DOCUMENT RESUME

ED 119 687 IR 003 153

AUTHOR McIntosh, Naomi E.

TITLE Data-Banks in Theory and in Practice; A Case Study in

Higher Education.

INSTITUTION Open Univ., Walton, Bletchley, Bucks (England).

PUB DATE 75
NOTE 22p.

EDRS PRICE MF-\$0.83 HC-\$1.67 Plus Postage

DESCRIPTORS *Academic Records: Adult Education: Computer Oriented

Programs; Correspondence Courses; *Data Bases; External Degree Programs; Higher Education;

Information Needs: Information Storage: *Management

Information Systems; Student Characteristics;

*Student Records

IDENTIFIERS Great Britain; *Open University

ABSTRACT

In a home-based institution like the Open University of Great Britain where enrollment is open and participation patterns are erratic, maintaining up-to-date, accurate records of students is imperative. A systematized data base is required for the administration of admissions, fee collection, materials distribution, monitoring and marking of assignments, grading, and policy planning. Since data collection at the Open University began before the potential uses of the data were established, the system has in some instances been unable to provide needed information, but the system is being updated and modified to insure easy access to more complete information. The possible uses of demographic and academic records are discussed in detail. (EMH)



ED119687

AND IN PRACTICE - A CASE STUDY IN HIGHER EDUCATION

BY NAOM! E. McIntosh
The Open University, London

Data-banks do not grow up on their own. They are constrained by the objectives of the researchers or administrators who build them, and by the institution or environment that they relate to.

Ideally they imply and require systematic planning. In practice this may be difficult to achieve. Unlike normal institutions of higher education, who receive students with few exceptions for a specific limited time period, which has a known end, the British Open University is committed to an open-ended flexible system in which students may study different numbers of courses in different years, take time off and return a year or so later and carry on their studies over an indefinite period.

Different departments inside the University need information from different sources for different purposes. Some of it is course-based and some of it is student-based.

Some of this was envisaged and planned for from the inception of the University. Some was not, and has grown up haphazardly. The University's computer system is predominantly gared to meeting administrative needs and informational needs have until now been met either as a by-product of administrative processes or through survey research. This paper looks from the point of view of the survey researchers at how this data base has been built up, and with what problems, and how it compares with the theoretical outline of feedback requirements prepared in 1970.

24,000 students commenced studies in 1971, and numbers are now up to 45,000. The basic student record file carries 300-400 items of information for each student, ranging from occupational and educational background, address changes, access to study facilities and progress and attainment during their studies.

N.E. McIntosh

Survey researchers at the Open University were confronted in 1969 with a request from computer consultants, retained to assess and plan for the likely future computer needs of the Open University, to outline their envisaged needs for data processing over the next few years. At that time one could only answer in general terms.

Twenty-thousand odd students were to start studying in the first year, to be followed by similar numbers each year, ...orking in an open-ended, flexible system which permitted different numbers of courses to be taken, in different years, with time off in between courses if desired. All this could only mean one thing: a data-bank, with facilities for constant up-dating, banking and Colleagues in survey research understood the problem, but indicated that facilities at that time, at least in Great Britain, were not available. This was, in effect, a software rather than a hardware problem. Computer experts had had no need to become familiar with the problems posed by "messy survey data", and did not appear to understand that there might be such The one page questionnaire to be completed at that stage enabled answers to be given only in terms of the mechanical requirements of a data-bank:

- ability to add, bank, up-date at frequent or irregular intervals
- variable length records
- all information able to be inter-related
- large core storage
- much survey type data and therefore survey type analysis

But this was only one view of the University's requirements, and was greatly overshadowed at that stage, by the urgent necessity to get administrative systems, essential for the efficient working of the University, into operation. This absolute necessity was not aided by the limited capacity of the data processing system permitted within the funding allocated to the University, or by the limited time-scales within which the stem had to be set up.

Of course, in some ways the problems are analogous to a hospital with medical records for patients stretching back over decades and liable to be revived at any time without warning, or



to a social service department dealing with families with multiple problems over time. But the complexity of the system, combined with the numbers, the scatter and the time-scale made the problem formidable.

The numbers likely to apply were unknown, and even the application form had to be designed to:

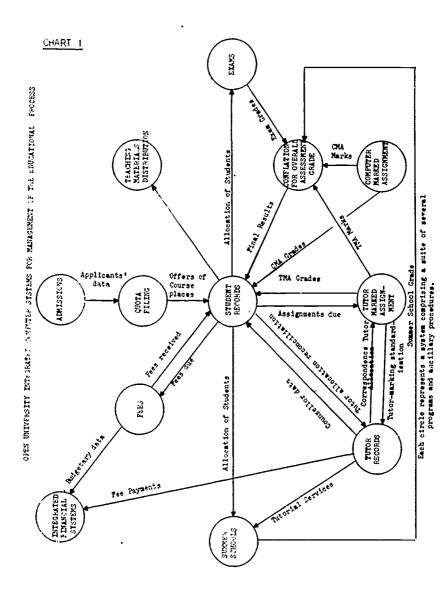
- tell the University something about the students, and enable selection if this proved necessary or desirable
- 2) be computerised to assist the admissions process
- 3) become the first section of the student record file

The admissions process formed one critical part of the early work of data processing. Other parts, of less immediate importance to the researchers, but more critical for the University, were the fee payment system and the system for the distribution A third critical area was monitoring of teaching materials. and marking of assignments. This includes information both about the return of assignments and about marking. Some are marked by computer, and some by tutors, and records of tutors also have to be kept. The range of computer processes which together add up to the sum of the University's operation are shown in chart 1. Basic decisions about this structure were reached early on in 1970, and the detail of individual systems was then worked out subsequently. Of course, some of the systems have little relevance to the researchers, and this paper will concentrate on those that are relevant.

On the face of it, this is an enviable system. In reality the systems do not allow interface between the different student files without purpose-written programmes; and access, except for agreed regular administrative routines, has been extremely difficult.

It is possible to identify three main groups in such a situation which are likely to affect, or at least have an interest in, the base of data to be gathered and accumulated. Note that the decision to gather data does not assume or imply the decision to accumulate.





ERIC

These three groups are:-

- the administrators (or officials)
- the data processing staff
- the policy makers (in this case the academics)

The interests and objectives of these groups may not be the same, and may well be in conflict.

For example:-

- a) the administrators may wish to impose a system which appears tidy, but is less efficient in data processing terms.
- b) the data processing staff may not be aware of the implications of the political or social objectives of the institution.
- c) the policy-makers may not even know yet what it is they need to know!

At minimum the objectives of the first two groups are likely to be short-term while that of the third group is more likely to be long-term. Constraints on resources may be felt to be more real by one group rather than another.

One example suffices to illustrate the complexity of this In principle, applicants for places at the University were to be categorised according to occupation. Housewives, in this case, create difficulties. To categorise them according to their past job, the original suggestion, might be unfair. Since the University provides a second chance, their intended job might be a better guide: if one was to look at "national need" an oft-quoted criterion in the Open University's early days, married women returning to "new" occupations might well prove important. In the event, academics and researchers with these long term aims in mind included, on the first application form, both past and intended occupation. To "data processing" these appeared unnecessary refinements, particularly since in the first year they were not used in the admissions process. So they recommended forcibly that they should be dropped: and The research connotations of the long term objectives took second place to the short term pressures for efficiency and space on the computer file. Two years later they have been



N.E. McIntosh

re-instated, but valuable trend information has been lost.

This paper, written as it must be from the stand-point of the researcher, cannot do justice to the problems faced by other parts of the organisation. It can try to recognise that they exist 'and at best may occasionally have contributed to their solution. At worst the researcher's demands could be seen as the straw that broke the camel's back. It is not reasonable to discuss here the whole early design of the data-processing system except in as much as it affects the researcher.

The first decisions about data capture had to be made before any decisions about retrieval and storage were known. application form provided the first main input to the student record file via the admissions file. It recorded some demographic data - sex, date of birth, marital status and terminal age of education, occupation, TV and radio ownership and highest level of educational qualification. But, of course, it was designed primarily to facilitate the admission process and not to be the beginnings of a data-bank. The admissions file recorded all applications in the date order of the application and manipulating of the selection process was done by varying the order of the file. Weightings could be made only by jumping people forwards or backwards in the queue for a limited number of variables, e.g. course choice, region, occupation and age. The significance of this was that the ultimate selection was effectively first-come, first-served in the queue. (The admissions process is discussed in greater detail in the "Report on the 1971 Admissions cycle".) 1.

This paragraph illustrates the multi-purpose nature of all records, and shows how impossible it is to pretend that a base of data is sytematically built up in an integrated manner. Having said this, it is possible to disentangle some threads of order.

- Administrative processes had to be carried out
- Policy decisions had to be implemented
- Long term evaluation had to be planned for

Information all stems from students. It may stem from information they give, either compulsorily or voluntarily. Alternatively, it may stem from observation and recording of their actions. Of course information about academic staff,



7

full, and part-time is important, but it is in a very real sense-secondary, as without students there would be no staff! An' important early decision related to the amount of information one could compulsorily require all students to give as a necessary part of the system. Effectively, information believed necessary was required compulsorily. Information believed desirable has to be obtained in other ways.

And it is one thing to ask for the information once - it is another matter to keep it up-dated. A data-bank, as with longitudinal studies, not only has to record entry behaviour, but changes over time. An obvious administrative example, a necessity in the Open University situation, is a classic unobtrusive research measure - change of address. There is nothing to lose and everything to gain for a student in providing this information. It is an administrative necessity. It can also provide, for the researcher, an indication of pressure.

The same carrot be said for occupation. A student may change his occupation several times. He may consider this of no concern to the University. But "occupation" already has multiple uses.

- initial occupation is used in the selection process
- initial occupation is used as one of the indices of student progress
- changes in occupation may measure the achievement of individual's academic or personal objectives
- aggregate changes in occupation may measure the achievement of social objectives

Occupational data are complex to collect, categorise and up-date. And adult students may well feel that this verges onto the area where data-banks and privacy become an issue.

The University started its life at a time when data-banks and privacy were the subject of much public discussion. The question of what was or was not reasonable to ask and record of thousands of independent working adults was therefore important not just for administrative reasons, but also because of possible political repercussions.

The first sets of decisions related therefore to the information that should be stored for each student. An early listing prepared by the computer consultants showed the following



N.E. McIntosh

tentative listing, and the likely source of the information. The timing of this listing (18th July 1969) coincided with the drafting of the first application form, the first main source of information.

Information to be stored - first outline

Personal information	Input	
Student number Student name Title Address Telephone number Sex	Admissions	H H H
Birth date Marital Status Age and number of children Number of other dependents Occupation Physical disabilities Country of Origin Nationality	Application " " " Admissions Application	n n n n file
Education Previous Education (Non-Open University) Previous Courses & Grades (Open University)	Application To be up-dat	

Open University Allocation

Region	Admissions file
Viewing Centre Counsellor	By region By region
BBC Reception	Application form

<u>Grades</u>

Assignment (Examination	_	Tutor-record Exam.student		
			9	

Fees

Summer School Information	S.S. student grade form
Financial Information	Financial office
	ringucial office

Source: Urwick Diebold - July 1969



It would be interesting to insert, for comparison, the listing of the main student record file in its current form. Suffice it to say that it is now over ten pages long!

In the event, it was decided that some of the information listed should be gathered later or in different ways. application form could not physically hold all the questions, and even if it could, some of them were felt to be inappropriate as "application stage" questions for what was, after all, an open university. Nationality and mother tongue were examples Another contentious area was the amount of information that should be asked about education. Since the application form is the main public document seen by all applicants, to ask too much detail about previous education might well give the appearance to the applicant that the university is not in fact open. On the other hand, to ask too little may lead to the charge of irresponsibility since the university would not be able adequately to advise on the applicant's preparedness for study.

Not all the information gathered on the application form was input on to the admissions file, and - a second filtering - not all the data from the admissions file was transferred on to the main student record. One particular example - the algorithmic grading to assess educational preparedness used in the admissions process in the first two years - was not transferred to the student record file, and has been lost. (Note that this grading was not used to select for entry, but was used to determine whether or not an applicant should register for one rather than two courses in one year). Such a record would have provided an important alternative indicator of likely success or failure to set against the more conventional recording of educational qualifications and background. This data loss illustrates again the conflict in priorities between data processing and research, when the long term evaluation needs are not immediately apparent to those concerned with day-to-day operations.

Both the application form and the provisional registration form were seen as primarily administrative documents, and it is not surprising therefore that the main responsibility for their design rested originally with the administration. The basic informational implications of these forms were not realised at that stage let alone the implications of these decisions for subsequent research and evaluation. An early Senate meeting (August 1969) faced with the draft application form, had to choose between about ten occupational groups, chosen "pragmatically" by



the Vice-Chancellor and 211 occupational groups - The Registrar General's groups - chosen technically correctly, but quite unrealistically by the administration. It was at this stage that the researchers were first invited to assist in achieving a compromise, and were involved in decisions about how to ask for information and what should be recorded. May 1970 saw the drafting of the next key document - the provisional registration form. This was designed to up-date the information transferred from the application form and also to ask additional questions which appeared crucial, but had been inappropriate or difficult to ask earlier. It aimed to cover those indicators that were generally agreed to be likely to prove critical for student survival.

I say "generally agreed" since the responsibility for the design of the form again laid with the administration and data processing functions, which had the main say in choosing the questions and drafted the coding frames. They accepted three "research" questions as additional candidates for inclusion at this stage - mother tongue, type of last school attended, and previous part-time education. The majority of the information on the provisional registration form did, in fact, have research as well as administrative purposes, although we have not had the resources to analyse some of it to this day. We have successfully used it to provide the main indicators against which we have chosen to analyse our regular student progress statistics. These indicators currently are -

- date of birth
- sex and occupation
- region of study centre
- highest educational qualification
- previous part-time education
- terminal age of education
- special indicators of difficulty mother tongue access to BBC special disabilicies etc.

We also have recorded, but have not used

- marital status
- type of last school attended



- phone ownership
- professional qualifications

Considering that this range of information is available for all students automatically, it could be considered that we were already better off than normal educational institutions, and no further information was necessary. However, the institution is not normal, and the researchers felt strongly that adult students working part-time studying through a multi-media system might be so constrained by their patterns of work and study circumstances as to affect their capacity to study. An area of information concerned with likely study patterns was thought to be critical. In principle, the administration accepted the desirability of this. However, the original idea of printing the additional research questions on the back of the provisional registration form proved not to be possible, and some alternative way of seeking this information had to be found. The types of information suggested at this stage (1st May, 1970) included -

- ease of access to libraries
- time leaving for work and returning home
- place in which to study
- access to tape-recorder, record player etc.
- anticipated student problems

Also under discussion was the whole area of cumulative indicators of student progress derived from student's study patterns. Outlines of ideas about feedback were developing in parallel with this, but at this stage feedback was seen as primarily related to student assessment and not as an important function in its own right. It was intimately related to the work of the Examinations and Assessment Committee, a completely different group of people from those involved with student registration. Feedback was seen primarily as course-related, and not as related to the functioning of the whole University. Indeed it is still seen in this narrow sense by some parts of the University now.

A paper, prepared in May 1970, attempted to take an overall look at feedback requirements, and this summary did something to bring together the disparate groups involved in developing different parts of the system. It is salutary to look at this outline of feedback requirements to see, three years later, how many of them have been met, how many have been proved to be unnecessary, or thought now to be undesirable, and how many have fallen by the way-side.



The Deputy Secretary and Registrar commented some months later, in the light of the May paper. 3

"The feedback and communication system in the Open University situation is complex. It is complex because of the remoteness of the students and because of the fragmentation of the teaching services. At least six - University standing committees and Project Working Groups have been involved in discussion of parts of the system excluding the four Foundation Course Teams. This division of labour has been necessary in order that limited parts of the system could be developed within the time available, but it has also had the disadvantage that overall co-ordination has not always been achieved.

In the last resort much of our teaching system will be dependent upon personal and informal relationships between the student and his tutors. All the University can hope to do is to provide a system which allows information to flow quickly and effectively; the information which is recorded will only be a small sample of the whole. A paper prepared for the Examinations and Assessment Project Working Group comes nearest to asking all the questions..."

The Deputy Secretary, took at that stage, the objectives of the feedback system to be:-

- to signal individual student problems and to allow remedial action to be taken
- to signal individual tutor problems and allow remedial action to be taken
- to signal course units' problems and allow remedial action to be taken
- to allow the inter-relationship of student, tutorial and course unit problems to be seen as a whole.

An extract from that paper is shown on the 27,28 and 29. The right hand column gives an indication of what information has been collected and what has not, and whether it has been done so as part of the official information system, or additionally through research procedures.



Extract of paper prepared for Examinations and Assessment P.W.G.

FEEDBACK FOR ASSESSMENT PURPOSES

FEEDBACK will need to cover three informational areas:

- 1. The unit and the course.
- The tutors.
- 3. Individual students and their performance.

These will be needed to meet the following objectives:

- (a) (i) Monitoring the course performance.
 - (ii) Monitoring the unit's performance.
 - (iii) Monitoring the test item's performance - subjective.
 - objective.
- (b) Monitoring the tutor's performance
- (c) Monitoring the student's performance on each unit cumulatively

FEEDBACK may also be categorised according to the users of the feedback.

- 1. The individual student.
- The tutor and the tutorial staff.
- The course team.
- 4. The regions (including counselling service)
- The administration.

Decisions on the amount and frequency of feedback to each user will need to be taken and <u>possibilities</u> have been indicated in the right hand columns according to the above categories.

Information required may be subdivided into three categories:

- A. Information to be banked initially.
- B. Information to be added periodically, to the record.
- C. Information to be recorded assignment by assignment.



1. Feedback to student might include:-

Individual mark
Average mark
Average length of time on unit
Computer commentaries on errors

2. Feedback to tutor might include:-

Overall distribution of scores per assignment, plus interpretations Student's subjective reactions Tutor's subjective reactions

3. Feedback for course unit analysis might include:-

Distribution of total scores
Distribution of scores, item by item
Length of time spent on unit
Length of time spent on course cumulatively
to date

Distribution of students subjective indicators - difficulty, motivation etc.

Distribution of return dates
Tutors subjective views of unit
Broadcast and study centre access
Work patterns
Etc. etc.



Categories of information					_			
	Users of feedback					Sources of feedback		
	S	T	C	R	A	OR	RD	NO
A. Information to be banked initially (from central records)								
 Identification number Course(s) subject and number being studied Region Study Centre Counsellor Tutor (unless Science) 						X X X X		
2. Information on study conditions - time back from work - access to library - access to BBC 2 TV - access to VHF radio - space to study at home - ability to attend study centre		X X X X X		X X X X	X X X	X X	x x x	
3. Demographic data from background questionnaire and recordsB. Information to be added		x	Х	х	х	х	Х	
periodically 1. Non-assessment assignments								
- Summer schools assessment - Examinations - by item - overall grade	x x	x x	x x	x x		x x		
 Information from special ad hoc investigations e.g diagnostic tests personality tests etc. 			х		х			x

S - Student T - Tutor C - Course R - Region A - Admin. OR - Official records RD - Research Data NO - Not obtained



Categories of information (continued I)		Heave of facilities					Sources of		
	S S	Users of feedback					feedback OR RD NO		
C. <u>Information to be</u> recorded assignment by assignment		_		,					
1. From the individual STUDENT									
1. IdentificationStudent numberUnit numberCorrespondence tutor						X X X			
2. Broadcasts: TV programme seen - midweek - weekend - neither			X X X		x x x		X X X		
- Radio programme heard - mid-week - weekend - neither			X X X		X X X		X X X		
3. Study centre visited:		Х		х			х		
 4. Work pattern: Time spent on this unit (irrespective of when?) Time spent studying this 			x				х		
week - Average time per unit - Cumulative time on course - Assignment received on			X X X				х	x x	
time or not			x			х			
5. Student attitudes:Difficulty of unitAmount of material in			x	х			x		
unit - Getting behind or not - Motivation to continue			X X X	X X X			X X	x	

S - Student T - Tutor C - Course R - Region A - Admin OR - Official records RD - Research Data NO - Not obtained



Categories of information (continued 2)	Users of feedback					Sources of feedback		
	S	Т	С	R	Α	0R	RD	NO
 6. Continuous assessment: Self-assessment tests (are these to be collected?) Objective tests - overall score, marks per item Tutor-marked assignments - distribution of marks: overall - by tutor - for each student 	x x x	x x	x x	X X		x x x x		х
II FROM THE TUTOR 1. Tutor's range of marks per assignment 2. Tutor's number of students and drop-out rate (except science) 3. Time taken to return to work 4. Subjective view of assignment: * - level of difficulty * - amount of material			x x x	x x x	x x x	x	X X	x x

^{*} experimental work is starting on this.

S - Student T - Tutor C - Course R - Region A - Admin. OR - Official records RD - Research Data NO - Not obtained



Of course, that listing does not contain all the information that needed to be collected, or indeed all that was collected. Since it looked at feedback within the overall framework of the assessment system, it concentrated on the "course unit" - the amount of work to be covered in one week on any one course. (Foundation (first) level courses were all originally planned to run for 36 weeks, with 36 course units taking on average ten hours a week to study). It was possible in those early days to link feedback with assessment since the assessment system had not been worked out in detail, and all the plans were based on the assumption that for each week, i.e. each unit's work there would be a written course unit, TV and radio programmes and some form of assessment. The original idea, then was to obtain the feedback for most of the regular study information from the assessment form, and require all students to complete It soon became clear that not only it as a matter of course. were there going to be variations between course teams in their use of computer and tutor-marked assessment procedures, but also that there were certainly not going to be assignments related to every unit. Assignment based feedback could not therefore provide all the information listed in Section C.1. 2-5. Clearly the information had to be unit-based.

Systems analysts and researchers then combined to work out a simple document-read system based on course units which would automatically provide from all students the possibility of both individualised and aggregated data for the different groups who needed it. However, it proved impossible to implement these plans for integrated automatic feedback, and so although much of the same data have been obtained, they have had to be obtained outside the system by means of additional research projects. In particular although still document-read, they had to be devised, read and analysed by systems external to the University. Thus merging of data, particularly demographic data, although possible becomes more difficult.

This merging problem also applies to all the data from the longitudinal studies of students' educational, occupational and social background.

These studies were funded initially by the Social Science Research Council, not by the University and the University did not have the data processing capacity to analyse them internally. The computer records give however a unique bonus to the researchers in providing not only a virtually perfect sampling frame, but also a wide range of demographic and other variables



against which to check the characteristics of non-respondents.

Ways in which information from this data-base have been already used within the Open University are described in a companion paper to this one.⁴. Examples of forms and questionnaires from which the data are derived can be found elsewhere.⁵. This paper has concentrated on the problems involved in setting up a data-base. The presentation will supplement this discussion with findings drawn from that data-base.

There remains one question - when does a data-base become a data-bank?

The Data Processing Officer writing in September, 1970 identified the characteristics of the data-bank from the point of view of research and evaluation, as follows:-

"1. Growth

Apart from the build up of the data for the initial group of students over 5 years, data will be added from numerous surveys conducted with other student groups and even with control groups from the general population.

2. Need for flexibility

It is impossible to define at this time what surveys and analysis will be wanted as the research project develops. Current design and operations of the project must not be allowed to prejudice the future.

High volume

5

No substantive figures can be given yet, but it is known that there are 25,000 students in year 1 and the data recorded for each will run to thousands of characters.

4. Capability of direct access

Access to the data-bank cannot be restricted to serial mode (i.e. magnetic tape) because of the need to process the data by characteristics rather than by the individual student.



5. <u>Multiple sources</u>

The data-bank will comprise selected data from the "administrative" files augmented by surveys, interviews etc. The "administrative data" will in turn arise from a variety of files set up by D.P. Division.

6. <u>Security</u>

We shall be bombarding our students throughout their courses to supply us with data about themselves, some of it of a highly personal nature. This will be done at a time when public awareness of the potentialities of computer data-banks will grow. The data-bank will be operated within the arrangements approved by Council for computer personal data.

7. <u>Hardware requirements</u>

Basically this is access to a machine with large core store, moderate speed processor and substantial random access backing store, preferably on exchangeable units. This machine must have the capability to exchange magnetic tapes with the Open University ICL 1902A."

None of these characteristics have changed, and although, the University's computer installation has been enhanced somewhat, it still cannot meet the researcher's needs. The University's own needs have grown so rapidly that it is currently requesting funds for a major revision of the system in the next triennium. Current plans for an integrated data-base will incorporate, it is hoped, all the researcher's needs - but not for several years.

In the meantime, one might re-phrase the question. When is a data-bank, not a data-bank? When you cannot access it.



RE FERENCES

- 1. The Open University. "Report on the 1971 Admissions Cycle". April, 1972.
- 2. The Open University Committee paper. Clinch J. "Feedback and communications with reference to the teaching processes" 14.9.70.
- The Open University Committee paper. Lewis, B. N. & McIntosh, N.E. "Computerised assessment procedures and feedback". May, 1970.
- 4. McIntosh, N.E. "The use of survey research in the planning of higher education". ESOMAR 1973.
- 5. McIntosh, N.E. "Research for a new institution the Open University" published in "Innovation in Higher Education". SRHE 1972.

