

ARE ESSAY GRADES FAIR?

By Leonora Archer

EDITOR'S NOTE: *Mrs. Archer is a secondary schoolteacher and a graduate of the University of the West Indies/ College of The Bahamas Education Programme at the College of The Bahamas.*

Introduction

Examinations have a very important role in today's society. At the school level, they provide, among other things, help for teachers in their guidance of students into suitable courses of study. The courses of study will, in turn, determine the future of many students. It is therefore important that the examinations given to these students be valid and reliable — that is, measuring what they set out to measure and doing it consistently.

Here in The Bahamas, the National examination (BJC) given at the end of junior high school, is used by the majority of educators as a basis for selection of students into high level courses. While most of the examinations offered at BJC level include essay-type questions, it has been noted that BJC examiners, in English Language in particular, allocate more than fifty percent of the total marks to questions requiring essay-type answers. This seems to indicate that educators consider the essay tests as a reliable measure of students' writing ability.

The present research was therefore undertaken to determine whether the essay test provides a reliable measure of the free writing ability of junior high students.

The literature reviewed indicated that essay tests are highly unreliable. Various studies (Palmer, 1961 Burton, et al, 1975) have shown that the score a student receives on an essay test is determined to a great extent by: The person who marks the script, the topic on which the student writes, and the particular marking scheme used in scoring the essay.

The investigation therefore set out to examine whether these findings would be applicable to the Bahamas. Answers were sought to the following three questions:

- (1) Would marks given for an essay test vary when the scripts were graded by different examiners?
- (2) Would marks given for an essay test vary when different topics were given to the same students?
- (3) Would marks given for an essay vary when different marking schemes were used?

Research Procedures

In order to obtain this information, twelve essays written by six junior high school students on two topics selected from the 1978 BJC English Language examination were marked by eighteen experienced BJC English Language examiners. The examiners were divided into three groups with each group using either the mid-analytic (partially detailed), the BJC (completely detailed), or the holistic (general impression) marking scheme.

Each essay received six independent ratings within each method of scoring, and the examiners were asked to grade the essays and award them pass/fail judgements in accordance with the requirements at BJC level.

The Findings

Reliability of the examiners. — It was found that the score a student received varied from examiner to examiner, topic to topic, and from one marking scheme to another.

Examiners were unable to agree on the scores, ranks, and pass/fail judgements awarded to each of the twelve essays. Each essay received scores ranging from a very poor rating to a near excellent one, even in the same marking scheme. As table 1 shows, an essay gained a score of 7 from one examiner, while another examiner awarded the same essay 30 points — both out of a possible 40. Another essay received an "E" and also a "B" rating from examiners using the same marking scheme.

The examiners also could not agree on the quality of the essays. One examiner gave an essay a second place rank, while another examiner felt that the essay was the worst of the twelve. Another essay was considered the best of the twelve

essays by one marker and was given a tenth place rank by another.

Agreement among the examiners was not found on the pass/fail status of the essays. With the exception of one of the essays, the essays were passed and failed depending on the particular examiner who marked the script.

Reliability of the test. — It was also found that the majority of students performed better on a particular topic than on the other. Most of the students gained a higher score on the narrative essay than on the descriptive essay. Differences were also seen in the ranks and pass/fail judgments the students received for each of their two essays. One student received a fifth rank average for his narrative essay and an eighth rank average for his descriptive essay. He also gained eleven passes for the narrative essay compared with only six passes for the descriptive essay.

Reliability of marking scheme. — Allowing for the variability among markers, it was apparent that the particular marking scheme used also made a difference in the scores the essays received. Most of the essays gained lower scores and fewer passes from examiners using the BJC (detailed) marking scheme. Only one of the 12 essays received an average score of over 20 in the BJC marking scheme, while mid-analytic examiners awarded scores of 20 or more to at least 5 essays. It was also noted that none of the essays received the same average mark in all three marking schemes.

TABLE 1
DISTRIBUTION AND RANGE OF MARKS AWARDED
TO FOUR OF THE TWELVE ESSAYS

Essays	Markers																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
I	35	31	21	27	19	23	20	24	18	10	16	20	C	C	B-	B	C	B-
II	26	28	28	23	15	15	23	15	20	9	21	17	D+	C	D	D	B	B-
IV	10	0	19	19	12	12	21	10	9	3	6	12	D-	E	E	E	E	E
X	19	19	17	27	15	10	30	14	18	7	12	17	C	D	D	C	C+	C

Discussion

It has been shown in this research that the score a student receives on an essay test is determined more by the particular examiner, the topic and the marking scheme than by what he has actually written. It is necessary to notice the low level of agreement found among individual examiners using the same marking scheme when judging the quality of an essay. This, no doubt, indicates that the rating given an essay by only one examiner cannot be taken as a true evaluation of a student's writing ability.

One must also note the variations found in the students' performance on the two essays. As these two topics were taken from the same essay test (BJC, 1978), it would seem to suggest that students should not be given an unlimited choice of topic on an essay test as they may be incapable of selecting the topics on which they are best able to write. It must be mentioned, too, that a student's ability to organize, integrate and express his ideas is greatly influenced by the degree of difficulty of the given topic.

The present research also points to the fact that marking schemes alone will not reduce the unreliability of essay marking. The observed variations within the detailed (BJC and mid-analytic) marking schemes seem to indicate that (contrary to the views of some Bahamian educators) detailed marking schemes do not provide a greater measure of agreement among examiners. Instead, the highest level of agreement on the quality of the twelve essays was found among examiners using the holistic (general impression) marking scheme.

On the basis of the findings of this research, it therefore seems necessary for educators and testmakers to re-examine the whole concept of essay tests as an integral part of the BJC English Language examinations. Consideration must be given to either increasing the reliability of the essay test (allowing multiple rather than single gradings of each essay) or de-emphasizing its importance (allocating a smaller percentage of the total marks to questions requiring essay type answers).

Failure to do this could mean that educators are being unfair to the hundreds of students whose very careers are determined by BJC English Language examination results.