MULTIPLE CHOICE TESTS AS AN ASSESSMENT TECHNIQUE FOR HUMANITIES AND SOCIAL SCIENCES

Paper read at a seminar on:

"A background to University teaching in Zambia" (May 2 3rd-24th, 1972)

by Wim M.J. van Binsbergen

ABSTRACT

Objective educational tests, especially the multiple choice (MC) technique seem at least as valid and reliable as more traditional types of assessment; they save time, in the long run; and we have computers to deal with them. Still, MC tests are not being used in our HSS, -apparently for the following reasons: most lecturers are unfamiliar with the technique; they distrust it because, allegedly, it would only be relevant to rote learning and not to -'understanding"; and many of them consider essay-writing to be the main method both for training and for assessment.

By means of a pilot experiment (december 1971, involving 46 first year sociology - S110 - students) I tried to show that MC tests can actually be useful in (the first few years of) undergraduate sociology teaching; with implications, of course, for related fields in HSSi

The experiment was subjected to three for&s of evaluation:

a. Educational aims. The experimental test paper in its totality proved to assess the' candidates[1] capabilities according to no. less than *7* out of the 8 explicit aims of our S110 teaching; the evident exception being students' writing skill. The experiment made it clear that quite difficult and fundamental problems of the social sciences (at least at first year level) can successfully be phrased into MC questions.

b. Comparison with other study results. Statistical analysis (as against the examination grade and the'final course grade) showed that the experimental MC test was, as an assessment technique, fully comparable to essay-type assignments.

c Internal numerical evaluation of the question paper showed that the discriminatory power of most questions was satisfactory, and pointed out, at the same time, which questions were less adequate.

Details are described in a separate report, copies of which will be available at the seminar. This report includes the testpaper, analyses and evaluations; it gives many clues as to how a MC test suitable for HSS might be constructed, analyzed, evaluated and improved, and it pays some attention to such problems as guessing, marking and professional consensus.

If seriously and skilfully handled, MC tests can form, a good assessment technique, for HSS as well.They can favourably complement (though not entirely supersede) essay-writing. Once a corpus of adequate questions will be formed (a time-consuming process), a MC test can easily cope with any number of candidates, at the same time ensuring greater objectivity than essay-type assessments (within one MC test many more topics can bo covered; the candidates' performance is judged with an impersonal answer key; the MC test **easily** provides[1] a general yard-stick for all students in a course - even ii they are taught by different tutors; inadequate questions are exposed by the answer analysis; the candidate cannot waffle hinself out of the problem).

The sociology department has now decided to incorporate MC tsting in the S110 curriculum for 1972-73; thus within a year we shall:be able to make more definitive recommendations about the advantages and limitations of MC tests for HSS.

THE UNIVERSITY OF ZAMBIA
SCHOOL OP HUMANITIES AND SOCIAL SCIENCES

MULTIPLE CHOICE TESTS AS AN ASSESSMENT TECHNIQUE FOR HUMANITIES AND SOCIAL SCIENCES

May 197 2

Wim M.J. van Binsbergen
Sociology

Contents

1« Introduction.

During the last few years, much attention has been given to the development of so-called "objective educational tests", among which the multiple choice technique is the most well-known and most widely used*

There are various reasons for this preoccupation :

The rapid increase of numbers of students in secondary and tertiary educational institutions created many practical problems. Traditional forms of assessment (such as essay-writing, oral examinations, and - in case of

mathematics and the natural sciences - problems to be fully written out • by the candidate) are increasingly diffioult to handle for the teaohingb staff.

Moreover there is increasing awareness that the aims and goals of formal education should be explicitly formulated and classified, and that the actual teaching should be rationally directed towards these objectives. This development stimulates questions to be raised as to the validity and reliability of current educational test techniques (including traditional forms of assessment); modern social sciences methodology, and statistics, are more and more used to tackle these problems.

Thirdly, the growth of technology and of applied mathematics.enables us now to assign complicated, hitherto time-consuming, tasks to electronic computers.

Of course, on top of all this, a certain element of snobbery and technological magic flourishes. But multiple choice techniques are are not simply good because they are fashionablel

There is a growing body of international literature on all these and related problems. Meanwhile, multiple choice tests are being used widely, in our University as well (notably in Natural Sciences). Having observed the use of multiple choice tests in the University of Amsterdam (the Netherlands) in such departments as Sociology, Anthropology and Physiology, and realizing what possible gains could be derived from an introduction of this technique in our School of Humanities and Social Sciences, I gladly accepted the challenp'" of my colleagues in Sociology and constructed an experimental multiple choice test.

This paper reports on the experiment. It describes and evaluates the test I administered. In general, it explores the advantages and limitations of multiple choice tests within the School of Humanities and Social Sciences. And finally, i t may give some hints about the construction of this type of tests, for those amongst my colleagues who are willing to be persuaded that,

however limited, multiple choice tests provide a useful addition to more traditional forms of assessment. Although the present paper, gained much from extensive discussion with my colleagues in Sociology, responsibility for it s contents is entirely mine.

2. What is a multiple choice test?

The essence of a multiple choice test is that the candidate is offered, for each question, a choice out of a.number of fully specified alternatives. Usually only one alternative is the right answer to the question, but all alternatives should look more or less acceptable, attractive, to somebody who has no full knowledge and understanding of the topic. .

By varying the number of choices, and by ingenious coding, on the question paper, of combinations of alternatives, many types of multiple choice tests can be generated.

In choosing a particular type of test, various considerations are relevant. We must keep within the limits of the program and of the implementation of the computer; this point only very seldom creates difficulties. It is however, difficult to find attractive, but clearly false alternatives -a reason to keep the number of alternatives low. The candidate might try to guess the right alternative - a reason to keep the number of alternatives high (see below, 7) . And finally, the question should present technical difficult" with regard to logic, reasoning, grammar and style only in so far as we want to measure the candidate[1] s ability to handle these problems within the framework of his subject proper; we do not want to test, as such, the ability to answer multiple choice questions - a reason to refrain from confusing set-ups like multiple negations, coding of combinations of alternatives, etc.

In view of these considerations, many people working in this field hold that the most satisfactory ultiple choice test has four alternatives, one of which is the right one, and lacks too complicated phrasing and coding. The test much contain a fair number of questions, (all of which to be answered by each candidate), and for proper evaluation the number of candidates sitting the test should not be too low.

-

3. Description of the experiment

The test was designed for S 110 students, 1771.

In view of the experimental character of the test, participation had to be entirely voluntary. However, the experiment could only be evaluated if a sufficient number of students would participate, i f they would be sufficiently motivated, and if, by revising their notes etc., they would prepare seriously.

The students themselves created favourable conditions in these respects, by asking (begin November) for a "mock examination", to precede the final examination in January 1972. When I presented the experiment as a "mock examination", of my four tutorial groups (about 80 internal S 110 students), three groups were enthousiastic to participate in the experiment, once the principles of a multiple choice test were explained to them.

To avoid time table clashes, the test was administered during the regular tutorial hour of each group. In the three sessions, a total of *k&* students sat the test. In each session, the first 10 minutes would be spent on instruction (with the aid of an instruction sheet, appendix I)• Each candidate received a computer card, on which he himself would punch his computer number, test number, and the number of the alternative chosen at each question. After the students would have started on the computer card, I would go around, to check whether everybody had understood the instructions, and to help individually i f necessary. Soon all students could cope with this new task. Virtually all students were able to complete the test (15 questions) within 30-2*0 minutes. To avoid the questions leaking out too much, test papers were handed back after the test, along with the computer card and the instruction sheet.

The questions covered significant topics out of the whole S 110 syllabus as taught in 1971. I had not myself lectured this course, ao I derived my question topics from:

a. difficulties as shown during my own tutorials, and in the related assignments.

b . hand-outs distributed in tutorials.

c, the written correspondence lectures differing, of course, i n many
details from the oral 1971 lectures),

d. the required textbooks for S 110, 1971»

From this material a rather long list of possible question topics was compiled. For each topic I attempted to construct a question. Many attempts failed, for a variety of reasons? the question would be too difficult; too easy; I failed to find three false but persuasive alternatives; I failed to find one alternative that was clearly the only right one; the question would hardly be relevant to the aims of our teaching in S 110, etc. Drafting, editing and selecting questions demands many hours and some measure of inspiration. As it was only an experiment, and as the question paper should be covered within one hour, I confined myself to 15 questions - which is actually rather few for a multiple choice examination.

The fact that I had no full details about what had been taught at the lectures, made that the test contained one question (iii) about differential fertility" a sociological notion with which the students happened to be unfamiliar at the time of the test.

4. Evaluation in terms of educational aims.

Let me try to outline the .goals and aims S 110 as they are clear from the present syllabus. Very briefly, the student who has successfully completed S 110 can be expected to have the following characteristics:

He (respectively she) is able to read, to understand, to summarize and sometimes to evaluate, basic social science texts in English.

He has some understanding of the scientific approach and method in general; of the place of sociology in the taxonomy of sciences; of the characteristics and limitations of sociology.

[p. 6]

He knows some specific core concepts of sociology, has an awakening awareness of the interrelations between these concepts, and is familiar with of a number of theories in which these concepts appear,

Using these concepts, he has acquired familiarity with some descriptive sociologicaluterature concentrating on Africa, and on Zambia in particular.

He is able to show his understanding by applying his sociological outlook and concepts in the analysis of selected social stituations. .

He has had a first introduction to sociological research methods.

He is, to some extend, able to write about what he read, heard in the lectures and tutorials, and about his own analysis. Doing so, he uses clear and reasonably correct English, and he makes efforts to structure his argument in a consistent way.

If we agree that this sums up the aims of S 110, then a multiple choice test can be considered to be useful and valid if it is pertinent to at least some of these major points. The same reasoning holds true for the other courses and subjects in the School of Humanities and Social Sciences, even when their objectives differ rather widely from the ones outlined here

for S 110.

It is clear that a multiple choice test does not assess a student's skill in writing, and does not allow him the space, the freedom, to show off all the read and knows, like in an essay. On the other hand, hardly anybody'would claim that a multiple choice test is unable to assess knowledge of facts, and even knowledge of the definitions of important concepts. But to be a satisfactory student requires more than just-,this knowledge - as is clear from the above list of aims. .

Now, a common allegation, especially among teachers working in the field of humanities and social sciences, is that a multiple choice test is completely unable to assess a student's understanding. The use of this technique would, allegedly, be confined to rote learning.

My main challenge, when conducting this experiment, was to refute this allegation. My test paper included a number of questions that have hardly anything to do with sheer memorizing but which, on the other hand, directly appeal to the much more comprehensive intellectual capabilities that are the main objectives of academic training.

Table I is a, necessarily subjective, attempt to estimate the relevance of the question paper to the aims of the course as outlined above. I do not think that I have to explain extensively the considerations behind the ratings shown (+ means: rather relevant;*means: rather irrelevant). Oneexample will suffice: question vi i is relevant to the items reading (a key word being "completely" in the second altenative); concepts and theories (the concept of "open", vs. "closed" "systems of social stratifications") ; description (contemporary South African society); "the other items being practically irrelevant in this case.

If one accepts the ratings in table I as more or less correct, then i t will be clear that the only educational aim which i s not covered at all by the present test is: writing; and that hardly any question"'if at all, is entirely a matter of memorizing. Quite difficult and fundamental problems, such as the nature of science, of a definition, of a pair of ideal types, can apparently be framed into a multiple choice test. And this with at least one great advantage over the essay questions: the false but rather attractive alternatives deprive the student from the opportunity to talk himself out of the problem - an opportunity so often misused in the case of traditional types of assessment.

5 Evaluation i n the view of other study results obtained by the same students

Assuming, for convenience, the validity of more traditional forms of assessment, we can investigate the validity of the present multiple choice test by comparing the test results with other study results (grades) obtained by the same students.

Before we can make this comparison, we have to assign a certain numerical value to the test results. Here the following questions can be raised:
Should the same weight be given to all right answers?
Should the candidate be penalized for unattempted questions?
Should the candidate be penalized for wrong answers?
If we choose to penalize the candidate for a wrong answer, should the same weight be given to all wrong answers?

A few examples may illustrate the relevance of these points. We might feel inclined to give more credit for right answer i f this involved a considerable amount of understanding (e.g. questions iv, vi), instead of just the knowledge of a definition (question iii) , or the handling of a simple concept (question 3 i> An unattempted question may either mean that the candidate does neb know the answer (in which case he would have been penalized under the traditional forms of assessment), or that he was short of time (which perhaps should not be penalized). In questions where reading or comprehensive understanding are important, choosing a certain false alternative may have very wide, and gloomy, implications, and we might want to penalize severely (e.g. vi. 1, xi.4)j but in other cases choosing a wrong alternative may be mainly a matter of a lapse of memory (ix, iii), not revealing so much about a student's probable future within the social sciences, - and in this case we would penalize mildly, if at all.

The multiple choice program of the UNZA Computer Centre allows for all these considerations to be built in. By assigning a certain numerical value to each answer, and adding these values, a percentage distribution and a ranking of the total performance of the candidates can be made. Of course, different solutions of the above questions produce different percentage and ranking distributions.

However, i t is extremely difficult to weigh all alternatives properly, and to express this weight numerically. The weight should not only depend on the objective difficulty of the question, but also$_0$n the amount of attention given to the problem concerned, in lectures, tutorials, readings and assignments; the attractiveness of the alternative' their phrasing, etc. For this reason, and in order not to compel3ate the experiment too much. I assigned the value 1 to all right answers, and

the value 0 to wrong answers and to unattempted answers

The final course grade is meant to be the best single representation of the degree in which- a student's following of the course has been successfully Therefore, a student's performance in the multiple choice experiment should primarily be compared with his final course grade. The rank correlation (Spearman seems the best approach here. Hovever, before we are able to draw sornd conclusions, it is wise to relate the final course grade to still other study reasults of the students concerned, within the same course. The final grade is the resultant of a number of traditional-type assessments: the grade obtained in the examination of January 1972 (counting for 60 - 70$)> and the continuous assessment average (counting for ***kO*** - 30$, and derived from the marks for nine tutorial assays written throughout the academic year 197j.» It would highly strengthen the case for the multiple choice technique, if our quantitative analysis would show that the present multiple choice test, despite its experimental character, is about as closely related to the final course grade, as.some of the other forms of assessment.

Table II gives, for the sample consisting of the *kB* candidates sitting the multiple choice test, the rank correlation coefficient (RS) and'"in order to assess the significance of RS _,the associated t value, between the final course grade on the one hand and on the other, successively:
the examination grade;
the continuous assessment average;
the mark for essay no.7 (on traditional systems of social stratification and rank), due September 1971?
the mark for essay no. 9 (on tribalism and national integration), due November 1971;
the percentage gained in the multiple choice test.

The two essays are chosen arbitrarily out of the total of nine. For further comparison the above mentioned grades are also related to the examination grade (table II column 2); here essentially the same pattern shows.

Admittedly the present material is far from ideal for such a relatively-simple approach. The various grades and marks form a seriea on a time axis, and reflect the continuous development of the students - a development that in mjmy cases is highly influenced (for "better or worse) by the very fact that a student knows his. previous grades. Hcweover, much of the correlations shown is purposely built-in: deliberately a great weight is given in the final grade to the examination grade, and some weight is given to the average continuous assessment grade. So we can hardly bo surprised that positive correlations occur here! The remarkable fact is, on the contrary, that in three out of four cases (row entries 2, 3 and 4**,** table II) the correlations although significant, are still rather small. This appears to be due, largely, to the lack of consistency in the performance of the individual student. However, the multiple choice test did not count towards the final grade, and is relatively independent from the other variables; therefore the present analysis is rather justified.

Table II now shows that the multiple choice test can well compete with a tutorial essay, as a measure of a student's capabilities (as shown by the final course grade). We might paraphrase the outcome of the analysis thus: a student who wrote reasonable essay no»7 was not much surer of finally passing the course than wos a student who did reasonably well in the multiple choice test; and the latter had slightly better prospects of passing the course than had student who wrote a reasonable essay, n. 9 This is a solid argument for the validity of the multiple choice technique as a method of assessment.

Of course, the encouraging result does not "prove" that a multiple choice test is as good as, or better than, essay writing. If we make students write essays, we do not only want to assess whether they have acquired, through reading and revising, the expected knowledgod and understanding (for this a multiple choice test might do); we also want to teach them how to write about the topics covered in the course - and this a multiple choice test can never do. But should we really always combine this training in writing with the assessment (including formal marking) of knowledge and understanding? A combination of multiple choice tests and more technically-oriented training in essay writing seems preferable.

6. Internal numerical evaluation of the question paper.

The preceding two sections aimed at demonstrating the validity of the multiple choice test in view of considerations extenal to the tost itself: educational aims, and other study results. So far the argument tries to convince ths reader that multiple choice tests can be worth while; but it does not indicate how one multiple c oice test can be better than another, and how we might improve a certain test. This is the question of internal evaluation.

For this purpose, the computer program produces a num erical answer analysis, given below as table III.

In this table, DF stands for "discriminatory factor." This is a numberical index for the degree in which a particular question distinguishes between [l;]good" and[5] we ale[1] candidates. The DF of a certain question is iigh, i f candidates who throughout the whole test obtained a relatively high score, tend to answer this particular question right, and, alternatively, if candidates who throughout the whole test obtained a relatively low score, tend to answer this particular question wrong. People working in this field developed quite sophisticate' types of D F; however, the DF in the present program of our Computer Centre is extremely simple: the number of correct answers by candidates in the bottom third of the class is subtracted from the number of correct answers by candidates in the top thi' of the class and the difference is divided by the sum of these two numbers and then multiplied by 100 Thus, the factor can go from - 100 to + 100. Generally speaking, the more positive the factor for a question the better it is " (Memorandum, Computer Centre) .

It will be notieed that this approach does not take into account the actual number of candidates answering a question right; It treats as equal all candidates in the top thirdrespectively in the bottom third; it ignores the performance in the middle third; it does not assess statistically the significance of the difference bifeween the top third and lottom third; it does not use the information acquired, to omit the worst questions (the ones that hardly, or even negatively discriminate) from the final computation of the percentage grade, and rank, of the candidates.

Anyway the present approach still gives some idea about the quality of our test. In tableIII, DF ranges from + 100 to + 13, no DF being negative. This is rather reassuing.

Until, however, a more sophisticated DF is adopted, it is wise to look at the actual distribution of answers on each question (Table III; the right answer is indicated by +). How does this distribution help us to evaluate our test paper?

In view of our rule that, per question, only one alternative is right, but all alternatives should be more or less attractive, we can apply the following criteria in evaluating multiple choice questions:

No alternative should be rejected by all candidates (score 0 on this alternative in Table III) .

No alternative should be chosen by practically all candidates (score on this alternative in Table III tending to 46).

According to these criteria, the questions vii, viii, ix, xii and xv are deficient. These happen also to be the ones with the lowest DP s; the explanation for this seems to be that, i f one alternative is already completely unacceptable, or, respectively, if one alternative is very highly attractive, the question becomes so easy that even many weak candidates can answer it right. Probably, these questions can be improved by minor alterations (different wording, substitution of a few different alternatives).

Naturally, the four alternatives to a question can never be completely equally attractive. However,
c. a question can be considered as satisfactory if at least two alternatives get each a reasonable amount of scores (say: at least 20$),

This is the case for the remaining 10 questions(all having DF'S $^{A}kO)$ In 9 out of these 10 cases, the right alternative is ranking among these two or three much - chosen alternatives. This is a desirable situation, at least if we assume the following* that the majority of our students can live up to the level of this University, that our teaching is more or less adequate, that the candidates prepared seriously for the test, and that the wording of the test paper minimizes needless technical difficulties (see above,
2). The only exception is question v. The explanation here is that very few candidates wore aware of the meaning of the word "cyclical" used in the right alternative; substitution of a synonym will improve this question.

7. Three further problems.

Without aiming at an exhaustive discussion, I now want to draw attention to three further problems: guessing, marking, and the laek of professional consensus.

Guessing. Even if we tell candidates not to guess, we can be pretty sure that most of them guess some of their answers, including some right answers. Hence the demand that all alternatives should be roughly equally attractive: only so we can assume that guesses are about evenly distributed over all alternatives. If this condition is fulfilled, the average "expected value" a candidate can gain by mere guesswork will be

E_{av} = 100% / number of alternatives per question.
(in the present experiment, E_{av} = 25%). Theoretically, a candidate might answer all answers right by mere guessing - but the probability of this happening

is completely negligible. Still an unusually lucky candidate might get, by mere guessing, considerably more than E_{av} .How much he- gets at most dsrend on the number of questions in the testpaper and on the number of alternatives per question' a good estimate is provided by the upper confidence limit in a binomial distribution, e.g. on the 95% level .

Guessing could be discouraged by imposing (or just feigning to impose) penalties on wrong answers. In general guessing does not appear to be a serious problem in the multiple choice technique, provided that the number of questions is not too small and that the number of alternatives per question is at least 4.

Marking. The computer program gives for each candidate the percentage (actual credit / total possible credit) and his rank position as compared to other candidates. It has a complicated scaling procedure to convert these values into regular grades. The computer differs from the average lecturer in that the former is not reluctant to give many fails and very high pass-marks. So we have to consider whether the rather narrow distribution of grades (lecturer) or the much wider distribution (computer) reflects more adequately the actual distribution of abilities and performance among the candidates. In view of this problem and in view of the fact that all nuestions (regardless of their DF) contribute in the same way to the final mark, it seems advisable not to leave the actual grading to the computer, but to convert personally the percentages into grades, after inspection of the computer output. This can be done in a few minutes.

Professional consensus. By their very nature, the humanities and social scienoes display a much wider range of acceptable professional approaches than, e.g., physics and chemistry. Many problems in our field are still a matter of the appreciation of individuals or "schools", rather than of general acceptance. In sociology, this is especially clear with regard to methodology and the epistemological status of the profession. Thus the problem arises that not all socialogists, however well qualified, would agree about the correct answer to certain questions (e.g. vi). Although this is a tantalizing problem at the higher professional level, in practice it appears to be hardly relevant for composition of multiple choice questions for first or second year students: for these we do not present with all existing different approaches, but give them only a selection of those viewpoints which we consider both valid and easy enough to grasp. It is in those first years that the multiple choice technique seems particularly appropriate. In the third and fourth year of undergraduate study, and ***a*** fortiori in postgraduate study, we would like to see students operate in a more independent and creative frame of reference than provided by multiple- choice assessment - but there the marking of essay- type

assignments does not exercise the same kind of pressure on the teaching staff, and is more intrifcsically rewarding for the lecturer*

8. Conclusion

Multiple choice tests appear to be a good assessment technique, if seriously and aptly prepared, analyzed anr[n] evaluated* The use of this technique is mainly in the first and second year of acadomic teaching. Whan supplemented by essay-writing, this technique appears to have the following positive aspects:

a. Teachers are stimulated to make up their minds still more explicitly about the aims and actual content of their teaching

b. Greater objectivity is reached, because within one test many more topics can be covered, at various levels' of knowledge and understanding , and the one - sided emphasis on writing skill is absent..

c. Greater objectivity is reached, because a candidate's performance is judged not by individual teachers (subject to human sympathies, antipathies, ideosyncrasies, moodiness, fatigue etc.) but by an impersonal apparatus, using a fixed answer key. This point is notably relevant for the common situation in which the students in a large course arc divided into a considerable number of tutorial groups, most groups being tutored by a different lecturer. When the continuous assessment entirely consists of essays, it is extremely difficult to harmonize the differ nt standards since like of the lecturers involved; especially since in a University like ours, with its lack of continuity in teaching staff, consensus about marking is unlikely to evolve. Incidental multiple choice testing of all students in the course can be very helpful as it provides a general yard-stick throughout all tutorial' groups.

d. Greater objectivity is reached, because inadequate questions (which use to haunt both multiple choice and essay-type question papers) are exposed by the answer analysis.

e.⁝ Although it takes much more time and effort to prepare a multiple choice test than to prepare a traditional essay-type question paper, hardly any time is spent on marking; the test can cope with a nearly infinite number of candidates; if circulation of the questions among students is prevented, adequate questions, once prepared, can be pooled and used at more than one occasion. All this would be to the benefit especially of large courses (notably first year courses, such as A110 !) .

In view, among other things, of these considerations and of the experiment as a whole, the Sociology subject decided to follow up the present experiment by incorporating regular multiple choice testing in the S110 curriculum for 1972-73, and by reducing the number of essay-type assignments. Thus i t will take long before definitive recommendations can be made about the advantages and disadvantages of multiple choice assessment in sociology and related subjects in our University. By that time, however, our discussion will have a sounder basis than the present, necessarily extremely limited, report.

Meanwhile I would suggest that a study group for multiple choice techniques be created within the School of Humanities and Social Sciences.

	i	ii	iii	iv	v	vi	vii	viii	ix	x	xi	xii	xiii	xiv	xv
a. reading	-	+	-	-	+	+	+	-	-	-	-	+	-	+	-
b. science	-	-	-	-	+	+	-	-	-	+	+	+	-	-	-
c. concepts and theories	+	-	+	+	-	+	+	+	+	+	+	-	+	+	+
d. description	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-
e. own application	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-
f. research methods	+	-	-	+	-	-	-	-	-	-	-	+	-	-	-
g. writing	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Table I. Content analysis of the question paper in terms of educational aims.

	Final course grade		Examination grade	
	r_S [1])	t [2])	r_S [1])	t [2])
Examination grade (January 1972)	+0.90	13.83	+1	∞
Cont. assessment grade (March-November 1971)	+0.62	5.10	+0.41	2.98
Essay 7 (trad. Ranking systems), due September 1971	+0.52	4.05	+0.38	2.72
Essay 9 (tribalism and national integratioon), due November 1971	+0.33	2.31	+0.29	2.00
Multiple choice December 1971)	+0.42	3.08	+0.36	2.56

Table II . The final course grade and the examination grade as related to performance in multiple choice test and in other assessments.3)

1) corrected for ties

2) N.B.: $t_{N=46;\ 95\%} = 2.41$

3) cf. Siegel, S, n.d., Nonparametric statistics: for the behavioral sciences, McGraw-Hill / Kogakusha, p. 202 ff .

	alternative					
question	1	2	3	4	blank	DP
i	19+	6	:7.	12	2	+83
ii	10	7	4	24+	1	+71
iii	10	11 +	3	16	6	+100
iv	19 +	3	4	15	5	+83
v	27	9	6+	4	0	+50
vi	6	14	11	15+	0	+40
vi i	26 +	12	0	8	0	+20
viii	5	4	28+	8	1	+26
ix	7	5	31+	3	0	+13
x	14	3	24+	4	1	+46
xi	17	22 +	2	5	0	+47
xii	1	5	0	40+	0	+15
xiii	2	15	24+	4	1	+50
xiv	20	16 +	4	4	2	+45
xv	1	34+	0	9	2	+23

Table III. Answer analysis

APPENDIX I

INSTRUCTIONS.

Write your NAME on the. lefthand edge of the punch card, where you see the words "printed in USA".

Punch holes can be made with a ball point pen or pencil, by pushing.

Punch your computer number in the first 4 columns of the cards (columns 2-4-6-8).

N.B. ADULT EDUCATION STUDENTS do not punch in these first 4 columns, but put A E behind their written names.

4-. All students punch : 1 in column 10

For every question in t e question paper, punch the right alternative (l ¸ 2. 3. or 4) in the column (12, 14, 16 etc.) indicated.

Each question has only one right alternative. Read the question well: sometimes the right alternative means a wrong statement, depending on the question. Per column, only one out of the numbers 1, 2, 3, 4, should be punched. Per column, do not use the numbers 0 and 5-9.

DO NOT GUESS, But always try to make the best possible choice. First eliminate alternatives that are obviously wrong; then decide which one, of the remaining two or three, is the best one.

Before punching your choice, you must be pretty sure about your answer. If after punching you want to change your mind, do the following:

draw a circle around the wrong punch hole

punch the right hole

fill the wrong hole with one of the loose chips, push with your nail, and it will stick.

Proceed rapidly and try to anwer as many questions as possible.

After completing the test, turn your punch card and remove all the loose chips from the back side. Otherwise these will damage the computer.

10. After the test, return: PUNCHCARD, TEST PAPER, INSTRUCTION SHEET

Successl

APPENDIX II .

TEST PAPER

column 2-4-6-8 : your computer number column 10 : punch 1

column 12: question I

The table below gives the composition of a certain village, as to number of persons per household:

number of persons number of such households

per household

2	2
3	3
6	1
7	3
9	2
10	2

Total 13 households

In this village, 6 persons per household is

1 , the mean size of the household

2 the modal size of the household

3 the median size of the household

4-. alternatives 1,2,3, are all false.

column 14 question II .

Which one of the following statements about Ankole society (Bahima, Bairu etc.) is NOT true?

1 Bairu were exposed to the extortion and military intimidation by Bahima: for Bairu, the only means of protecting themselves was by direct appeal to the Ankole king.

2 Although formal marriages between Bahima and Bairu did not occur Bairu women lived as concubines with Bahima men.

3 No one born from Bairu parents could ever rise to become a member of the Bahima group, and (e.g.) own productive cattle.

[p. 21]

4. In order to preserve their position as an elite, the Bc.hima were completely endogamous, not contracting marriages with any other group within the Ankole Kingdom.

Colum 16: Question III

Differential fertility is:

1 The process by which genetic variation occurs within the hereditary characteristics of a certain population, as a basis for the emergence of different races.

2 The difference in birth rate between various classes in a society, allowing members of certain classes to fill positions for which other classes were no longer able to provide the personnel.

3 is the tendency, in most human populations, for women to outnumber men by 2 to 5%**.**

4 the general term, under which we discuss the characteristic difference in demographic structure between, respectively, traditional societies, transitory societies, and modernized societies.

Column 18: question IV.

Investigation 500 rural immigrants in a certain town, a sociologist established a correlation (correlation coefficient = +0.91) between income and duration of continuous residence in town. Which one of the following is the only RIGHT conclusion?

1 There is some connexion between income and duration of residence, but further analysis is needed as to the nature of this connexion.

2 The level of income is the cause of duration of residence (for only the successful immigrant, who is able to improve his income, will stay In town)

3. No valid statement whatsoever about the relation between income and duration of residence can be made.

4. The level of income is the effect of duration of residence (for living in town the immigrant gradually acquires both skills and contracts that warrant him a better income).

Column 20 : question V*

"In sociology we define a group as a group of persons who interact with one another and who share common valuesi " This definition is UNACCEPTABLE ***i***

1. Because it fails to indicate that groups have certain goals, reflected in the common values of the members, and imposing upon the members specific types of interaction.
2. Because it fails to draw attention to a number of different approaches to the concept of " group" among sociologists and social psychologists .

3. Because it is essentially cyclical in defining something by itself

4. Because it gives too much emphasis to the role of values in the analysis of social groups

Column. 22. question VI .

"Modern empirical sociological analysis has proven that Weber's concept of status group (Stand) is true to such an extent that it should be taken as a starting point in all stratification studies."

1. This is a sound statement, because alternative concepts in this field (such as Marx's concept of class) have been refuted.
2. This is a sound statement, because status is the actual moving force of individuals in society.

3. This statement is unsound because conflict based on different relations to the means of production is proven to be the dominant feature of social stratification.

4. This statement Is unsound because concepts are entirely arbitrary, and cannot be shown to be true or false.

Column 24 ; question VII .

Looking at the strafideation in contemporary South Africa, which one of the following conclusion is the BEST one?

1. It is a combination of closed system based on colour and a rather open system based on class.
2. It is a closed system, for the position every individual occupies within society is completely determined by the race he is born into.

[p. 23]

3. It is essentially an open system because there exists an incessant flux of both imnigrantsand emigrants across the frontiers of the country.

4» The system is entirely closed because the principle by which it is governed are formalized in elaborate laws (imposing very severe sanctions) , ideology (Apartheid), and religion.

Column 26: question VIII.

Consensus in sociological writings means:

1 an inventory of the members of a certain community under study, as the basis for further sociological inquiry.

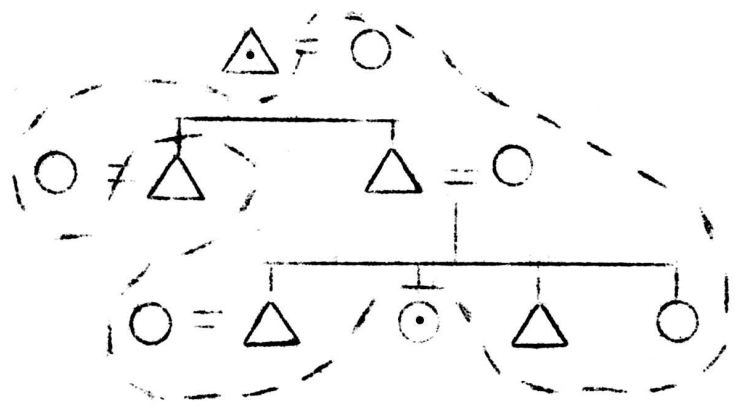
2 the recurrent activity by which citizens of a state take part in the political process.

3 The common adherence to certain more or less fundamental values, by virtually all members of the social unit under study.

4» The allegiance of subjects vis-a-vis their rulers, especially as found in traditional political systems.

Column 28: question IX.

The following persons are living with a in his homestead: his mother, wife, widow of his brother (not remarried), an unmarried daughter, an unmarried son, a married son and the latter's wife. (See diagram below; the dotted symbols represent deceased persons) .



Now determine whether these people in A's homestead constitute:

1 compound family 3. an extended family

2 a unilineal descent group 4. neither 1, nor 2, nor 3

Column 30: question X:

Comparing the modern sociological approach to the state with Marx's view of -the state, we conclude:

1 . Marx claimed that the state was the main instrument in unifying the mass of workers and bringing about the classless society; whereas the modern approach is more critical of the state as an institution.

2. The modern approach is mainly an elaboration on the central question in Marx's approach: the legitimacy of violent force.

3. Marx viewed the state as the direct reflection of production relations, whereas in the modern view the state is a largely independent institution which cannot be completely explained by simple reference to just one type of structural relations outside the political sphere.

4. Marx's approach to the state turned out to have some limited relevance: modern political sociology applies it to communist countries only.

Column 32: Question XI:

Toennies' paired concepts of community (Gemeinschaft) and association (Geselschaft) are useful in the sociological analysis of present - day society because:

1. they show that all human societies, despite their wide variety, can easily be fitted into two major categories

2. although they are too simple to form a complete picture of any actual society, still they are helpful in suggesting a general frame of thought within which we can-direct our research and make our own refinements.

3 in modern urban society the community-like (Gemeinschaft-like) aspects will soon disappear completely - therefore it is good that we try to study them before it will be too late.

4 Toennies' approach can be an answer to the alarming decline of morality that is chara teristic of modern city life.

Column 34: question XII

When we compare sociology with such sciences as physics psychology, geology and mathematics, we conclude that:

1. Sociology is science because it is based on experiments, carried out under highly controlled circumstances, by trained investigators, with the assistance of statisticians and computer specialists .

[p. 25]

2. sociology is a science because it is based on empirical data, meticulously collected by fieldworkers, along the guidelines provided by such theorists as Marx, Durkheim and weber.

3. sociology is not a science because it is not based on experiments but rather on a crude assembly of primary facts, brought together by a dubious and uncontrolled process of selection, and the analysis of which is affected by value judgements and ideologies.

4. sociology is a science because it tries to build up a systematic body of tenable statements, based on the application of explicitly defined concepts to the empirical social reality, and within the general framework of scientific method and logic.

Column 36: Question XIII.

Considering the sociological concepts of family and household we conslude that:

1. .for all practical purposes the two concepts coincide.

2 a family is a household exclusively composed of persons linked by ties of marriage, kinship and adoption.

3 a family tends to be a household, but not all families are necessarily households.

4 alternatives 1,2, and.3 are all wrong.

Column 38 question XIV

Indicate to which one of the'following text examples the concept of "status incogruence is best applicable: -

1 “Belonging to different castes, the two lovers had to face the fact that a marriage would be entirely impossible”

2. "There was an almost unbearable .tension in the room where the final meeting between union leaders and management was held. Although they were clearly aware of their power, yet the members of the strikers committee could not help to feel embarrassed by their poor clothes and their unfamiliarity with certain sophisticated expressions the executive manager used.

3. "Had she been living in her own country, she would have enjoyed all the rights of full citizenship; but now as a foreigner she was just dependent on local regulations and the goodwill of the immigration officers.

[p. 26]

4 "Looking back at the splendid career he had behind him, he felt that inequality in society might indeed have the effect of placing the most talented people in the most important positions, at the same time compensating them for the sacrifices they had to make during their long period of formal training."

Column 38 question XIV.

In the diagram below, the shaded symbols together constitute:

1 a patrilineal kingroup

2 a matrilineal kingroup

3 an extended family

4 a bilateral kinship system.