Z-score to T-score enabled the researcher to discover that student (A) who scored 93.00% was 60.75 on T-score. Meanwhile, Ukwuje (2003), Tan and Michel (2001) and Schern and Asley (2006) supported the finding that raw score on any test has a limited inherent meaning. That is to say, raw score does not actually tell the performance of student on a particular test.

Moreover, it becomes more worrisome when such candidates are being compared with average raw score. This finding on assessment identified that student E who was ranked 5th on the raw score. Score table was ranked 6th on T-score table Ukwuije (2003) stated that raw score provide answers to questions such as which student had the best overall performance? Or on which test did student do best or worst? The study findings confirmed that students performed best chemistry instead of psychology as presented on raw score average (see table 3). Meanwhile, the worst performed course was Biology (X- = 44-95) when presented with raw score. However, on conversion to T-score, the worst performed course was mathematics (T-score average = 33.89). This finding was confirmed in the works of Lund Research (2013), Logston (2014). Obviously, the study identified that there was no significance difference between the average raw score and T-score and T-score average on different courses. One thing about T-score, it places students on equal variability before companion. When students' scores are presented on equal variability, actual best performance is easily identified. This study, therefore, was in line with Schoen and Ansley (2006) that standard scores are more directly interpretable than raw score. This is parameter when a student takes either an individual or group administered test, the results are made available to both parents and teachers.

2. CONCLUSION

Based on the findings of the study, the study concluded that:

- Standard scores present students' performance better than the use of raw score.

- Students' performance can best be compared by the use of standard scores than raw scores.

3. RECOMMENDATION

Based on the conclusions and findings, the study recommended that:

- Management should be looked into the use of standard scores in presenting results for admission.
- Before award should be given to the best students, such scores be converted to standard scores.
- A seminar, workshop and symposium should be organized to train staff dealing with results of students.

REFERENCES

Anees, A. (2013). A study of academic achievement in relation to intelligence of class vii students. *Excellence International Journal of Education and Research. 1(3).p239-248.*

Anonymous (n.d). Understanding standardized test scores. Retrieved February 16, 2005 from www.geneseo.edu/~balajthy/coursework/educ512/formaltesting.html.

Cross, L. H. (1995). Grading student. *Practical Assessment, Research and Evaluation, 4(8).* Retrieved February 16, 2015 from http://pareonline.net/getun-asp? V= 48n = 8.

Flanagan, D. P. & Caltabiano, L. F. (2004). Test scores: A guide to understanding and using test results. *Helping children at home and school II: Handouts for Families and Educators. National Association of School of Psychologist. Bethesda, MD 20814-(301)657-0270. Retrieved on February 16, 2015 from www.nasponline.org/communications/spawareness/testscores.pdf.52-81.*

Glutting, J.J. (2002). Glutting's guide for norm-referenced test score interpretation, using a sample psychological report. *Retrieved on November 19, 2014 from c://users/test/downloads/HANDOUT FOR NORM-REFERENCE TEST SCORE INTERPRETATION.htlm.*

Hijazi, S. T. & Naqvi, R. S. S. M. (2006). Factors affecting students' performance: A case of private colleges. *Bangladesh e-Journal of Sociology 3(1)1-10*

Koretz, D. (2000). The impact of score differences on the admission of minority student: An illustration. *National Board on Education Testing and Public Policy.1(5):149-181. Retrieved on February 16, 2015 from www.bc.edu/research/nbetpp/publications/vins.html.*

Kyoshaba, M. (2009). *Factors affecting academic performance of undergraduate students at Uganda Christian University*. A master degree dissertation on educational management of Makerere University.

Lund Research (2013). Standard score. Laerd statistics. Retrieved on February 16, 2015 from https://statistics.laerd.com/statistical-guides/standard-score-4.php

Scheon L. H. & Ansley T. N. (2006). Manual for test use, interpretation, and technical support. *IOWA Testing programs: 1-18.*

Siegle, D.(n.d). Standard scores. Retreived on February 16, 2015 from www.gifted.ucom.ed/.../research/

Ukwuije R. P. I. (2003). *Peanuts educational statistics.* Port Harcourt. Celwil Nigeria.

Woodcock Johnson 111 Diagnostic Reading Battery (n.d). Scoring and interpretation of the woodcock Johnson III Diagnostic Reading Battery. WJIIPic.jpg. Retreived on February 16, 2015 from www.genoseoredu/inbelonglly/courceswork/educ512/rormaltesting-html retrieved 16/2/2015