Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.8, No.10, 2017

Multilevel Assessment of the Predictive Validity of Teacher Made Tests in the Zimbabwean Primary Education Sector

Zadzisai Machingambi, M.ED (Psy), BSC, Hons (Psy) & B.ED (Admin)
Faculty of Applied Social Sciences, Zimbabwe Open University, P O box 1210, 68 Hellet Street Masvingo, Zimbabwe

Abstract
The principal focus of this study was to undertake a multilevel assessment of the predictive validity of teacher made tests in the Zimbabwean primary education sector. A correlational research design was adopted for the study, mainly to allow for statistical treatment of data and subsequent classical hypotheses testing using the spearman's rho. The variables that provided the bedrock for the study were aggregate test scores for pupils' performance in:

i. Midyear tests and
ii. End of year tests, in 2016.

The four null hypotheses that underpinned the study were tested on the basis of correlation coefficients computed using test scores generated through the aforementioned tests. After subjecting the hypotheses to the court of empirical evidence, through significance testing, some depictable results emerged. The major finding was that homogeneous results were observed, as all the four null hypotheses upon which the study was based were rejected. The major finding attested to the conclusion that teacher made tests in the Zimbabwean primary education sector were valid for prediction purposes. The study also observed that if tests being appraised were found to be valid, then the same tests were also reliable. In the light of the foregoing, recommendations on the broad applications of teacher made tests in educational programming were promulgated
Keywords: Predictive validity, Multilevel Assessment, Predictor, Predicand, Criterion, Null Hypothesis, Significance Testing, Educational Programming.

1. Background to the Study

The concept of predictive validity has a fairly long history. It is traceable to as early as A.D 200 when written tests became evident in China. However the scientific study of the predictive validity of educational tests was pioneered in the 19th century in the western developed world. From then onwards, research studies in this area appeared to have been much more developed and more frequent in western countries than in developing countries. The studies yielded data that led to the development of a thorough and perceptive knowledge of a variety of test validation procedures. The principal validation procedures were tailored to establish inter alia: content validity, concurrent validity, construct validity and lastly predictive validity which was the focal point of this investigation, (Kubiszyn and Borich, 1993). A somewhat different state concerning research on the predictive validity of educational tests was found to obtain in the developing world where research in this area was at its nascent stage.

Early conceptions of validity were phrased exclusively in predictive terms. However the different forms of assessing the validity of educational tests took centre stage at different epochs for example content validity has been dominant in the past 20 years. From the Zimbabwean perspective fairly interesting developments were evident. Before the advent of independence in 1980, there was a measurement and evaluation unit (MEU) under the then department of Native education (DNE). This unit administered scholastic aptitude tests, inter alia. These tests were meant to predict students' scholastic potential in mathematics and English as basis for making instructional, selection, placement, screening and other important educational decisions. Determination of predictive validity in Zimbabwe and elsewhere in the world was done by comparing test scores. Comparing scores was able to show for example that pupils who had high standing on the prediction test had a high standing on a criterion test (Gronlund, 1995). It was found that if a consistent pattern of students' performance in the two tests, was discernible, then the test being validated had high predictive validity. However it became evident that these procedures were too rudimentary, yielded unimpressive data and less informative. Thus this study involved the calculation of validity coefficients, which give relatively more accurate and concise information on educational measurements (Guilford and Fruchter, 1981).

On the Zimbabwean educational landscape, systematic research on the predictive validity of tests seemed to be relatively inadequate. Only a few investigations on the validity of tests have been attempted. Furthermore studies on the predictive validity of educational tests tended to be conducted by government sponsored commissions and not by scholars. In this respect the system wide approach adopted appealed more to policy makers in their quest to improve the quality of the whole system of education. This tended to overshadow any ingenious attempts at developing research perspectives by individual scholars, whose utility can be manifested at specific levels of the system of education. In the light of the foregoing the researcher found it

inevitably compelling to engage in the assessment of the predictive validity of teacher made tests in Zimbabwe.

2. Statement of the Problem

At all levels of the Zimbabwean primary education sector many critical students selection, placement, curriculum, educational counselling, instructional and administrative decisions are made on the basis of the teacher made tests. Thus this study attempts to ascertain the predictive validity of these tests. The quality of educational decisions is inextricably dependent on the validity of teacher made tests, inter alia, hence the need for this investigation.

3. Conceptual Framework

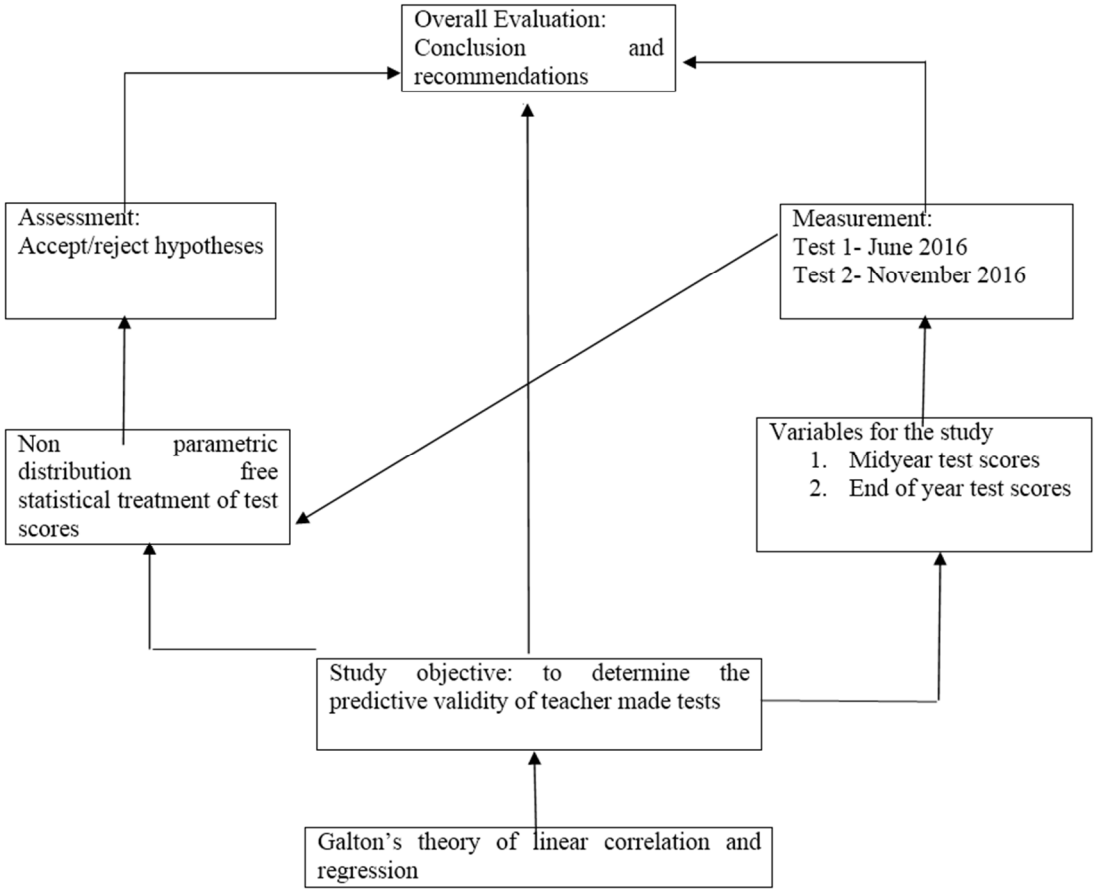
Teacher made tests in Zimbabwe, are the tests generated by teachers themselves for a variety of applications in the classroom and in school settings. Teacher made tests fall under the broad category of educational tests. Many authors agree that a test is a measuring device, procedure or sample behavior that tells us something and not everything about some class of behavior (Gronlund and Linn, 1995; Gay, 1980; Cohen and Swerdlik, 2005). Teacher made tests in Zimbabwe are also referred to as classroom tests. Teacher made tests differ markedly with standardized tests in terms of universality and scope of application. Whereas teacher made tests (TMT) are criterion referenced and used basically for formative evaluation, standardized tests are norm-referenced and principally used for summative evaluation particularly at the end of a cycle of educational experiences, for example primary education, (Mpofu,1990).

Gronlund and Linn (1995) observed that regardless of the type of assessment used or how the results are used, all assessments should possess certain characteristics; the most essential of these is validity. This preposition serves to underpin the significance of this study, and draws us to the concept of validity in particular. Validity in general is concerned with how well a test instrument measures what it is designed to measure (Ormrod, 2000).

According to Thondike and Hagen (1997) a test is only valid for a specific purpose. To this effect this study only sought to establish the predictive validity of teacher made tests in the Zimbabwean educational sector. Many authors (Gay, 1980; Thorndike and Hagen, 1997; Kubyszn and Borich, 2000; and Gronlund, 1985) regard the major function of predictive validity when applied to psychological testing and assessment as that of predicting an individual's performance on some subsequent criteria. Ormrod (2000) highlights four basic assumptions that underlie any predictive validity study, and indeed that guided this investigation. The first precondition is that at least two variables must exist for prediction to be possible. Secondly, there must be a time lapse between the application of the test being validated and the criterion measure. The correlation coefficient obtained is likely to be influenced by factors such as intervening learning between measurements and improvement in maturation levels, reliability of the instruments used and the distribution of scores.

The variables that were treated in this study were the aggregate marks for the four examinable subject areas: Mathematics, Shona, English and General Paper. The aggregate marks for the subsamples of twenty pupils were collected in June and November 2016 respectively and subsequently correlated. The four subsamples were collected in four districts of Masvingo Province. The four rural primary schools selected were typical of most rural primary schools across Zimbabwe. The diagram below attempts to place the current study into a clear theoretical and conceptual perspective.

Diagramatic illustration of the theoretical and conceptual framework that guided the study



4. Hypotheses

Some hypotheses were considered on the basis of the main problem for the study. These have been stated below in null form for the purposes of statistical treatment of data.

$\boldsymbol{Ho_1}$: There is no significant relationship between grade two pupils' performance in midyear tests and end of year tests.

$\boldsymbol{Ho_2}$: Grade four pupils' midyear test scores do not correlate significantly with end of year test scores for the same pupils.

$\boldsymbol{Ho_3}$: The correlation coefficient for primary grade five pupils' performance in midyear tests and end of year tests is not significant.

$\boldsymbol{Ho_4}$: There is no significant relationship between primary pupils' performance in midyear tests and end of year tests at grade six level.

5. Significance of the Study

This study was considered significant in several respects. It attempted to assess the predictive validity of teacher made tests in the Zimbabwean education sector. Specifically the study sought to establish the predictive validity of teacher made midyear tests across four levels of the Zimbabwean primary education system. The study was considered significant because efforts to improve validity of tests inevitably result in an improvement of reliability of the tests in question (Omrod, 2000). However the other way round is not always correct, a reliable test is not necessarily valid (Guilford and Fruthcher, 1981).

The study was also significant from a methodological stand point. It applied an unusual technique of studying relationships between variables. Instead of employing the traditional methods of comparative analysis using mean scores or and percentages, this study adopted a correlational statistical design to allow for more precise and intense study of the phenomenon under investigation (Cohen and Swerdlik, 2005).

This study was considered important in the sense that it was confined to a specific sector and levels of education as opposed to studies which to focus on the whole system of education. Specifically it was beamed at grades 3, 4, 5 and 6 of the primary education sector. Consequently knowledge generated through the study was deemed critically useful to policy makers, school managers, teachers themselves, central government, multilateral agencies in general and the Zimbabwe school examination council (ZIMSEC) in particular. However measurement of predictive validity of teacher made tests was done within the context of the dynamics of the

education system that prevailed at the time of the study. Hence it is these critical considerations that underpinned the study.

6. Review of Related Lliterature

Before and after independence in 1980 no visible effort was apparent to measure the validity of teacher made tests in the Zimbabwean primary education sector (ZPES). As a result, systematic research in this area appeared to be missing. Hence in exploring related studies on the validity of teacher made tests, a recourse was made to studies conducted in other countries, mostly western countries were in research in this area seemed more developed.

After a series of studies, Gay (1980) reported on the graduate record examination that was used to select students for admission to graduate schools. The major assumption was that students obtaining a score of 1000 and above had a higher probability of succeeding in the graduate school. The study cited above differed from this study in that the time interval for prediction of success was fairly long. However the predictive function of the graduate record examination (GRE) and teacher made midyear tests in the current study were found to be essentially similar in that they provided useful insights into the possible performance of students in future criteria of success. It follows therefore that the graduate record examination (GRE) subsumed a wide range of faculties and lacked predictive validity in any specific areas. Similarly the aggregate scores that formed basis for computing correlation coefficients, were derived from a wide range of subject disciplines that constitute the primary education curriculum.

In 1973, Marjundar headed the National Council in New Dehli, India which undertook to test, a priori, and the predictive validity of the newly constructed mathematics creative test series that was being developed under the Department of Sciences for utilization in the national talent search scheme. The study relied on two tests, one which served as predictor and the other as the criterion. The predictor test was administered to group of 60 higher secondary school students in August 1973. The same group was retested in December of the same year and results were subsequently analysed and published. The findings revealed that the null hypotheses upon which the study was based were rejected since the mathematics creative test proved to be a good predictor of performance in the criterion test. Quite outstanding parallels are discernible between the two studies i.e the one described above and the current one. Firstly the time lapse and timing between measurement of the predictor and criterion variables was approximately the same i.e about six months stretching from midyear to end of year. Secondly and more importantly the purpose of the study was similar to the current study in that it sought to establish the predictive validity of educational tests. Lastly the two studies shared common ground in that they were premised on null hypotheses.

In the year that followed, Bennet, Seashore and Wesman (1974) as reported in Hopkins (1976) conducted an investigation in the United States of America to find out the extent to which a battery of differential aptitude tests (DATS) scores correlated with occupational and academic achievement. The study concluded that differential aptitude tests were good predictors of occupational success. However in yet another study MacNemer (1964) in Hopkins and Stanely (1981) illustrated that the superiority of differential aptitude tests for predicting academic success was small since there exists a great commonality among abilities required to succeed in most academic disciplines. The picture that emerged was that both the general aptitude tests (GAT) and graduate record examinations (GRE) tended to obscure prediction in specific areas of interest. This study shares the same limitation since in all instances overall performance was considered for both individual pupils and for groups.

MacGall (1977) set out to investigate the extent to which the childhood I.Qs predicted adult educational and occupational success. The findings were that I.Qs obtained for a sample of children between 3 and 18 years of age were found to be significant predictors of educational and occupational status at 26 years of age and older. The study also found that I.Qs obtained at the age of 5 correlated highly with adult I.Qs giving a correlation index of 0.5 or higher. The foregoing conclusions seem to concur with the findings of a study by MacNemer in 1964. The two studies observe that, there exists a great commonality among abilities required for academic and occupational success and intelligence is one such ultimately underlying factor. However the dilemma in Zimbabwe is that psychometric measurements of such aspects as I.Qs etc are not commonly carried out in the school life cycle of most children.

It was possible to conclude from these studies that factors that affect predictive validity of teacher made tests are varied. Quite interestingly, most prediction studies tended to produce results showing significant correlations between the test being validated and the criteria of success, (Torrance, 1972; Kelvin et al., 2008; Kinyua and Okunya, 2014). This study then sought to establish whether teacher made tests at primary school level were good predictors of pupils' performance in end of year examinations.

7. Methodology

The correlational research design was adopted for this study. It was deemed appropriate since it satisfies one of

the primary purposes of social sciences, that of discovering relationships amongst phenomena with the ultimate view to predicting and controlling their occurrence (Robson, 1993). Assessment of predictive validity was based on performance scores generated through teacher made tests. Consequently a questionnaire was designed to capture scores on pupils' performance. It was on the basis of these test scores that correlation coefficients were computed to determine if teacher made tests were valid for prediction purposes.

7.1Sample

The study was based on a sample of 80 students. This was further broken down into subsamples of 20 students per level. The four subsamples of 20 pupils were distributed among the four districts from Masvingo Province that formed the focal point of this investigation. The subsamples were generated randomly while the four districts and four schools were purposefully sampled for convenience and feasibility reasons. Both the teachers who set the teacher made tests and the students upon whom the performance was analysed were typical of teachers and pupils elsewhere in the country. Consequently results yielded by the study were basis for credible generalizations applicable to the Zimbabwean primary education sector.

7.2 Instrumentation

The major information gathering instrument was the questionnaire. Basically the questionnaire was in the form of a score sheet on which scores were entered. Scores entered were aggregate marks on pupils' performance in the four disciplines of the primary education curriculum. Two sets of scores were elicited for each pupil for performance on the predictor and criterion tests. The questionnaire in the form described above was used to collect information on four subsamples obtaining at only four levels of the primary education sector. Both the predictor and criterion tests were teacher made tests.

7.3 Data analysis

The variables that were treated in this study were aggregate marks for the four examinable faculties: Mathematics, English, General Paper and Shona. Two sets of aggregate marks were generated for each subgroup of 20 students sampled at each of the grade levels (3-6). Overall performance scores collected in June and November, 2016 were subsequently correlated using the spearman's rho computational formula:

$$\text{Rho} = 1 - \frac{6\sum D^2}{n(n^2-1)} \quad .$$

Thus null hypotheses were then tested at the 0, 05 level of significance using Guilford and Fruchter's table K. It was then possible to accept or reject the null hypotheses.

8. Major Findings of the Study

This section reports on the major findings of the study. Along the process outstanding observations based on interpretation of the findings are also unveiled.

8.1 Summary of findings on Ho_1 using spearman's (rho).

Table 1. Aggregate tests scores for pupils' performance in teacher made midyear and end of year tests at grade 3 level (N=20)

Student Number	Sex	Aggregate mark/midyear tests(predictor)	Aggregate Mark/end of year tests (predicand)
1	F	262	278
2	F	283	260
3	F	187	225
4	F	168	190
5	F	70	115
6	F	188	262
7	F	228	260
8	F	241	267
9	F	103	178
10	F	262	282
11	M	79	90
12	M	225	238
13	M	271	291
14	M	120	181
15	M	281	308
16	M	213	251
17	M	158	193
18	M	152	140
19	M	238	269
20	M	269	312
Possible mark		400	400

The data on table 1 above was used to calculate Spearman's correlation coefficient on table 2 as basis for testing the first null hypothesis (Ho_1) .

Table 2. Calculation of Spearman's rho to determine the predictive validity of teacher made tests at grade 3 level of the Zimbabwean Primary Education Sector.

X	Y	Rank(X)	Rank(Y)	D	D^2
262	278	16	16	0	0
283	260	20	12	+8	64
187	225	8	8	0	0
168	190	7	6	+1	1
70	115	1	2	-1	1
198	262	9	13	-4	16
228	260	12	12	0	0
241	267	14	14	0	0
103	178	3	4	-1	1
262	282	16	17	-1	1
79	90	2	1	+1	1
225	238	11	9	+2	4
271	291	18	18	0	0
120	181	4	5	-1	1
281	308	19	19	0	0
213	251	10	10	0	0
158	193	6	7	-1	1
152	140	5	3	+2	4
238	269	13	15	-2	4
269	312	17	20	-3	9
					$\sum D^2$ =108

Spearman's rho = $1 - \frac{6\sum D^2}{n(n^2-1)}$

rho = $1 - \frac{6\times108}{20(20^2-1)}$

Therefore rho = 0.92

Table 2 above depicts the calculation of the spearman's correlation coefficient (rho). The hypothesis that

stated that: there was no significant relationship between grade 3 pupils' performance in midyear and end of year tests was tested. The observed value of 0.92 was located within the critical region, since at the 0.05 confidence level and for 20 degrees of freedom a critical value of 0.377 was established. Hence the null hypothesis was rejected. In conclusion midyear teacher made tests were valid for prediction purposes at grade 3 level.

8.2 Summary of findings on Ho_2 using the Spearman's (rho)

Table 3.Aggregate test scores for pupils' performance in midyear and end of year tests at grade 4 level (N=20)

Student number	Sex	Aggregate mark/ midyear tests (predictor)	Aggregate mark/ end of year test (predicand)
1	F	215	209
2	F	258	217
3	F	250	216
4	F	249	245
5	F	228	208
6	F	318	296
7	F	271	264
8	F	218	178
9	F	301	285
10	F	344	310
11	M	314	295
12	M	236	219
13	M	193	211
14	M	200	196
15	M	159	150
16	M	242	190
17	M	250	219
18	M	196	196
19	M	224	244
20	M	88	96
POSSIBLE MARK		400	400

Data presented on table 3 above provided the basis for calculation of the Spearman's correlation coefficient as reflected on table 4 below. The second null hypothesis (Ho_2) was then tested.

Table 4. Calculation of the Spearman's (rho) to determine the predictive validity of teacher made midyear tests at grade 4 level.

X	Y	Rank(X)	Rank(Y)	D	D^2
215	209	6	8	-2	4
258	217	15	11	+4	16
250	216	14	10	+4	16
249	245	12	15	-3	9
228	208	9	7	+2	4
318	296	19	19	0	0
271	264	16	16	0	0
218	178	7	3	+4	16
301	285	17	17	0	0
344	310	20	20	0	0
314	295	18	18	0	0
236	219	10	13	-3	9
193	211	3	9	-6	36
260	195	5	5	0	0
159	150	2	2	0	0
242	190	11	4	+7	49
250	219	14	13	+1	1
196	196	4	6	-2	4
224	244	8	14	-6	36
88	96	1	1	0	0
					$\sum D^2$=200

Spearman's rho = $1-\frac{6\sum_D 2}{n(n^2-1)}$

rho = $1-\frac{6\times200}{20(20^2-1)}$

Therefore rho = 0.85

Table 4 above shows the computation of Spearman's correlation coefficient (rho). The null hypothesis

(Ho_2) that stated that: there was no statistically significant relationship between pupils' performance in midyear and end of year tests was tested. For 20 degrees of freedom and at the 0.05 alpha level, a critical value of 0.377 was established. Since the observed value of 0.85 is located in the critical zone, the null hypothesis was rejected. Conclusively midyear teacher made tests were found to be valid for predicting pupils' performance in end of year tests at grade 4 level.

8.3 Summary of findings on Ho_3 using Spearman's (rho).

Table 5. Aggregate test scores for pupils' performance in teacher made midyear and end of year tests at grade 5 level (N=20)

Student number	Sex	Aggregate mark/midyear tests(predictor)	Aggregate mark/ end of year tests(predicand)
1	M	163	217
2	M	191	235
3	M	183	247
4	M	212	259
5	M	251	303
6	M	159	206
7	M	180	208
8	M	171	179
9	M	160	198
10	F	151	188
11	F	182	234
12	F	158	187
13	F	245	294
14	F	211	221
15	F	188	235
16	F	202	261
17	F	183	221
18	F	159	216
19	F	161	214
20	F	174	260
POSSIBLE MARK		400	400

Table 5 above presents aggregate scores on pupils' performance in midyear tests and end of year tests written in June and November of the same year (2016) respectively. It was on the basis of statistical data in table 5, that the predictive correlation coefficient was computed using the Spearman's (rho) to find out whether the midyear tests were a good predictor of pupils' performance in end of year tests.

Table 6. Calculation of Spearman's (rho) to determine the predictive validity of teacher made midyear tests at grade 5 level.

X	Y	Rank(x)	Rank(y)	D	D^2
163	217	7	9	-2	
191	235	15	14	1	4
183	247	13	15	-2	1
212	259	18	16	2	4
251	303	20	20	0	4
159	206	4	5	1	0
180	208	10	6	4	1
171	179	8	1	7	16
160	198	3	4	-1	49
151	188	1	3	-2	1
182	234	11	12	-1	4
158	187	2	2	0	01
245	294	19	19	0	0
211	221	17	11	6	36
188	235	14	14	0	0
202	261	16	18	-2	4
183	221	13	11	2	4
159	216	4	8	-4	16
161	214	6	7	-1	1
174	260	9	17	-8	64
					$\sum D^2$=210

Spearman's rho = $1-\frac{6\sum D^2}{n(n^2-1)}$

rho = $1-\frac{6\times210}{20(20^2-1)}$

Therefore rho = 0.84

The third null hypothesis had predicted no relationship between pupils' performance in teacher made midyear tests and end of year tests at grade 5 level. Statistical data presented on table 6 above was used to run a test of significance on Ho_3. For 20 degrees of freedom and at the 0.05 probability level, a critical value of 0.377 was read off from the table of critical values. Consequently the null hypothesis Ho_3 was rejected. The observed (rho) of 0.84 is greater than the table value of 0.377, and located within the critical region. In conclusion teacher made tests administered to pupils at grade 5 level were a good predictor of the same pupils' performance in end of year tests.

8.4 Summary of findings on Ho_4 using Spearman's (rho)

Table 7. Aggregate test scores for pupils' performance in teacher made midyear and end of year tests at grade 6 level of the Zimbabwean Education System.

Student number	Sex	Aggregate mark/midyear tests(predictor)	Aggregate mark/end of year tests(predicant)
1	M	167	175
2	M	160	179
3	M	138	152
4	M	129	155
5	M	232	195
6	M	133	168
7	M	303	323
8	M	149	148
9	M	175	174
10	M	269	210
11	F	239	235
12	F	171	185
13	F	183	196
14	F	280	285
15	F	178	189
16	F	244	245
17	F	168	201
18	F	231	255
19	F	202	227
20	F	187	191
Possible mark		400	400

The statistical data presented on table 7 above was used to calculate the coefficient of correlation using Spearman's (rho).

Table 8. Computation of Spearman's coefficient of correlation (rho) to assess the predictive validity of teacher made tests at grade 6 level.

X	Y	Rank(x)	Rank(y)	D	D^2
167	175	6	6	0	0
160	179	5	7	-2	4
138	152	3	2	1	1
129	155	1	3	-2	4
232	195	15	11	4	16
133	168	2	4	-2	4
303	323	20	20	0	0
149	148	4	1	3	9
175	174	9	5	4	16
269	210	18	13	5	25
239	235	16	15	1	1
171	185	8	8	0	0
183	196	11	9	-8	64
280	285	19	18	1	1
178	189	10	9	1	1
244	245	17	16	1	1
168	201	7	12	-5	25
231	255	14	17	-3	9
203	227	13	14	-1	1
187	191	12	10	2	4
					$\sum D^2$=186

$$\text{Spearman's rho} = 1 - \frac{6\sum_D 2}{n(n^2-1)}$$

$$\text{rho} = 1 - \frac{6\times186}{7980}$$

$$\text{Therefore rho} = 0.86$$

Information on table 8 above was used to run a test of significance on the fourth and final null hypothesis Ho_4 that ruled out a statistically significant relationship between test scores obtained by grade 6 pupils for midyear and end of year tests. Thus the observed Spearman's rho of 0.86 is greater than the rho critical of 0.377 read off from the table of critical values at the 0.05 probability level and for 20 degrees of freedom. Since the observed Spearman's rho of 0.86 is located within the critical zone, the null hypothesis was rejected. By way of conclusion the teacher made tests at grade 6 level possessed a good degree of predictive validity.

9. Discussion of Major Findings

Information gleaned from related studies and data collected from the field were subsequently analysed to find out if any clear patterns emerged. Generally there was a serious concurrence of data on almost all major focus areas, as reflected in the discussion that follows.

The first hypothesis had ruled out any significant relationship between pupils' performance in teacher made midyear and end of year tests at grade 3 level. The major finding on this hypothesis revealed a close relationship between pupils' performance in the two tests, hence the null hypothesis was rejected. The outcome of the significance test on the first hypothesis concurred with Gay (1980), whose study proved the validity of the graduate record examination (GRE) in predicting a high probability of success in the graduate school.

The second hypothesis similarly predicated no significance relationship between pupils' performance in the predictor and criterion tests at grade 4 level of the Zimbabwean primary education sector. Once more the hypothesis was put to the court of empirical justice. The significance test disapproved the hypothesis. Consequently the second null hypothesis was rejected. As a result pupils' performance in the predictor tests was significantly correlated to their performance in the criterion tests. Quite outstanding parallels were discernible between research results of this study and findings of a study by Majundar (1973). The referred study ascertained the predictive validity of the newly constructed mathematics creative test series that was used in the national talent search in India.

The third hypothesis postulated that there was no significant relationship between pupils' performance in midyear and end of year tests at grade 5 level. Similarly the hypothesis was put to the rigour of empirical testing. The null hypothesis was subsequently rejected. The correlation coefficient of pupils' performance in the two tests was not only high but also located in the critical region, attesting to the validity of midyear tests in predicting pupils' performance in end of year tests.

Finally the last and fourth null hypothesis had dismissed any significant relationship between pupils' performance in midyear and end of year tests at grade 6 level. Results from the significance test run on the hypothesis depicted a close relationship between pupils' performance in the two tests. Accordingly the null hypothesis was rejected. It was concluded that teacher made tests were valid enough to predict pupils' performance in a future known criterion of success. The finding on this hypothesis was in tandem with results of studies conducted earlier on, on the subject of predictive validity (Bennet, Seashore and Wesman, 1974; Mc Gall, 1977 and MacNemer, 1964).

An inspection of findings on all the four hypotheses depicts that teacher made tests were valid for prediction purposes across all the sampled four grade levels of the Zimbabwean Primary Education Sector. Relative stability and consistency in pupils' performance in the two tests was observed. Lack of positive perfect correlation between the two sets of scores could be attributed to improvement of maturation levels of pupils and intervening learning during the period between the two measurements.

10. Summary, Conclusions and Recommendations

The focal point of this investigation was to carry out a multilevel assessment of the predictive validity of teacher made tests in the Zimbabwean Primary Education System. The correlational research design provided a guiding framework for the study. Subsamples of 20 students per grade were randomly generated from four districts of Masvingo Province. Given the similarities in teacher and pupil characteristics across the country, teachers and pupils who participated in the study were deemed representative enough to be basis for credible generalizations. Significance tests run on the four hypotheses depicted homogenous results, in that they were all rejected. The results provided a solid background for conclusions and recommendations.

Conclusions

- Teacher made tests in the Zimbabwean Primary Education sector were found to be valid for predicting pupils' performance in future known criteria of success.
- If teacher made tests possessed an element of predictive validity, then the corollary was also true, the same tests were generally reliable, notwithstanding the fact that the other way round is not true, a reliable test is not always valid.
- The study also concluded that factors that affect predictive validity of teacher made tests were varied.
- The high correlation coefficients observed in computations above attested to the relative stability and

consistency of test scores generated through the predictor and criterion tests.

- If teacher made tests possessed the quality of predictive validity, then they were amenable to wide range of educational applications as articulated in the recommendations propounded underneath.

Recommendations

On the basis of the above research findings and conclusions, some recommendations that had a bearing on educational programming were promulgated. Thus it was recommended that:

- Placement decisions on pupils' entry into next grades must be made conveniently earlier using data obtained from teacher made midyear tests.
- Critical decisions on referral and repetition cases could be finalized on the basis of midyear test results.
- Schools should develop standardized batteries of midyear and end of year tests.
- Given their proven validity, teacher made tests could be used for selection of pupils into resource units and special classes.
- Teacher made tests be used instead of the western developed (WRAT) tests for the performance Lag Address Programme (PLAP).
- Teachers' colleges and universities should make testing and evaluation a critical component of their curriculum.
- Teachers should receive occasional and systematic inservice training on tests construction to embrace a wide range of validation procedures.
- Educationists should be empowered to make more informed and meaningful interpretation of test outcomes using advanced statistical procedures.
- Terminal evaluation of students at the end of the Primary Education cycle (PEC) be localized to articulate the unique, environmental, cultural, technological, economic and developmental factors obtaining in different communities.
- Further studies should be conducted to assess the predictive and other forms of validity at different levels of Zimbabwean education system.

References

Bogdan, R.C and Biklen, S.K. (1997*). Qualitative Research for Education*.
Boston: Allyn and bacon
Bogdan, R.C. and Taylor. S. (1975) .*Introduction to qualitative research*.
New York john Wiley and sons.
Borg, W.R. and Gall, M.D. (2000). *Educational Research: An introduction. New York*:
David Mekay Company, Inc.
Cohen, L. and Manion, L. (1980). *Research Methods in Education*. London Routledge.
Cohen, R.J and Swerdlick, M.E (2005) *Psychological Testing and Assessment* (6thed).
New York:McGraw-Hill.
Furgurson, G.A. (1981*). Statistical Analysis in Psychology and Education* (5thed).
Auckland: MacGraw-Hill International Company.
Gay, L.R. (1980). *Educational evaluation and measurement*.
London: Merrill Publishing Company.
Gronlund, N.E. (1981). *Measurement and Evaluation in teaching*. New York: Macmillan Publishing Company.
Gronlund, N.E and Linn, R.L. (1995). *Measurement and Assessment in Teaching* (7thed). London: Prentice Hall, Inc.
Guilford, J.P and Fruchter, B (1981) *Fundamental Statistics in psychology and Education*. Singapore: McGraw-Hill.
Harper,W.M. (1991). Statistics. Singapore: Longman.
Hopkins,K.D and Stanely, J.C (1981). *Educational and psychological Measurement and Evaluation*. New Jersey: Prendice Hall, Inc, Englewood Cliffs.
Hopkins, C.D. (1976). *Educational Research Structure for inquiry*. Columbus: Merrill Publishing Company.
Kelvin, M., Frances, P. G., Glen, P.A., Michaell, s., Paul, H. L., and Jane, K. S. (2008). The Thorny Nature of Predictive Validity Studies on Screaning Tests For Developmental Behavioral Problems. *Pediatrics.* October, 2008, Volume 122/ Issue 4.
Kinyua, K. and Okunya, L. O. (2014). Validity and reliability of Teacher made Tests: Case Study of Year 11 Physics in Nyahururu District in Kenya. (2014*). African Educational Research Journal*. Vol 2 (2) pages 61-71, May 2014 ISSN 2354-2160
Kubiszyn, T.and Borich, G. (1983). *Educational Testing and Measurement: A Classroom Application and Practice*. New York: Harper Collins.

Leedy,P.D. (1980). *Research Planning and Design* (2nded). New York: Macmillan Publishing Company.
Majundar,S. K. (1973). *Predidion of Achievement in Mathematics on the Basis of Scientific Creative test Scores. London*: Commonwealth Secretariat.
Miles M.B. and Huber man, A.M. (1984). *Qualitative Data Analysis: Sourcebook of new methods*.London: Sage Publications.
Mpofu, E. (1991). *Testing for Teaching*. Harare Longman.
Omrod, J.E. (2000). *Educational Psychology: Developing Leaners* (3rded). New Jersey: Merril Publishing Company.
Popham, W J(1975). *Educational Evaluation* New Jersey: Prentice Hall .Inc.
Robson, C. (1993).*Real World Research*: Resource for Social Scientists and Practitioner Researchers. Massachussetts: Blackwell Publishers.
Thordike,R.L and Hagen, E(1977). *Measurement and evaluation* (4thed). London: John Wiley and sons.
Torrance, P. E. (1972). Predictive Validity of the Torrance Tests of Creative Thinking. *Journal of Creative Behaviour*. Volume 6, Issue 4, Dec 1972, pages: 236-262. Doi:10.1002/j.2162-6057.1972.tb00936.x
Vonks,J.H.C. (1984). *Teacher Education and Teacher Practice*.
Amsterdam Free University Press

Bibliography

Good, C.V. (1972*). Essentials of educational Research Methodology and Design* (2nded)Englewood cliffs: Prendice-Hall Inc
Krejcie,RV. And Morgan, D.W. (1970) .*Determining Sample Size for research activities*. Education and psychological Measurement, 30, 607-610.
Mussen, P.H. (1983).*A Handbook of Child Psychology* (4thed). New York: John Wiley and sons.
Nunally, J.C Bernstern, J. H. (1994*). Psychometric Theory* (3rded) New York: McGraw Hill, inc.
Rees. D. G. (1995*) Essential Statistics* (3rded). London: Chapman and Hall.
Tuckman, B.W. (1975).*Measuring Educational Outcomes: Fundamentals of Testing*.
New York: Harcourt Brace Jovanovich Inc.