

SELF-EVALUATION BY HIGH SCHOOL PUPILS

Dr. O.C. Nwana
Director, Curriculum-Development Centre, University of Nigeria

In the traditional setting, evaluation of the attainment of school pupils has been done exclusively by the teachers and public examining bodies. Hardly any opportunity has been given to pupils to evaluate themselves in a regular and deliberate manner. The purpose of this study was to find out whether or not high school pupils can and do assess themselves, the nature of the assessment, its relationship with that done by their class teachers, and the effect of experience on the assessment. Motivation for this study was derived from the dialogue that followed a presentation made to the first Commonwealth Planning Conference on Public Examinations by the author in 1973. The report¹ of the conference had this to say on the topic:

"The value of self-evaluation and peer-evaluation does not seem to have been fully appreciated and further development in this area would be useful. Care must be taken, however, that these techniques do not lead to an excessively competitive and negative reaction, self-derogatory and critical rather than positive and appreciative. As education is likely to move increasingly in the direction of informal and individual learning, self-evaluation will acquire an increasing importance so that improvements in this technique will be of much practical utility."

INSTRUMENTATION AND DESIGN

A brief questionnaire was developed consisting of six questions each with multiple answers provided for the pupils to select. A copy of the questionnaire is presented in the appendix. Question One enquired whether the pupil was interested in the subject. The idea was to determine on a five-point scale (spanning from 'Very Highly Interested' to 'Very Much Dislike') the pupil's level of interest. Question Two enquired how well the pupil performed usually in the subject. In other words the pupil was asked to indicate at what level on a five-point scale (spanning from 'Very Good' to 'Very Poor') he rated himself. Question Three enquired whether the subject teacher was fair in rating the pupil's work in terms of whether he was 'Over-Rated' through being 'Appropriately-Rated' to being 'Under-Rated'. Question Four enquired what range of marks the pupil would award himself in the subject that truly represents his achievements, if given the opportunity to mark his work. This assessment was done using the widely known system of percentage-marks grouped into a ten-point scale (Spanning from '1% to 10%' through '55% to 60%' to '91% to 100%'). Question Five enquired whether the pupil thought he knew about how capable he was in school work. In other words he was to indicate the extent of his knowledge of his capability on a three-point scale (spanning from 'Exact-Knowledge' to 'No-Knowledge'). Emphasis here is on "Knowledge" of capability rather than "Level" of achievement as in the case of Question Two and Question Four. Question Six enquired who should assess the pupil's work - the teachers, the pupils, or both. The six variables are summarised as follows:-

- (1) Interest (INT)
- (2) Self-Rated Achievement (SRA)
- (3) Fairness of Teachers Ratings (FTR)

- (4) Self-Marked Achievement(SMA)
- (5) Knowledge of Capability (KC)
- (6) Who Should Assess (WSA)

A seventh variable namely 'Actual Achievement', (ACT) was not indicated to the pupils in the questionnaire but was obtained from school records of the Mid-Year Examination for Forms 1 and 3, and the Mock-School Certificate Examination for Form 5. Four school subjects were selected for the study namely English Language, Mathematics, History, and Biology. The choice of subjects was influenced by the desire to include arts and science subjects, as well as subjects that have relatively high percentage of offering in the secondary schools. This report embodies the result for English Language alone.

Three Schools in the Nsukka area of the East Central State of Nigeria were used - one a boys school, one a girls school, and one a mixed school. In each school, one class each of Form 1, Form 3 and Form 5 was randomly selected and used for the study. The first, middle, and last forms of the secondary school were chosen to enable the analysis of possible effects of maturity on the variables under consideration. Every member of each class selected was given the questionnaire to complete under the direction of the author. During the exercise, class teachers were asked to keep off their classes as their presence would have influenced the response of the pupils to the questions. The pupils were assured that all information given by them would be kept confidential. At the end of each session, the members of each class were given an opportunity to react to the questions individually and the instrument as a whole. The opinions expressed during these discussions proved useful in the interpretation of the results.

RESULTS AND ANALYSIS

The frequencies of pupils responding to the various items of the questionnaire were summed up on a form to form basis and tabulated for each variables. The Chi-Squared test was used in the analysis. In each case tested, frequencies were appropriately combined to avoid or reduce the use of correction formulae where expected frequencies are very small in magnitude. Table 1 shows the frequency distribution of Interest levels within the three forms.

Table 1
Frequency Distribution of Interest in English Language

	Interest	Form 1	Form 3	Form 5	Total
(a)	Highly Interested	66	72	49	189
(b)	Interested	28	13	20	61
(c)	Indifferent	1	2	3	6
(d)	Dislike	0	0	1	1
(e)	Much Dislike	0	0	0	0
	Total	95	87	73	255

The table shows that generally there was a high level of interest in English Language. A Chi-Squared test investigating the hypothesis that there was no difference in the pattern of interest of the three forms was accepted (Chi-Squared = 9.20 with 4 degrees of freedom) at the 0.05 and 0.01 levels of significance.

Table 2 shows the frequency distribution of Self-Rated Achievement within the three forms.

Table 2
Frequency Distribution of Self-Rated Achievement

	Self-Rated Achievement	Form 1	Form 3	Form 5	Total
(a)	Very Good	36	26	4	66
(b)	Good	46	41	23	110
(c)	Average	13	20	46	79
(d)	Poor	0	0	0	0
(e)	Very Poor	0	0	0	0
	Total	95	87	73	255

The table shows that generally the pupils feel that their level of achievement in English Language is good. A Chi-Squared test investigating the hypothesis that there was no difference in the pattern of achievement of the three forms was rejected (Chi-Squared = 55.96 with 4 degrees of freedom) at the 0.05 and 0.01 levels of significance. A coefficient of contingency was determined and gave 0.42 for the relationship between level of self-rated achievement and the form attained. This shows that the higher the form attained the lower the level of self-rated achievement in English Language.

Table 3 shows the frequency distribution of the Fairness of the Teachers Ratings within the three forms.

Table 3
Frequency Distribution of Fairness of Teachers Ratings

	Fairness of Teachers Ratings	Form 1	Form 3	Form 5	Total
(a)	Much Over-Rated	17	5	0	22
(b)	A Bit Over-Rated	29	13	1	43
(c)	Appropriate	45	65	58	168
(d)	A Bit Under-Rated	1	4	12	17
(e)	Much Under-Rated	3	0	2	5
	Total	95	87	73	255

The table shows that generally the ratings by the teachers were deemed appropriate by the pupils with a slight bias towards being a bit of an over-rating. A Chi-Squared test investigating the hypothesis that there was no difference in the pattern of fairness in the three forms was rejected (Chi-Squared = 38.16 with 6 degrees of freedom) at both the 0.05 and 0.01 levels of significance. A coefficient of contingency was determined and gave 0.36 for the relationship between level of fairness and form attained. This shows that whereas at the

lower forms the pupils thought that they were somewhat over-rated by their teachers, at the higher forms the pupils held the reverse view that they were somewhat under-rated by their teachers.

Table 4 shows the frequency distribution of Self-Marked Achievement with the three forms.

Table 4
Frequency Distribution of Self-Marked Achievement

	Self-Marked Achievement	Form 1	Form 3	Form 5	Total
(a)	81% to 100%	59	31	5	95
(b)	61% to 80%	35	54	48	135
(c)	41% to 60%	1	2	20	23
(d)	21% to 40%	0	0	0	0
(e)	1% to 20%	0	0	0	0
	Total	95	87	73	255

The table shows that generally the pupils would award themselves marks ranging from 61% to 80% i.e. high marks. A Chi-Squared test investigating the hypothesis that there was no difference in the range of marks awarded in the three forms was rejected (Chi-Squared = 80.45 with 4 degrees of freedom) at both the 0.05 and 0.01 levels of significance. A coefficient of contingency was determined and gave - 0.49 for the relationship between self-marked achievement and the form attained. In other words, pupils at the lower forms have tended to award themselves higher marks than pupils at the upper forms.

Table 5 shows the frequency distribution of the pupils' knowledge of their Capability in School work within the three forms.

Table 5
Frequency Distribution of Knowledge of Capability

	Knowledge of Capability	Form 1	Form 3	Form 5	Total
(a)	Exact Knowledge	47	31	43	121
(b)	Vague Knowledge	43	52	28	123
(c)	No Knowledge	5	4	2	11
	Total	95	87	73	255

The table shows that generally pupils' Knowledge of their capability in school work was predominantly equally between exact knowledge and vague knowledge. A Chi-Squared test investigating the hypothesis that there was no difference in the level of knowledge of the pupils' capability in the three forms

was accepted (Chi-Squared - 8.87 with 2 degrees of freedom) at the 0.01 level of significance. It was however significant at the 0.05 level of significance. In other words, pupils in all three forms generally held similar views regarding how knowledgeable they were about their capability in school work.

Table 6 shows the frequency distribution of Who Should Assess the pupils' work in the three forms.

Table 6
Frequency Distribution of Who Should Assess

	Who Should Assess	Form 1	Form 3	Form 5	Total
(a)	Teachers Alone	73	78	55	206
(b)	Teachers & Pupils	22	7	18	47
(c)	Pupils Alone	0	2	0	2
	Total	95	87	73	255

The table shows that generally the pupils are of the opinion that their work should be assessed by teachers alone. A Chi-Squared test investigating the hypothesis that there was no difference in the three forms regarding who should assess their work was accepted (Chi-Squared = 6.75 with 2 degrees of freedom) at the 0.01 level of significance. It was however significant at the 0.05 level of significance. In other words, pupils in all three forms generally held similar views regarding who should assess their work.

As stated earlier in the paper, the seventh variable was the Actual Achievement of the pupils in class. Due to the fact that each school set a different examination from the others, the data for this variable could not be pooled for the three schools. The data were therefore analysed on a school to school basis. Table 7 shows the frequency distribution of the Actual Achievement of the pupils within the three forms for one of the three schools. The raw-scores were converted to descriptive terms using the following arbitrary but commonly used scheme: - 71% and above = Very Good; 61% to 70% = Good; 41% to 60% = Average; 31% to 40% = Poor; 30% and under = Very Poor.

Table 7
Frequency Distribution of Actual Achievement

	Actual Achievement	Form 1	Form 3	Form 5	Total
(a)	Very Good	0	8	2	10
(b)	Good	4	15	2	22
(c)	Average	25	18	12	55
(d)	Poor	13	3	11	27
(e)	Very Poor	3	1	12	16
	Total	45	45	40	130

Due to the fact that the tests set for the three forms were different, it was not considered desirable or meaningful to compare the performance of the three groups. If a standardised test of achievement were to have been administered to the three forms, form to form comparison would have been appropriate. The variation in the distribution of performance shown in Table 7 can be accounted for jointly by the varied mean difficulty (or facility) of the tests set, and the level of industry of the groups. The present study was not designed to study the effects of these factors.

INTER-RELATIONSHIP OF THE VARIABLES

The study investigated the inter-relationships of the seven variables in question. A contingency Table was developed for each pair of variables to be correlated. As an illustration, Table 8 is the contingency table between Interest in English Language and Self-Rated Achievement in Form 1 in the three schools used for the study.

Table 8
Relationship between Interest and Self-Rated Achievement in Form 1

		Self-Rated Achievement					
		V. Poor	Poor	Average	Good	V. Good	Total
Interest	Highly Interested	-	-	9	25	32	66
	Interested	-	-	3	21	4	28
	Indifferent	-	-	1	-	-	1
	Dislike	-	-	-	-	-	-
	Much Dislike	-	-	-	-	-	-
	Total	-	-	13	46	36	95

For each contingency table, the frequencies were appropriately combined (where necessary), the Chi-Squared value determined, and the coefficient of contingency, C, also determined. In cases where Chi-Squared is not significant at both the 0.05 and 0.01 levels, the coefficients of contingency were not computed. Table 9 presents the results of the inter-relationships of the variables. Not all pairs of variables were correlated. Only pairs of special interest to the study were correlated. Due to the fact that the vast majority of the pupils were of the opinion that they were appropriately rated by their teachers (65%) it was thought unnecessary to correlate the variable 'Fairness of Teachers Ratings' with any of the others.

Table 9
Inter-Relationships of the Seven Variables

Variables		Form	X ²	d.f	C	Remarks
1.	INT & SRA	1	11.43	2	0.33	Sig 0.01 level
2.	" "	3	13.15	2	0.36	Sig 0.01 level
3.	" "	5	0.94	1	-	Not Sig
4.	SRA & SMA	1	9.42	2	0.30	Sig 0.01 level
5.	" "	3	22.87	2	0.46	Sig 0.01 level
6.	" "	5	12.09	1	0.38	Sig 0.01 level
7.	KC & WSA	1	0.18	1	-	Not Sig
8.	" "	3	0.35	1	-	Not Sig
9.	" "	5	0.11	1	-	Not Sig
10.	ACT & INT	1	5.24	1	0.32	Sig at 0.05 level
11.	" "	3	0.01	1	-	Not Sig
12.	" "	5	0.11	1	-	Not Sig
13.	ACT & SRA	1	2.36	1	-	Not Sig
14.	" "	3	0.14	2	-	Not Sig
15.	" "	5	0.11	1	-	Not Sig
16.	ACT & SMA	1	0.08	1	-	Not Sig
17.	" "	3	0.20	1	-	Not Sig
18.	" "	5	3.58	1	-	Not Sig
19.	ACT & KC	1	7.77	1	0.38	Sig at 0.01 level
20.	" "	3	2.27	1	-	Not Sig
21.	" "	5	0.02	1	-	Not Sig
22.	ACT & WSA	1	3.02	1	-	Not Sig
23.	" "	3	0.54	1	-	Not Sig
24.	" "	5	0.00	1	-	Not Sig

(a) Interest and Self-Rated Achievement

Table 9 shows that with the exception of Form 5 Interest is positively related to Self-Rated Achievement. In other words, in the lower and middle forms, pupils with lower interest in English Language tend to think that they have lower achievement than pupils with higher interest in the subject who tend to think that they have higher achievement. In the higher forms, interest and self-rated achievement are not related.

(b) Self-Rated and Self-Marked Achievement

Table 9 shows that in all three forms Self-Rated Achievement and Self-Marked Achievement are positively related. Those pupils who rate themselves low in achievement tend to award themselves low marks while those who rate themselves high in achievement tend to award themselves high marks.

(c) Knowledge of Capability and Who Should Rate

Table 9 shows that the level of knowledge of the capability of the pupils and the opinion regarding Who Should Assess their work are not related. Pupils with exact or vague knowledge of their capability hold the same opinion regarding who should assess their work in all three forms.

(d) Actual Achievement and Interest

Table 9 shows that with the exception of Form 1 the actual performance of the pupils is not related to their interest in the subject. Pupils with high and low interest levels performed equally well in the subject at the higher forms.

(e) Actual Achievement and Self-Rated Achievement

Table 9 shows that in all three forms, the actual achievement of the pupils is not related to their self-rated achievement. Put in another way, their concept of their level of achievement is not the same, generally speaking, as those held by their subject teachers in English Language.

(f) Actual Achievement and Self-Marked Achievement

Table 9 shows that in all three forms, the actual achievement of the pupils is not related to the marks they would award themselves if given the opportunity. This is similar to the situation in (e) and should be expected.

(g) Actual Achievement and Knowledge of Capability

Table 9 shows that with the exception of Form 1, the actual achievement of the pupils is not related to their knowledge of how capable they were. In Form 1, there was a tendency for those who have exact knowledge of their capability to be average or less in actual achievement while those who have vague knowledge of their capability tend to be average and above in actual achievement.

(h) Actual Achievement and Who Should Assess

Table 9 shows that in all three forms, the actual achievement of the pupils is not related to their opinion regarding who should assess their work. Poor, average, and good pupils hold equally the predominant opinion that their work should be assessed by teachers alone.

GENERAL DISCUSSION

The high level of interest shown in English Language may be in part accounted for by the fact that it is the medium of instruction in the schools as well as the official language of the country, Nigeria. One cannot get on in the country without the language; it is a symbol and often a measure of education. It is also a compulsory subject at the secondary school level. The rather limited variability within this variable is accountable in terms of its national importance. The subject has to be taught to pupils even if it means the use of ill-qualified tutors or tutors majoring in fields quite remote from the subject. Not to show interest in the subject is to admit failure in the whole programme of secondary and tertiary education.

That the pupils rated themselves lower the higher up they went in the school can be explained from point of view of maturity. The Form 1 pupils are recent graduates from the primary schools where they distinguished themselves as the cream of the school population. They tend to have a rather exalted opinion of themselves. By the time they have spent three to five years in the secondary school, they become more conservative, they begin to feel less sure of themselves. The threat of the School Certificate Examination makes the vast majority of them feel that they do not have command of even what they have learnt so well. Around the School Certificate period, the pupils feel 'so empty' and yet they have learnt 'so much'.

The view held by the majority of pupils that the ratings carried out by their teachers are appropriate is not based necessarily on the pupils understanding the principles of reliable and valid evaluation. It is based predominantly on the age-old respect and trust which pupils have in their teachers. When those pupils who thought that their teachers rated them appropriately were requested to explain why, most of them had nothing to say, while some replied 'he is the teacher, and he knows best'. On the relationship between Fairness of Teachers' Ratings and form attained it will be observed that high marks have often been used by teachers to gratify and motivate pupils at the lower forms. At the same time low marks have often been used to humiliate and motivate pupils at the higher forms. Suffice it to say that there are plausible educational reasons for and against this practice. The results of this study have shown that the pupils are aware of this fairly widely used teaching technique. In other words, it will be unwise to assume that the pupils are still being 'fooled' either by over-rating or under-rating their achievement.

It is interesting that the pupils think that they have a fairly good knowledge of their capability (47.45% with exact knowledge and 48.24% with vague knowledge) yet they are of the opinion that their work should be assessed by teachers alone. One would have expected a fairly positive correlation between the two variables. Those with exact knowledge may be expected to ask that pupils assess their work; those with vague knowledge may be expected to ask that both the pupils and the teachers assess their work; while those with no knowledge may be expected to ask that only the teachers should assess their work. The zero to near zero correlations obtained need to be explained. Discussions that followed the administration of the questionnaire clearly showed that the pupils distrusted one another - they would not let their fate be determined by their fellow colleagues. To summarise their views it can be said that "some will give high marks to their friends and relations" while "some will give low marks to their enemies and academic rivals". This feature of extraneous factors influencing the peer-ratings is currently being studied by the author.

The relationship between Self-Rated Achievement and Self-Marked Achievement has been shown to be positive. But it is low in magnitude. The limited relationship may be in part due to the fact that there was a general tendency to award rather higher marks than would have been expected from the level of self-rated achievement. The limited variability in both variables (no pupil rated himself poor or very poor, and few rated themselves as average) may have reduced the correlation between the variables.

Actual achievement did not relate to self-rated achievement. This result could be contrary to expectations. But the nature of the subject, English Language, may have something to do with the low correlation. As the medium of instruction in the schools, some pupils are prone to assume, without varification, that they are good in the subject. On the contrary, as a foreign language, some pupils are prone to assume, again without varification, that they cannot ever be so good in it. These two factors acting at cross purposes can grossly distort the view which the pupil has regarding his level of achievement in the subject, and consequently can affect adversely the correlation between this variable and the actual achievement of the pupils. The correlation between actual achievement and self-rated achievement in English Language is perhaps the most vital information to be derived from this study.

It is so because it is also a principal aim of this study and related studies to investigate the advisability of using the pupils' assessment of themselves in overall evaluation of their attainment. The present results would suggest that the assessment of school pupils in English Language should continue to be done by teachers alone being that the pupils' assessments of themselves are quite unrelated to those of their teachers. A basic assumption of this conclusion would be that the teachers' assessments constitute an ideal or criterion against which any other measures of the pupils' attainment must be evaluated. This assumption cannot always be upheld as it is well known that the reliability of marking essay-type answers in English Language is rather low. This is more so with teachers un-tutored in the techniques of improving the marking of essays as is the case with the group of teachers involved in the study. The present use of the teachers' ratings, i.e. actual achievement, as criterion has no other justification than that the teacher is the 'Master' whose opinion on academic matters has over the ages been thought of as being weightier than that of the 'Learner' or the parents.

To throw more light on the issue of low correlation between actual achievement and self-rated achievement, the reliability of the responses to self-rated achievement was determined. The questionnaire was administered to the same pupils three weeks after the first administration for Form 1, only. A contingency table was developed between the responses in the first administration and those of the second administration of the questionnaire. This is presented in Table 10. Chi-Squared was computed and gave 27.19% for 4 degrees of freedom. This is significant at both the 0.05 and 0.01 levels. A coefficient of contingency was determined and gave 0.47 for the relationship between responses in the two administrations of the questionnaire. The result suggests that the questionnaire was reliable with regard to self-rated achievement.

Table 10
Reliability Table for Self-Rated Achievement

Second Testing	First Testing					Total
	V. Poor	Poor	Average	Good	V. Good	
Very Good	-	-	-	15	17	32
Good	-	-	4	21	18	43
Average	-	-	7	10	1	18
Poor	-	-	2	-	-	2
Very Poor	-	-	-	-	-	-
Total	-	-	13	46	36	95

This result suggests that the low correlation between actual achievement and self-rated achievement may have been due to the low reliability of the achievement scores awarded by the teachers. One can therefore predict that with better marking schedules the relationship between actual achievement and self-rated achievement can improve.

Actual achievement of the pupils did not relate to their opinion regarding who should rate their work. One would have expected that high achievers would tend to opt for assessment by pupils, average achievers to opt for assessment by both pupils and teachers, and low achievers to opt for assessment by teachers. That this is not so may be due to the effect of mutual distrust which influenced their responses.

LIMITATIONS OF THE STUDY

A major limitation of the study is the fact that the number of pupils involved is not as large as could be most desired resulting from the rather limited geographical area covered by the project. This is a consequence of the fact that the study is essentially a pilot study one of whose objectives is to explore the field and pave the way for the more extensive studies to follow. The conclusions are therefore tentative in nature. Another factor that affected the results was that interest and attainment in English Language are affected by a multiplicity of variables amongst which one can name as examples the location of the school (whether urban or rural) the influence of the local vernacular language, the qualification and experience of the teachers, the aptitude of the learners, the availability of materials etc. These and similar relevant facts were not and could not be controlled in this exploratory study. Their single and joint effect on the results cannot therefore be assessed. Replication of the study under more controlled circumstances will throw more light on the effects, if any, of these variables on pupil assessment of themselves.

Editor's Note

This is the first part of a series. Similar studies were carried out by the author on Mathematics, Biology and History.