APTITUDE TESTS, SOCIO-ECONOMIC BACKGROUND AND SECONDARY SCHOOL SELECTION: THE POSSIBILITIES AND LIMITS OF CHANGE

BY

H.C.A. Somerset

DISCUSSION PAPER NO. 249

INSTITUTE FOR DEVELOPMENT STUDIES
UNIVERSITY OF NAIROBI
P.O. Box 30197
NAIROBI, KENYA.

MARCH 1977

Mr. Somerset, until recently a Visiting Senior Research Fellow at I.D.S., is now working in the Examinations Section of the Kenya Ministry of Education.

IDS/DP 249

APTITUDE TESTS, SOCIO-ECONOMIC BACKGROUND AND SECONDARY SCHOOL SELECTION: THE POSSIBILITIES AND LIMITS OF CHANGE

by

H.C.A. Somerset

ABSTRACT

A group of verbal reasoning questions, of the type often used in intelligence tests, were included in the English paper of the Kenya secondary school selection examination (C.P.E. - Certificate of Primary Education) from 1971, and from 1974 scientific reasoning questions were included in the general paper. In early 1974 a full-scale item analysis of the 1973 selection paper was made. This paper presents some results from the analysis carried out for the years 1973 to 1976 of the English paper and the science section of the general paper.

The scores of pupils from Nairobi high-cost primary schools and from rural low-cost schools are compared. Three interesting results of this analysis are presented and some possible explanations offered. For one thing, pupils from the Nairobi high-cost schools perform up to 70 per cent better on English items testing knowledge of specialised words, expressions and idioms, whereas in the science paper, descriptive items testing knowledge of specialised and technical terms produce a mean difference of less than 12 per cent. Secondly, the verbal reasoning items in the English paper give the Nairobi high-cost pupils a smaller advantage than the achievement items; whereas in science, the reasoning items give a bigger advantage than the descriptive and explanatory items. Finally, the observation items in the science paper produce an especially large performance gap between the two types of schools.

These and other results suggest that the huge performance advantage enjoyed by Nairobi high-cost pupils in the English paper can be ascribed entirely to two sources: first, their greater familiarity with the language, and second, the superior quality of the teaching they receive. In both papers, questions which test higher-level intellectual skills, such as the ability to reason, are particularly sensitive to the effects of teacher quality, rather than reflecting the innate capacity of the pupils.

IDS/DP 249

APTITUDE TESTS, SOCIO-ECONOMIC BACKGROUND AND SECONDARY SCHOOL SELECTION: THE POSSIBILITIES AND LIMITS OF CHANGE

by

H.C.A. Somerset

Throughout most of the 1960s the main preoccupation of educational planners in newly independent African countries was the training of local high-level and middle-level manpower. New secondary schools and universities had to be built, equipped, and staffed quickly; this could be achieved only by diverting resources of money and manpower from other uses. Inevitably, the primary schools suffered. A high proportion of the ablest primary school teachers were filtered off to take up jobs in the secondary schools or in educational administration, and many of the new recruits who took their places lacked real commitment to primary teaching as a career. Primary teacher training was very often seen by school leavers as an inferior substitute for further formal education and as a stepping-stone to employment in other better paid and more prestigious occupations.

Faced with declining and increasingly uneven standards of primary education, many countries became concerned about secondary school selection. Clearly, pupils from low-quality schools could not be expected to reveal their full potential in conventional examinations, which measured only achievement in school subjects. Aptitude tests seemed to offer an inexpensive and effective solution to the problem. Such tests, it was said, were relatively immune to the effects of socio-economic and school-quality factors; they could provide a measure of the pupil's capacity for future development, rather than of his current attainment. Hence they could be used to improve both the efficiency and fairness of selection. Gifted pupils who had been badly taught would no longer be excluded from secondary school.

The validity of these claims had been the subject of heated, if intermittent, debate for more than forty years. As early as 1922, Walter Lippman attacked the intelligence testing movement in terms which foreshadow at some points the recent more radical critique of writers such as Bowles and Gintis:-

> The danger of the intelligence tests is that in a wholesale system of education, the less sophisticated or the more prejudiced will stop when they have classified and forget that their duty is to educate. They will grade the retarded child instead of fighting the causes of his backwardness. For the whole drift of the

low intelligence quotients as congenitally and hopelessly inferior....

Most of the more prominent testers have committed themselves to a dogma which must lead to just such an abuse. They claim not only that they are really measuring intelligence, but that intelligence is innate, hereditary and predetermined.

...Intelligence testing in the hands of men who hold this dogma could not but lead to an intellectual caste system in which the task of education had given way to the doctrine of pre-destination and infant damnation. If the intelligence test really measured the unchangeable hereditary capacity of human beings, as so many assert, it would inevitably evolve from an administrative convenience into a basis for hereditary caste. (Lippman, 1922, pp. 19-20)

By the 1950s, however, the balance of liberal opinion was generally favourable to intelligence testing. The claims of the more extreme enthusiasts that intelligence tests measured innate potential were rejected; but it was generally agreed that the tests reduced the effects of environmental handicaps and hence gave a more valid measure of the child's real ability than conventional tests of attainment. In Britain, this opinion was buttressed by an influential study carried out by Floud and Halsey (1957, pp. 33-39). These workers determined the social class origins of grammar school entrants in one county for the years 1952, 1953 and 1954. During this period the controlling county education authority removed an intelligence test from the eleven plus examination used for grammar school selection and substituted an English composition paper. Floud and Halsey summarised their results as follows:

> ...the changes in procedure have resulted, at any given level of ability, in a slight but persistent diminution of opportunity for working-class boys and a corresponding increase in opportunity for those at the higher social levels.

Formal aptitude tests were introduced into the examination batteries used for secondary school selection in many independent countries in West, Central and Southern Africa from the mid-1960s. In Kenya, however, the approach was more cautious. A group of verbal reasoning questions, of the type often used in intelligence tests, was included in the English paper of the selection examination (Certificate of Primary Education - C.P.E.) from 1971, and from 1974 scientific reasoning questions were included in the general paper. Evaluation of the results of the changes was not possible until early 1974, when a full-scale item analysis of the 1973 selection examination was made.

THE SAMPLE

1. The rural low-cost (R.L.C.) sample. This consists of a 5 per cent random sample of all Schedule A (low-cost) schools in rural districts of the country. All C.P.E. candidates in the selected schools are included.

Prior to independence, Schedule A schools catered exclusively for pupils of African origin. They are now technically open to all ethnic groups, but in fact virtually all the pupils who attend them come from low-income African families. Schedule A schools now charge no fees at all in the first four standards; from standards 5 to 7, the fees are Shs. 60 (US$7) per annum. Building fees and other charges, however, usually bring the total cost of attendance up to about Shs. 100 to 150 per annum. At least 95 per cent of the primary schools in Kenya fall into the rural low-cost category. In 1973, this sample consisted of about 250 schools and about 8,300 pupils. By 1976, the number of pupils had risen to about 8,600. (The exact numbers vary slightly from paper to paper.)

2. The Nairobi low-cost (N.L.C.) sample. This is a 20 per cent sample of the schedule A schools under the control of the Nairobi City Council. A few schools which teach an experimental modern mathematics syllabus are excluded. Nairobi Schedule A schools charge the same fees as Schedule A schools in other parts of the country, but their buildings and equipment are usually superior. They also employ only qualified teachers, whereas in the country as a whole about 32 per cent of primary teachers are unqualified. This sample numbered a little over 800 pupils in both 1973 and 1976.

3. The Nairobi high-cost (N.H.C.) sample. The third sample is made up of all pupils who attend the eight Nairobi City Council Schedule C schools. These schools were originally established for European pupils only, but a short while before independence in 1963 the racial criterion for admission was abandoned. However, the other two criteria -- competency in English and ability to pay the fees -- were retained. By 1976, about 90 per cent of the pupils sitting C.P.E. from these schools were of African origin. Most of the remainder were Asians; only a handful were Europeans. The total cost of attending a Schedule C school is about Shs. 800 ($100) per annum. In 1973, the Nairobi high-cost sample totalled a little over 500; by 1976, this number had risen to about 600.

Because places are restricted, competition for entry to Schedule C schools is intense. Many parents now enrol their children in high-cost pre-primary schools for one to three years in an attempt to improve their

chances. In these pre-schools pupils are taught in English, often by teachers who speak English as their first language. The fees charged are usually much higher than in the Schedule C primary schools; they range from about Shs 1,200 to well over Shs. 3,000 per annum. The annual output from these institutions in Nairobi is now well over 2,000 children, which is about three times the number that the Schedule C primary schools can absorb. Hence attendance at a high-cost pre-primary school is now a virtual prerequisite for a Schedule C place. This in turn means that access to the Schedule C schools is effectively restricted to children from higher-income families. Many families of quite modest means could perhaps find the money for the Schedule C fees; but the fees charged in most high-cost pre-primary schools are another matter. Pupils who have attended both a high-cost pre-primary school and a Schedule C primary school are generally more fluent in English than in any other language. Most of them use English as their main medium of communication, in everyday conversation as well as in the classroom.[1]

For practical reasons two other types of primary school were omitted from the C.P.E. item analysis. Schedule B schools are the former Asian schools. They are all located in urban areas, and are generally intermediate in character between Schedule A and Schedule C schools. There are also a number of private primary schools run by the Catholic Church which enter their standard 7 pupils for the C.P.E. examination each year. These schools charge fees ranging between Shs. 1,800 and 3,000 per annum. For convenience, we can refer to them as Schedule D schools, although this designation is quite unofficial. Three of the four Schedule D schools in Nairobi are attached to secondary schools. This means that very few Schedule D pupils ever fail the C.P.E., in the sense that they fail to gain a secondary school place. Until quite recently, the majority of pupils in Schedule D schools were Europeans and Asians, but over the last few years the proportion of Africans has increased rapidly, party as a result of pressure on the Schedule C system. In the lower standards, the proportion of African pupils in most classes is now about 70 per cent.

At the top of the status hierarchy are the high-cost private schools, run on a proprietary basis. These schools do not enter their pupils for C.P.E. at all; instead they prepare them for entry to private secondary

---

1. I am indebted to Mr. O.N. Gakuru for his generous permission to quote data from his current research. For further details, see Gakuru, 1977.

schools, mainly in Britain, but also to an increasing extent in Kenya. Fees in these schools range up to Shs 10,000 per annum for a day pupil; for a boarding pupil they are much higher. As yet, very few Kenya African pupils have enrolled at these schools.

The Certificate of Primary Education is the examination which terminates primary education. Nominally, the primary course runs for seven years, but the median period taken by pupils to pass through the primary standards is between eight and nine years. In 1976 nearly a quarter of a million pupils sat the examination, and of these, about 30,000 found places in government-maintained secondary schools at the beginning of 1977. The continuation rate for Kenya as a whole is thus about 12 to 13 per cent. But for pupils from the Nairobi Schedule C schools the continuation rate (mostly into high-cost government secondary schools) is at least 70 per cent, and probably much higher. Schedule C pupils who do not meet the performance requirements for a government secondary place usually continue their education at a high-cost private secondary school.

The examination consists of four papers, which are all given on the same day. Three of the papers -- English, mathematics and the general paper -- are in multiple-choice form. Four alternative answers are provided to each question, so that a candidate who makes no attempt to work through the questions but simply guesses should score 25 per cent by chance alone. The questions for these papers are not pre-tested, but the results of the item analysis of the previous year's questions are used during moderation. Until 1975, the English paper contained 70 questions, but this number has now been reduced to 40. The mathematics paper consists of 50 questions, and the general paper, 80 questions, made up of history (20 questions), geography (20 questions) and science (40 questions). These three objective papers are marked by computer, but the fourth paper, English composition, is marked by teams of secondary school teachers.

THE ANALYSIS

Item analysis is based on the raw marks. Each of the three samples is analysed separately. The proportions of candidates giving each of the four answers to each question is calculated, together with a simple discrimination index. The formula for the index is

$$D = \frac{(\bar{M}_R - \bar{M}_W) - 1}{SD}$$

where M_R is the mean total mark in the subject being considered of pupils answering the question correctly, M_W the mean total mark of those answering

incorrectly, and SD the standard deviation of the total mark distribution. One point is deducted from the differences between the means to allow for the mark gained from the question under consideration. Hence a D index of 0 indicates that the candidates who answered the question being considered correctly scored exactly the same mean mark on the other questions in the paper as the candidates who answered incorrectly. Similarly, a D index of 1 indicates that the difference in means between the two groups was one standard deviation. Because of limitations of computer time, other more complex statistics have not been calculated.

There is an important point to be made about item analysis. The technique is essentially conservative, in that it judges the efficiency of each item in terms of an internal criterion. It depends on the crucial assumptions that the human attributes which the test or examination should measure have been defined, and that the total score on the test is a useful measure of these attributes. If either or both of these assumptions is violated, then the value of the item analysis will be severely reduced. For example, if an examination paper is made up mainly of items which test memory and recall, then the total score on that paper will be a measure of that ability. Item analysis will then define the component questions as efficient to the extent to which they assess memorisation skill. But if the purpose of the examination should have been to tap other intellectual skills (for example, the ability to draw inferences from data or to see relationships) then we would be quite unjustified in concluding that the items with high discrimination indices were good ones. Item analysis can establish only the technical efficiency of the questions in an examination: it can tell us nothing about their relevance or usefulness. The technique is an aid to professional judgement, not a substitute for it.

The 1973 English Paper

The first detailed item analysis of the C.P.E. early in 1974 produced encouraging results with respect to pupils from different types of schools. The discrimination indices of the verbal reasoning items in the 1973 English paper were a little higher than the indices of achievement items of comparable difficulty. But more important than this, the verbal reasoning items gave the privileged Nairobi high-cost pupils a smaller performance advantage over the low-cost pupils -- especially those in rural areas -- than most types of achievement items. The relevant data are set out in Table 1.

Table 1. C.P.E. English 1973 — Mean difficulty levels of the various item types in the three types of school.

Item Nos.	Item Type	Mean Difficulty Levels			Differences	
		Nairobi high-cost	Rural low-cost	Nairobi low-cost	NHC-RLC	NHC-NLC
56-63	Comprehension I: short story	74.01%	32.55%	35.77%	41.46%	38.24%
64-70	Comprehension II: descriptive passage	68.31%	34.66%	37.55%	33.65%	30.76%
25-40	Word insertion I: conversational English	92.37%	58.80%	62.76%	33.57%	29.61%
1-24	Word insertion II: descriptive English	77.00%	51.12%	51.00%	25.88%	26.00%
41-45	Paraphrasing	89.42%	64.94%	66.98%	24.48%	22.44%
46-55	Verbal reasoning	60.08%	35.93%	35.20%	24.15%	24.88%

As might be expected, the Nairobi high-cost (N.H.C.) pupils have a substantial advantage over low-cost pupils in questions of all types, but the performance gap is much larger in some sections of the paper than in others. If we compare N.H.C. pupils with rural low-cost (R.L.C.) pupils, the gap is as large as 41.46 per cent in the first comprehension section, based on a short story; it drops to 33.65 per cent in the second comprehension section, which involves a passage of straightforward, descriptive English; to 24.48 per cent in a short paraphrasing section, which tests mainly knowledge of grammar and syntax; and it is smallest, at 24.15 per cent, in the verbal reasoning questions.

The patterns are very similar, if not quite so striking, when we compare the high-cost Nairobi pupils with low-cost pupils in the same city. Pupils in the high-cost schools have an advantage of more than 38 per cent in the first comprehension section, but less than 25 per cent in the verbal reasoning and paraphrasing sections.

By contrast, the performance differences between the two low-cost samples -- Nairobi and rural -- are never more than three or four percentage points. Moreover, one sample does not enjoy a consistent advantage over the other. This result has been repeated in every C.P.E. subject and in every year that the examination has been analysed. Clearly, the major cleavage in C.P.E.

performance is between high-cost and low-cost schools. By comparison, the differences between low-cost schools in Nairobi and those in the rural areas are of only minor importance.

For simplicity, we shall concentrate in the remainder of this account on the differences between the Nairobi high-cost and rural low-cost samples. As we have already mentioned, R.L.C. schools make up more than 95 per cent of all primary schools in Kenya. In general, results from the N.L.C. sample will be discussed only when they diverge significantly.[2]

Analysis of the separate questions in the English paper suggested several reasons why the performance gap was so much wider in some sections of the paper than in others. For example, one of the questions based on the first comprehension passage tested knowledge of the word 'scolded'. Virtually all of the N.H.C. pupils had this word within their vocabulary; virtually none of the R.L.C. pupils. The proportions answering correctly were 90.4 per cent and 29.6 per cent respectively, so that the performance gap was 60.8 per cent. Two other questions from the same section asked pupils to interpret the idiomatic expressions 'he should have plucked up his courage' and 'he was not of their standing'. These two questions produced differences of 51.9 per cent and 46.3 per cent respectively.

But the biggest performance difference of all came from a question in a section testing understanding of conversational English. The candidates were asked to fill in blanks to complete a dialogue. One man asks another about some visitors who had arrived the previous evening:-

'But how are they? They must be tired ——— their long journey.'

Only 18.1 per cent of the rural low-cost pupils correctly inserted 'after' into the blank to complete the sentence, as compared with 91.0 per cent of the Nairobi high-cost pupils. The performance gap was thus 72.9 per cent.

The word chosen most often by R.L.C. pupils to fill the blank was 'of'. The expression 'tired of' as a loose synonym for 'bored with' or 'impatient with' is known to most Kenyans who use English as a second language. Indeed it is commonly used by teachers when talking to pupils: 'I am tired of your noise'; 'I am tired of your poor work'. But 'tired after' is heard far less often. The distinction in meaning between the two expressions is entirely lost on the

R.L.C. pupils.

Other questions in the English paper, especially some of those involving verbal reasoning, gave the Nairobi high-cost pupils far less of an advantage. More than half the verbal reasoning items were constructed from information or experiences which seemed likely to be more familiar to rural than to urban pupils. In one reasoning item the N.H.C. advantage was as low as 5.6 per cent. In this question, candidates were asked to infer the direction of a boy's school from his house from information about his shadow as he walked to and from school in the mornings and evenings. Most rural pupils walk long distances each day, and have plenty of time to observe the behaviour of their shadows. Most Nairobi high-cost pupils, by contrast, are driven to school by car.

These results suggested implications for policy. The very large performance advantage enjoyed by the pupils from Nairobi high-cost schools could be reduced significantly if the paper were modified in two ways:

1. The achievement sections of the paper should focus on testing knowledge of straightforward, everyday English. Questions involving uncommon words, phrases and idioms, which are not often used or understood in rural areas, should be omitted.

2. The proportion of questions testing higher-level verbal skills, such as the ability to draw inferences from given facts, should be increased. Correspondingly, the proportion testing memory and recall should be reduced.

The General Paper in 1973 and Subsequent Years

Although these conclusions were derived from analysis of the English paper, they seemed equally applicable to other subjects tested in the C.P.E. particularly those in the general paper, science, history and geography. The item analysis report for the 1973 examination noted that the questions in the general paper worked well in a technical sense. Most of the items had high or fairly high discrimination indices; moreover, they were well spread over the full difficulty range. But these considerations did not seem the most important:-

> The major point to be considered is the content of the questions rather than their technical characteristics. Most of the items test the candidate's ability to recall factual material. Only a few test higher-level cognitive skills: the ability to understand relationships of cause and effect, for instance, or the ability to use factual material to draw inferences. Similarly only a few items test his ability to observe significant features of his local environment. Issues

Descriptive items test knowledge of straightforward, single facts. They require no preparation by the candidate except memorisation, and no intellectual operation in the examination room apart from recall. In science, descriptive items are usually concerned with the questions 'what' and 'which'; in geography, they can also involve the question 'where'; and in history, the questions 'who' and 'when'.

In science the fact tested by the item may be a specialised or technical term. Three descriptive items of this type from the 1973 science section are given above. In other descriptive items, the fact being tested is expressed in everyday language and does not require knowledge of highly specialised terms:-

> A mammal that feeds during the night is
>
> A. cow B. waterbuck C. gazelle D. bat
>
> (Item 80, 1973 paper)

In the 1973 paper, as many as 19 of the 40 questions were descriptive items testing the candidates' memory of specialised scientific terms. In the following year the number of items of this type was reduced to four, and since then they have been eliminated from the paper almost entirely. Descriptive items which do not test specialised terms have been retained, but a high proportion of them now have a strong rural bias. For example:-

> For good health we need to eat a balanced diet, containing energy-giving foods, body-building foods, and foods which protect us against diseases. Which of the following groups of foods gives us the best balance?
>
> A. Maize and beans, bread and butter, tea with sugar
> B. Ugali with meat, cake, a bottle of soda
> C. Sausages with potatoes, bread, tea with sugar
> D. Maize and beans, cabbage, milk.
>
> (Item 66, 1975 paper)

In descriptive items, facts are treated essentially as isolated entities. But in explanatory items, candidates are expected to show understanding of how the fact fits into a wider context of cause-and-effect relationships. Explanatory items involve the questions 'how' and 'why'; and the word 'because' frequently appears in the stem. In short, they are concerned with understanding as well as with knowledge.

> Leopards are yellow in colour with black markings. This is an advantage to the leopard because the colours

A. Frighten the leopard's prey
B. Keep it cool in the hot sun
C. Make it difficult to see in the places where it usually lives
D. Are attractive and make its skin valuable.

(Item 80, 1976 paper)

The proportion of explanatory items in the science section of the general paper has risen modestly, from about one-sixth in 1973 to about one-quarter in more recent years.

The third group of questions consists of those testing observation. An experiment of some kind, sometimes illustrated with a diagram, is described in the stem of the question, and the candidate is asked to say what he would expect to happen. Strictly speaking, the distinction between observation items and the other types concerns the source of the information rather than the kind of intellectual process involved. The data obtained from observation may be either descriptive or explanatory. It is always possible, of course, for pupils to learn the answers to observation-type questions from their textbooks or teachers' notes, without having conducted any practical investigations. An item was classified in this category only if it seemed clear that pupils who had actually made the observation would have a big advantage:-

The following pictures show several ways of connecting a bulb to a battery. Which one will light?

(Item 65, 1975 paper)

The experiment described in this question requires some simple equipment: a torch cell, a bulb, and a length of wire. Other observations, however, can be carried out using only the resources of the everyday environment:-

Owino and Masale are two boys of about equal strength. Owino was trying to close the classroom door by pushing very near to the hinges. Masale was trying to push the door open by the handle from the other side. What is most likely to happen?

A. Masale opens the door easily
B. Owino shuts the door easily
C. The door neither opens nor closes
D. The door breaks.

(Item 55, 1976 paper)

The proportion of observation items in the paper increased from less than 10 per cent in 1973 to 20 per cent in 1974 and since then has been maintained at about the same level.

The final group of questions tests scientific reasoning. The candidate is given a set of facts or observations, and then asked to draw an inference from them. In other words, he must show that he can carry out the processes by which elementary scientific conclusions are established from data. Most or all of the information needed is supplied in the stem of the item.

Some items require deductive reasoning:-

> Salad oil floats on paraffin and paraffin floats on water. Therefore:-
>
> A. Water is lighter than paraffin
> B. Salad oil will float on water
> C. Paraffin is lighter than salad oil
> D. Water will float on salad oil.
>
> (Item 67, 1974 paper)

In other items, the reasoning required is inductive:-

> Use the information given below to answer Questions 78 and 79. Scientists in Machakos District experimented with two kinds of maize, called Katumani and Tabora, to see which kind would give the bigger crops, and how far apart the seeds should be planted. In some plots the seeds were planted 60cm apart along the rows and 90cm apart between rows and in other plots they were planted 30cm apart along the rows and 90cm apart between rows.
>
> | They obtained the following results: | Weight of crop, per hectare |
> |---|---|
> | Tabora seeds, planted 60cm by 90cm apart | 1,600 kg |
> | Tabora seeds, planted 30cm by 90cm apart | 2,400 kg |
> | Katumani seeds, planted 60cm by 90cm apart | 2,100 kg |
> | Katumani seeds, planted 30cm by 90cm apart | 2,900 kg |
>
> 78. Judging from these results, which kind of maize is likely to give a Machakos farmer heavier crops?
>
> A. Tabora maize
> B. Katumani maize
> C. The two kinds of maize give equally heavy crops
> D. It is impossible to reach any conclusion from these results.
>
> 79. Using the same results, which of the following is correct?
>
> A. Both kinds of maize gave heavier crops when planted 30cm by 90cm apart
> B. Both kinds of maize gave heavier crops when planted 60cm by 90cm apart
> C. Tabora maize gave a heavier crop when planted 30cm by 90cm apart, but Katumani maize gave a heavier crop when planted 60cm by 90cm apart
> D. The distance which the seeds were planted apart made no difference to the weight of the crop.
>
> (1975 paper)

No reasoning items were included in any C.P.E. science paper prior to 1974. They now make up about one-quarter of the questions.

It should be stressed that the C.P.E. science items do not always fit neatly into one of the four main categories we have described. Very often a certain amount of subjective judgement is involved in classifying them. For instance, the question concerning balanced diets quoted above as an example of a descriptive item contains also a reasoning element. The candidate must know the facts about the three main dietary components, but he must apply that knowledge to decide which of the given groups of foods provides the best balance among them. Similarly the boundary between descriptive items, testing knowledge of specific facts, and explanatory items, concerned with more general concepts, is by no means always clear.

The move in the 1974 science paper away from items dependent mainly on memory towards items testing higher-level intellectual skills was very successful as far as the discrimination efficiency of the paper was concerned. In the rural low-cost sample the lowest discrimination index for any item was 0.30, which was considerably higher than the minimum index in the other subjects. Table 3 demonstrates the relationship between item type and efficiency.

Table 3. C.P.E. science 1974 — discrimination efficiency of items by item type and difficulty (rural low-cost sample).

Difficulty Level	% Pupils Answering Correctly	Items with 'high' D Indices					All Items				
		Descr	Explan	Reas	Obs	Tot	Descr	Explan	Reas	Obs	Tot
Easy	70%+	0	3	0	0	3	4	5	0	0	9
Moderately easy	50-69%	0	2	2	1	5	5	4	5	4	18
Difficult	-49%	1	0	2	1	4	3	3	3	4	13
TOTAL		1	5	4	2	12	12	12	8	8	40

For a valid comparison, it is necessary to control the difficulty of the items, because the D index and the difficulty-level are negatively correlated (easier items tend to have higher D indices). A D index was classified as high if it was among the top 30 per cent of indices for items in the same difficulty category (easy, moderately easy, or difficult).

It can be seen that explanatory and reasoning items are considerably over-represented among the items with high indices and descriptive items considerably under-represented. Four of the eight reasoning items and five of the twelve explanatory items are in the high category, but only one of the twelve descriptive items.

The main reason for this result is almost certainly that there is a larger chance element involved in answering a descriptive item correctly than there is in answering an explanatory or reasoning item. No candidate, however competent and conscientious, can learn equally thoroughly all the facts he may be asked about in an examination. If he has worked hard memorising the botanical names for the parts of a plant, he may well have had inadequate time to learn the functions of the barometer, the anemometer and the hygrometer. Because the facts are essentially isolated from each other, there is little or no transfer of learning; each fact must be learned as a separate entity. Explanatory and reasoning items, by contrast, focus on the cause-and-effect relationships between facts rather than on the facts themselves. A capable pupil who has built up an overall understanding of the subject is less likely to be penalised in items of these types if he has forgotten, or perhaps failed to learn, some of the marginal points of detail.

The change in emphasis in the 1974 science section produced an improvement in discrimination efficiency. Its effects on the performance gap between Nairobi high-cost and rural low-cost schools can be seen from Table 4.

Table 4. C.P.E. science 1974 — mean difficulty levels of the various item types in Nairobi high-cost and rural low-cost schools.

Type of Item	Number of Items	Proportion Answering Correctly: Nairobi high-cost	Proportion Answering Correctly: Rural low-cost	Difference
Descriptive				
(a) Specialised terms	4	65.64%	55.97%	9.67%
(b) Non-specialised	8	75.57%	62.80%	12.87%
Explanatory	12	71.25%	60.09%	11.16%
Observation	8	72.68%	48.00%	24.68%
Reasoning	8	70.21%	51.35%	18.86%

It is immediately apparent that these results do not confirm the expectations formed from our analysis of the 1973 English paper. In English, it will

be remembered, the items which penalised rural low-cost pupils most severely were those testing knowledge of specialised words, expressions and idioms. With some such items, the performance gap reached 50 per cent or even 70 per cent. But in science, items testing knowledge of specialised scientific and technical terms have a mean N.H.C.-R.L.C. performance gap of less than 10 per cent. Far from penalising the rural pupils, these items appear to give them a better chance of competing on equal terms with the Nairobi high-cost pupils than questions of any other type. Similarly, the verbal reasoning items in the English paper produced a performance gap which was narrower than that for all types of achievement items; whereas the scientific reasoning questions produced a gap which was wider than that for any other type of science item except the observation items. The move away from descriptive items towards questions testing higher-level skills appears to have increased rather than reduced the advantage enjoyed by Nairobi high-cost pupils.

It seemed possible that the low N.H.C.-R.L.C. performance gap for the descriptive items testing specialised terms might be due in part at least to chance factors because there were only four items of this type in the 1974 paper. The 1973 C.P.E. science data were therefore re-analysed to provide comparable results. These are set out in Table 5.

Table 5. C.P.E. science 1973 — mean difficulty levels of the various item types in Nairobi high-cost and rural low-cost schools.

Type of Item	Number of Items	Proportion Answering Correctly		Difference
		Nairobi high-cost	Rural low-cost	
Descriptive				
(a) Specialised terms	19	63.76%	51.94%	11.82%
(b) Non-specialised	9	72.21%	59.34%	12.87%
Explanatory	7	77.46%	62.53%	16.59%

The 1974 patterns were essentially confirmed. With 19 specialised-term items rather than only 4 the mean performance gap was up by only 2 per cent and still lower than for items of the other two types. The only noticeable difference in the two sets of results is that the explanatory items produced a wider gap in 1973 than they did in 1974.

CONCLUSIONS

There are three aspects of these results which must be explained:

1. Why do items in the English paper testing knowledge of specialised words, expressions and idioms produce N.H.C.-R.L.C. performance differences of up to 70 per cent; whereas in the science paper, descriptive items testing knowledge of specialised and technical terms produce a mean difference of less than 12 per cent?

2. Why do the verbal reasoning items in the English paper give the Nairobi high-cost pupils a smaller advantage than the achievement items; whereas in science, the reasoning items give them a bigger advantage than the descriptive and explanatory items?

3. Why do the observation items in the science paper produce an especially large performance gap?

We shall deal with the third of these points first because it is the most straightforward. It may be recalled from Table 2 that all of the observation items in the science section of the 1974 general paper required special equipment of some kind. This equipment included a rubber sucker to stick on a wall, two glass rods and a burner, and a glass jar. Such equipment is far more likely to be available in a Nairobi high-cost school than in a rural low-cost school. In the 1975 and 1976 papers, a total of nine observation items which did not require special equipment were included. These items produced a mean performance gap of only 10.35 per cent, as compared with 19.93 per cent for the 13 items which did require equipment.

Turning our attention now to the first of the three questions listed above, we may suggest two closely-related reasons why rural low-cost pupils are so much less penalised by items testing specialised vocabulary in science than they are in English. In the first place, the vocabulary pool on which the items draw is both smaller and more clearly defined in the science questions than in English. Secondly, definitions of the scientific terms which must be learned are more easily accessible to pupils in rural low-cost schools than are similar definitions of English words and expressions.

As happens with most content-oriented examinations, an examination culture built up rapidly around the C.P.E. general paper after its inception in 1966. By 1973, teachers could predict with considerable accuracy the topics which were likely to recur from a study of the previous years' questions.

All the three questions quoted above as examples of descriptive items testing knowledge of specialised terms had appeared in similar or closely related forms in at least half the general papers set between 1966 and 1972. The question on the parts of a flower (no. 43), for instance, had been asked in different ways for four consecutive years, from 1968 to 1971. In the four questions, the terms 'stigma' and 'pistil' appeared three times each and the terms 'style', 'anther' and 'ovary' twice each. Not surprisingly, conscientious pupils in schools of all types learned the definitions of all these terms thoroughly. More than 55 per cent of the rural low-cost pupils answered the 1973 version of the question correctly which was only 12.5 per cent less than the proportion answering correctly from Nairobi high-cost schools. Similarly, 59.5 per cent of R.L.C. pupils answered the question on the use of the barometer correctly (no. 41); and 42.6 per cent the question on distillation (no. 42). The N.H.C.-R.L.C. performance gaps for these two items were 9.8 per cent and 6.3 per cent respectively.

A rough count suggests that a candidate who had memorised about 120 to 150 definitions would have been able to answer most of the descriptive items involving specialised terms which appeared in the science section of the C.P.E. general paper between 1966 and 1973. Of course this would be a formidable undertaking, especially for rural pupils who might never have seen a single piece of scientific equipment nor carried out a single experiment, but at least the candidate knew with considerable accuracy which definitions he was expected to learn. Moreover, he probably had a CPE guidebook to help him. At least ten such guidebooks are available in Kenya, and all set out quite explicitly to show C.P.E. candidates how to answer the questions which recur in the examination most frequently. They are revised each year, and any new piece of information or type of problem which appears in a C.P.E. question is quickly incorporated into the text. One guidebook was advertised in the Kenya press during 1977 in the following manner:

With English, however, preparation which focuses on previous papers is much less effective than it is with science and the other subjects of the general paper. We have already discussed the unusual words and phrases tested in the 1973 paper which produced large N.H.C.-R.L.C. performance differences. They included 'scolded', 'tired after' and 'he should have plucked up his courage'. None of these expressions had been tested in any previous examination. Similar examples from the 1974 paper include 'for company', which produced a performance gap of 56.8 per cent, and 'in their teens', for which the gap was 57.9 per cent. The range of possible choices is obviously much wider in the English paper than it is in science. Furthermore, many of the expressions tested in the English paper are essentially idiomatic and hence used more in spoken than in written English. For these reasons, revision from old examination papers and from C.P.E. guidebooks is of limited usefulness. Real competence in English is gained from using the language; in reading, in writing, and above all, in conversation. Because they live in an environment in which English is used not only in formal learning situations but also as an everyday medium of communication, pupils in high-cost schools have a huge advantage.

In summary, then, it seems that low-cost pupils were relatively successful in learning specialised scientific terms because the terms are limited in number, and because definitions of their meanings are accessible. In English, by contrast, there are many more words and expressions to learn, and for rural low-cost pupils access to their meanings is difficult.

While it is true that items testing knowledge of uncommon or idiomatic expressions tend to generate the biggest N.H.C.-R.L.C. performance differences in the English paper, there are at least two other types of questions which also penalise rural low-cost pupils heavily. We shall discuss each of these in turn.

In the first place, rural low-cost pupils have severe difficulties with any question which requires them to appreciate humour expressed in English. One of the comprehension passages in the 1974 paper was a humorous short story, not meant to be taken literally. It told of an Indian who had a pet fish which he kept in a tank. As the fish grew bigger the water in the tank had to be changed more frequently, which the Indian was too lazy to do. So he trained the fish to live on dry land, by taking it out of the tank for longer and longer periods every day. Finally the fish could stay permanently out of the water, and it learned to follow the Indian wherever he went, just like a dog. Fame followed, and many people came to visit them. But one day while on a journey they had to cross a bridge over a river. The fish fell

through a hole in the planking and was drowned in the water below.

Rural low-cost pupils found this story very difficult indeed. Their mean score in the 7 items based on it was only 31.4 per cent or only 6.4 per cent better than the chance guessing level. Nairobi high-cost pupils, by contrast, scored an average of 77.7 per cent, so that the mean performance gap was 46.3 per cent. This is the biggest advantage that Nairobi high-cost pupils have enjoyed in any set of questions in any C.P.E. English paper between 1973 and 1976.

A clue to the rural low-cost pupils' problems can be found in their responses to a question which tested their general understanding of the story:-

		Proportion of Pupils Giving the Answer	
		Nairobi high-cost	Rural low-cost
62.	The story		
	A. Describes how Indians tame fish	6.6%	34.4%
	B. Is amusing but unlikely to be true	73.3%	21.5%
	C. Describes an important scientific experiment	5.6%	25.8%
	D. Shows how clever some types of fish are.	11.7%	15.1%

It can be seen that well over three-quarters of the rural low-cost pupils treated the story literally; they thought it either describes how Indians tame fish, shows how clever some types of fish are, or describes an important scientific experiment. Only 21.5 per cent thought that the story is amusing but unlikely to be true. Pupils in high-cost schools, on the other hand, had no difficulty in recognising that the story was meant to be a joke, and 73.3 per cent answered correctly.

Most of the experience of low-cost pupils in the use of the English language is in serious, learning-oriented situations, where both laughter and scepticism would be inappropriate. English is the medium through which important, examination-relevant facts are communicated; and the task of the pupils is to learn the facts, not to question them. When they want to tell a joke, rural low-cost pupils (and their teachers) nearly always use their mother tongue.

Rural low-cost pupils are also at a severe disadvantage in questions

testing grammar and syntax, where the usage in standard English differs from the usage in Kenyan vernacular languages. For an example we may take a question from the 1974 paper. The candidates were required to fill in the blank in the following dialogue:-

Judith: (Have) you moved into your new house?

Mary: **Yes**, we (moved) last week. You must come and see it.

Judith: Thanks. I'd like _____ very much.

(The brackets enclose the correct answers to the previous two items.)

The answer patterns were as follows:-

		Proportion of Pupils Giving the Answer	
		Nairobi high-cost	Rural low-cost
A.	No word required	7.8%	28.1%
B.	To see	11.2%	53.4%
C.	To	76.7%	5.6%
D.	Seeing	4.2%	11.8%

In English it is, of course, perfectly correct to use to on its own without a verb stem, where it is clear from the context what the verb is meant to be. But in most, if not all, Kenyan languages this cannot be done. In Swahili, for instance, to see is kuona; ku on its own is meaningless. Thus the infinitive must either be repeated in full, or omitted altogether. It is clear from the answer patterns that most rural low-cost pupils attempted to apply the same principles in English as they would in their vernaculars; 53.4 per cent repeated the infinitive in full, and a further 28.1 per cent omitted it altogether. Hence only 5.6 per cent answered the question correctly, which is far below the level which would be expected from chance guessing. By contrast, more than three-quarters of the high-cost pupils knew the correct usage in standard English. For many of them there was probably little or no interference from the alternative vernacular usage because the vernacular is their second language.

We may now turn our attention to the last and most complex of the three questions outlined at the beginning of this section: why does the inclusion of reasoning items in the science section of the general paper penalise rural low-cost pupils rather than reduce their handicap, as is the case with English?

It will be recalled that reasoning items were first included among the science questions in the 1974 examination. When the item analysis report for this examination was written, it was suggested that the relatively poor performance of rural low-cost pupils in scientific reasoning questions might

have been due mainly to lack of adequate preparation. The backwash effects of the new items on teaching methods in R.L.C. schools might well reduce the performance gap in subsequent years. But this hope has not been fulfilled, as can be seen from Table 6.

Table 6. CPE science 1975 and 1976 — mean difficulty levels of the various item types in Nairobi high-cost and rural low-cost schools.

Type of item	C.P.E. science 1975			C.P.E. science 1976		
	NHC	RLC	Difference	NHC	RLC	Difference
Descriptive (non-specialised)	78.63%	67.27%	11.36%	66.79%	54.14%	12.65%
Explanatory	72.10%	53.26%	18.84%	59.70%	44.75%	14.95%
Observation	60.67%	47.86%	12.81%	74.08%	55.47%	18.61%
Reasoning	62.11%	44.90%	17.21%	66.52%	41.75%	24.77%

There is no sign at all that the advantage enjoyed by Nairobi high-cost pupils in scientific reasoning items is being eroded as rural pupils become more familiar with questions of this type: if anything, the performance gap is becoming slightly wider. In 1976 a 'C.P.E. Newsletter' was sent out to all schools explaining, among other things, how the science paper was changing: fewer questions testing memorised facts were being asked, and more questions testing reasoning, observation and understanding. This Newsletter seems to have benefited the Nairobi high-cost schools in their preparation of candidates more than the rural schools.

A clue to a possible reason for these results is suggested by the fact that the N.H.C.-R.L.C. performance differences for the reasoning items in English and science tend to be similar, as may have been noted from previous tables. The full data are set out in Table 7.

Table 7. C.P.E. English and science - N.H.C.-R.L.C. performance differences in reasoning items.

	1973	1974	1975	1976
Verbal reasoning	20.3%	25.86%	16.08%	22.44%
Scientific reasoning	-	18.86%	17.21%	24.77%

It seems clear that reasoning questions, whether scientific or verbal, give N.H.C. pupils a performance advantage of about 20 per cent. Achievement questions in the English paper tend to produce differences which are higher than this level, whereas descriptive and explanatory items in the science paper produce differences which are usually lower. If, for example, we rank the various item

types in the 1976 English and science papers in order of their N.H.C.-R.L.C. performance difference, the result is as follows:-

Table 8. C.P.E. English and science 1976 - N.H.C.-R.L.C. performance differences by item type.

Subject	Item type	NHC-RLC Difference
Science	Descriptive	12.65%
Science	Explanatory	14.95%
Science	Observation	18.61%
English	Reasoning	22.44%
English	Comprehension I descriptive - Harambee project	24.35%
Science	Reasoning	24.77%
English	Paraphrasing	25.41%
English	Comprehension II descriptive - Portuguese forts	27.12%
English	Comprehension III - short story	35.39%

An increase in the proportion of reasoning items in the two papers will simultaneously produce a narrowing of the performance gap in English and a widening of the gap in science.

We may tentatively offer the following interpretation of these results. Far from being relatively immune to the effects of teaching quality, higher level intellectual skills, such as the ability to reason, are particularly sensitive to such effects. In subjects such as science, history and geography, a bad teacher teaches isolated facts. He copies from a textbook onto the blackboard catalogues of names, dates, places and definitions, which the children then copy into their exercise books and learn by heart. Alternatively, he simply tells them to revise from their C.P.E. guidebooks. A conscientious pupil can thus answer descriptive questions in these subjects largely irrespective of whether he has been well or badly educated. A skilled teacher, on the other hand, tries to make facts understandable to his pupils by establishing cause-and-effect linkages between them. He also tries to show the pupils how to create these linkages for themselves through reasoning. A pupil who has not received this guidance will be handicapped in explanatory and, more especially, reasoning items, no matter how conscientious he is at revision.

With English, however, the situation is different, as we have seen. Real mastery of the usages and idioms of English comes only from immersion in an environment where the language is used consistently; and this experience is accessible only to high-cost pupils. The fact that rural low-cost pupils performed relatively better in the verbal reasoning items than in the English attainment items does not mean that reasoning ability is largely immune to teacher quality effects; it simply indicates that the reasoning items are less heavily dependent on language competence.[3]

The interpretation just suggested depends crucially on the assumption that the differences between the Nairobi high-cost and rural low-cost pupils in verbal and scientific reasoning scores are due to school quality factors, and not to differences in home background or even in inherited potential. To investigate this point, two samples of low-cost schools in one district of Kenya have been drawn; one consisting of 16 schools which performed particularly well in the 1976 C.P.E. and the other of 16 schools which performed badly. All the schools in the sample are day schools. The district chosen is fertile and densely settled, and nearly all families grow either coffee or tea as cash crops. Data from another study indicate that more than 80 per cent of pupils attend the school nearest to their homes. The samples are matched as closely as possible by geographic location, distance from urban centres, and types of cash-crop grown, to minimise socio-economic differences between them. (It has not been possible so far to obtain socio-economic data from the pupils themselves.[4]) Each sample consists of about 10 per cent of all schools in the district. The performance of these high-quality (R.H.Q.) and low quality (R.L.Q.) samples in all items of the 1976 C.P.E. is being analysed, to determine from which types of item the R.H.Q. pupils build up their advantage.

---

3. It can be suggested that these factors may also explain the results obtained by Floud and Halsey in their British study discussed earlier. The eleven-plus examination may have selected a higher proportion of working-class pupils for grammar school places when it included an intelligence test, not because the test measured intellectual potential, but because it was less affected by linguistic competence than the composition paper which succeeded it. Working-class pupils in Britain are restricted in their access to the standard (that is, middle-class) version of English in much the same way, if to a lesser degree, that rural low-cost pupils in Kenya are also restricted.

4. It is also relevant to note that schools change their quality status quite rapidly. One school which is classified in the low quality category in this study would have been in the high quality category if the study had been made in 1971. There has been a very rapid turnover of teachers in the intervening five years. The correlation between mean C.P.E. performance in 1962 and 1976 for all schools in the district which offered candidates in both years is less than 0.10.

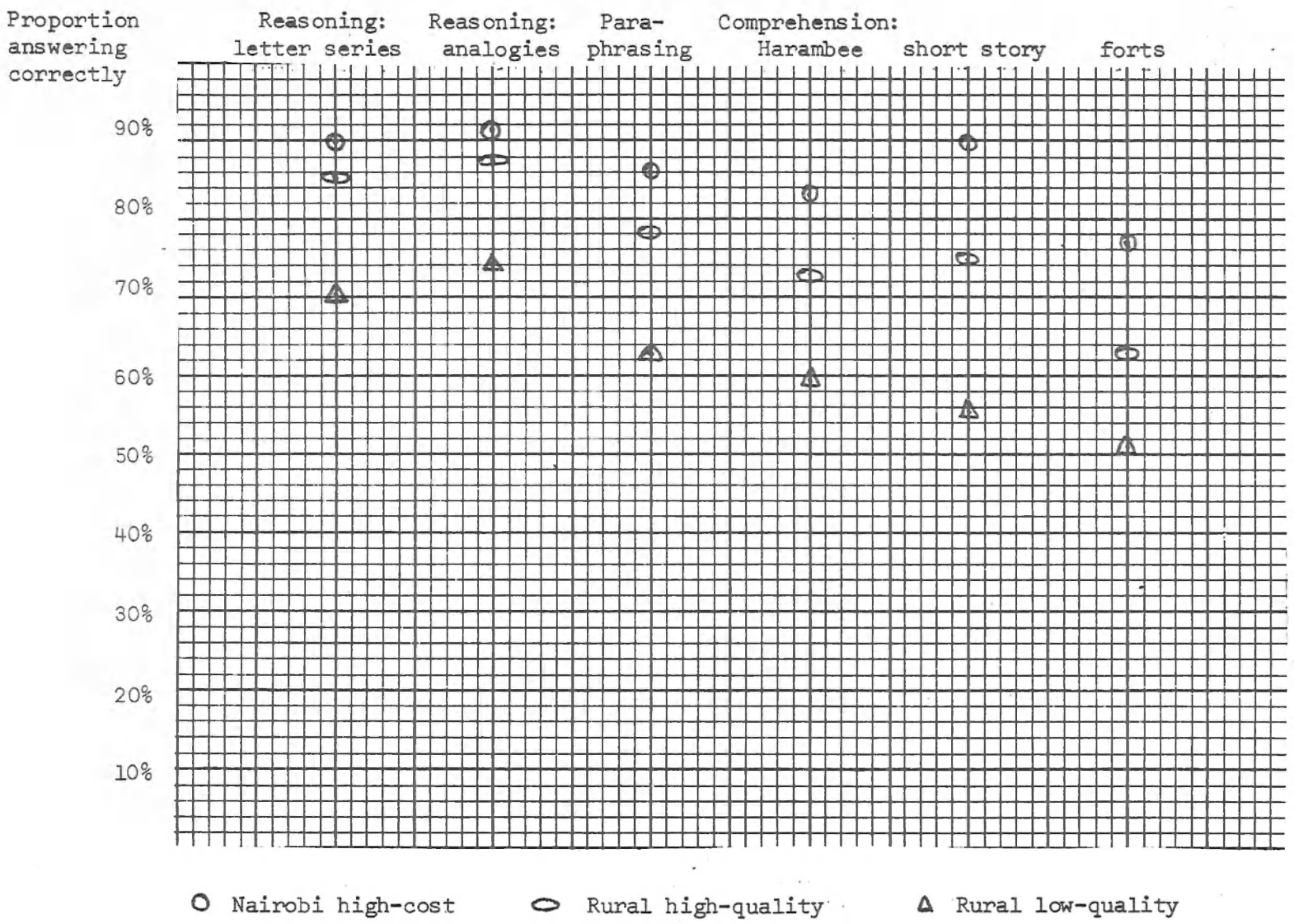
The first result from this analysis is given in Figure 1. It shows the average performance differences between the Nairobi high-cost, the rural high-quality, and the rural low-quality samples in the different types of items included in the 1976 CPE English paper. The N.H.C. means are represented by circles, the R.H.Q. means by ovals, and the R.L.Q. means by triangles.

It can be seen that the performance gap between the rural high-quality and rural low-quality schools is very similar for all types of items. For the reasoning items, the gap averages 13.7 per cent, for the paraphrasing items, 15.5 per cent, and for the comprehension items, 14.0 per cent. Thus good teaching appears to improve verbal reasoning scores nearly as much as it improves English attainment scores; there is no evidence that the pupils from the low-quality schools are compensated for their handicap by the reasoning items to any measureable extent.

It is interesting to note that the extent to which rural pupils from high-quality schools are able to close the performance gap on the Nairobi high-cost pupils is closely related to the degree to which the items test English language competence. One group of reasoning items, consisting of letter series, required virtually no knowledge of English. The alphabet was provided, so that candidates who did not know the correct letter sequence were not penalised. In these items, the rural high-quality pupils were only 3.9 per cent behind the Nairobi high-cost pupils. Similarly, in the analogies items, which required only knowledge of everyday words, the N.H.C.-R.H.Q. gap was 4.1 per cent. But in the comprehension sections some of the items required a command of the English language which no rural low-cost school can provide, however competent its teachers. In these sections, the mean N.H.C.-R.H.Q. performance difference rose to between 14 and 16 per cent. These results suggest strongly that the huge performance advantage enjoyed by Nairobi high-cost pupils in the English paper can be ascribed entirely to two sources: first, their greater familiarity with the language; and second, the superior quality of teaching they receive. There is no evidence whatever to suggest that N.H.C. pupils have superior capacity for abstract thought.

It is hoped to extend this analysis to other subjects, with the aim of separating the effects of school quality on C.P.E. performance from the effects of socio-economic background factors. But the data are complex, and must be the subject of another paper.

Figure 1. CPE English 1976 -- Mean difficulty level of items for Nairobi high-cost, rural high-quality and rural low-quality schools.



REFERENCES

1. Floud, Jean and Halsey, A.H. 1957. 'Intelligence Tests, Social Class and Selection for Secondary Schools'. British Journal of Sociology. 8, pp. 33-39.

2. Gakuru, O.N. 1977. 'Pre-primary Education and Access to Educational Opportunities in Nairobi'. Working Paper No. 321. Institute for Development Studies, University of Nairobi.

3. Lippman, Walter. 1922. 'The Abuse of the Tests'. Reprinted in N.J. Block and G. Dworkin. The IQ Controversy. New York, Pantheon, 1976, pp. 19-20.

4. Somerset, H.C.A. 1974. 'Who Goes to Secondary School? Relevance, Reliability and Equity in Secondary School Selection'. In David Court and Dharam Ghai, editors. Education, Society and Development: New Perspectives from Kenya. Nairobi, Oxford University Press.