THE DEVELOPMENT OF A CRITERION-REFERENCED TEST OF OCCUPATIONAL FUNCTIONAL READING ABILITY

OWEN DOUGLAS PARRY

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Sheffield City Polytechnic

September, 1980

CONTENTS

Appendix I	:	Tables of Data
II	:	British Standard Industrial Classification
III	:	Letters, Forms and Schedules
IV	:	Manual of the Item Banking System
V	:	Manual of the Vocabulary Banking System
VI	:	Provisional Manual of the Functional Reading Test, Form A
VII	:	Functional Reading Test, Form A
VIII	:	Extrapolated Quotients of the Edinburgh Reading Test, Form 4
IX	:	Flowcharts of Computer Systems
X	:	Published Papers
XI	:	Courses of Study Undertaken by the Candidate

ABSTRACT

The Development of a Criterion-referenced Test of Occupational Functional Reading Ability

Owen Douglas Parry

The Sheffield Occupational Functional Reading Project is discussed. The Project is placed in the context of rising concern about the qualities of school-leavers. The development of a test based on the reading requirements of employment is described. Functional reading tasks were identified for a wide range of jobs. Reading materials were collected during interviews with employers, officers of training organisations and staff of colleges of further education. Interviews were conducted with recent school-leavers.

Items were constructed using standard and novel item-types. All items were scrutinised by panels of employers and linguists and were piloted in schools. Items were amended or rejected as necessary. Two forms of a functional reading test, Forms A and B, were constructed. Test-retest procedures were used to assess the reliability of each form, using both product-moment correlations and agreement coefficients. Form B was not acceptably reliable. The effect of the novel item-types on the estimation of reliability was investigated and no significant effect was found.

Form A and the Edinburgh Reading Test, Form 4, were administered to 470 pupils. Form A had concurrent validity and internal consistency. Testees were followed up after leaving school. The employers of those obtaining jobs were asked to complete rating forms on their job performance. Due to local labour market conditions and a low return of these forms, the predictive validity of the test could not be established. Valid applications of the test are discussed.

Computerised question and vocabulary banking systems were developed to support test construction.

An analysis of errors made on Form A was undertaken. Implications for specific diagnosis are discussed.

The nature of further research in the area is considered and recommendations made.

ABBREVIATIONS

The following abbreviations are used in the text and appendices:

APU	Assessment of Performance Unit
B.S.I.C.	British Standard Industrial Classification
CONT-2	Content category 2
CONT-3	" " 3
CONT-4	" " 4
CONT-5	" " 5
CONT-6	" " 6
CRM (or CRT)	Criterion-referenced measurement (or test)
D.E.S.	Department of Education and Science
G.T.A.	Group Training Association
H. of C.	House of Commons Select Committee on Expenditure
I.B.S.	Item Banking System
ITB	Industrial Training Board
JOB-1	Job type 1
JOB-3	" " 3
JOB-4	" " 4
JOB-5	" " 5
JOB-6	" " 6
JOB-7	" " 7
LING-1	Linguistic task 1
LING-3	" " 3
LING-4	" " 4
MUSIC	McGill University System for Interactive Computing
NAEP	National Assessment of Educational Progress
SCP	Sheffield City Polytechnic
SOFRP	Sheffield Occupational Functional Reading Project
TES	Times Educational Supplement
VBS	Vocabulary Banking System

ACKNOWLEDGEMENTS

The excerpt from the British Standard Industrial Classification in Appendix II is reproduced by kind permission of H.M. Controller of Stationery.

Figure 7.1 is reproduced by kind permission of the Human Resources Research Organisation, Alexandria, Virginia, U.S.A.

Reading passages in the Functional Reading Test, Form A are reproduced by kind permission of the various copyright holders.

Personal thanks are due to the following:

William Latham, Co-ordinator, Language Development Centre, SCP, Supervisor;
Gordon Craig, Dean of Faculty of Education, SCP, Second Supervisor;
Brian Wilcox, Senior Advisor (Research & Evaluation), Sheffield Education Department, Advisor;
Peter Dutton, Liaison Officer, Sheffield Regional Centre for Science & Technology, Advisor;
Staff of the Language Development Centre, SCP;
Members of the Department of Professional Studies, SCP;
Members of the Department of Computer Services, SCP, especially Peter Messer;
Members of Sheffield Careers Service, especially Michael Threlfall;
Officers of Industry Training Boards and Group Training Associations;
Personnel and Training Officers, Employers and Employees from firms assisting in the SOFRP;
Headteachers, Staff and pupils of schools assisting in the SOFRP;
Members of the Employment Services Agency;
Janet Smith, Schools Council;
Christiane Forge for translations;
Morag Hunter, University of Leicester School of Education for comments on Chapter 19;

Sean Close, St. Patrick's College, Drumcondra, Dublin;
Keith Jones, Department of Mathematics and Statistics, SCP;
Margaret Fogden, Muriel Smith and others for typing;
Francesca Woods and Joy Tinsdeall for technical assistance; and
Angela Rees, for everything.

CHAPTER 1

THE SHEFFIELD OCCUPATIONAL FUNCTIONAL READING PROJECT

The Nature of the Project

The Sheffield Occupational Functional Reading Project (SOFRP) was established at Sheffield City Polytechnic in September, 1977, in association with the City of Sheffield Metropolitan District Council. SOFRP has four major objectives:-

(i) the identification and classification of reading tasks associated with job-seeking and employment in Sheffield, which a 16+ year old school leaver faces upon leaving school;

(ii) the construction and validation of a criterion-referenced test based on the reading tasks identified in (i) above;

(iii) the construction of associated diagnostic tests to provide detailed information on individuals whose scores are below some decision-related score or range of scores on the test developed in (ii) above; and

(iv) in consultation with local teachers, the development of appropriate remedial materials.

The research reported in this work relates in the main to the first two objectives listed above, whilst some work on the third is also reported.

Before describing the Project, it is first necessary to look at certain antecedent conditions and climates of opinions which have had an effect upon and are of direct relevance to the design and conduct of the research. Four such areas may be delineated: moves of a largely political nature in the sphere of education, problems of recruitment to industry and commerce, an increasing rate of technological change and a number of local conditions, all of which combined to make an appropriate setting for the study of the reading abilities required by school-leavers on starting work or looking for a job.

Employers and School Leavers

Over the last ten years, a public debate has been carried on concerning the standard of the British education system. At one extreme, there have been continual, scathing polemical attacks on a range of 'progressive' ideas and practices, including the raising of the school leaving age, comprehensive reorganisation, mixed-ability teaching and the Certificate of Secondary Education. The advocates of this extreme suggest that there has been a marked decline in the 'products' of the education system as measured by examination results, the 'quality' of school-leaver job applicants and the scores on various tests, and attribute this decline largely to the practices listed above. The other extreme condemns the former as reactionary and inappropriate to modern society. It is claimed that no real decline in standards has taken place, but that the instruments used for comparison over time are inappropriate or invalid. Proponents point to the wider range of activities available to the pupil to promote a better educated individual, and the increased access to higher and further education leading to the creaming off of the better pupils who were the job applicants of the past.

Such wide-ranging controversy needs careful consideration to sift the rhetoric from the genuine concern, the uninformed opinion from the detailed survey, and to discover the fallacy versus the logical argument. To do this, one must look at the claims and counter-claims and attempt to balance the argument in order to achieve an understanding of the specific problems. Fot it cannot be denied that the level of vehemence used in the public debate over education has been such that it could not be totally lacking in foundation, on either side. What those foundations are, one hopes to elucidate.

If one starts with the economic or political spheres, it is difficult to avoid the imputation that employers and politicians have no faith whatsoever in the British education system. For such is the newsworthiness of their complaints that praise is rarely printed. It would do them grave injustice to suggest that they have no praise, but their complaints have also been weighty.

Some selections from the educational Press will serve to indicate the nature of the concern of some employers and politicians over the standards of school-leavers in relation to employment. The personnel officer of a major bank asked recently: "Is it too much to expect that those leaving school at 16 should have the ability to communicate in their natural language in the spoken and written word and to have some understanding of what numeracy is about?" (Times Educational Supplement, 17th September, 1976). Dr. Keith Hampson, M.P. called for a crash programme to help 100,000 youngsters, who were so bad at writing and handling figures they were nearly unemployable (T.E.S. 26th June 1976). An ILEA inspector reported that many pupils leave at 16+ "lacking in the barest survival skills in terms of employability" (T.E.S., 17th September, 1976). The chairman of Wedgwood Limited reported this in an address, stating that young people did not have the basic tools of literacy and numeracy and too many were virtually unemployable (T.E.S., 14th January, 1977). Ten years earlier, a writer suggested that 80,000 school-leavers were unable to read effectively (Smart, 1968).

The problem is by no means new, however. The Norwood Report of 1943, for instance, says: "From all quarters, universities, professional bodies, firms and business houses, training colleges and many other interests and many individuals we have received strong evidence of the poor quality of the English of Secondary School pupils ... ". Dixon (1979) calls all the previously cited examples 'impressionistic' evidence, as 'comments from industry, commerce and public service are not based on nationally representative samples" (p. 4), but in contrast, the CBI disagrees. Presuming that their enquiries amongst their own members are fairly representative, they have stated that "Employers are becoming increasingly concerned that many school leavers, particularly those leaving at the statutory age, have not acquired a minimum acceptable standard in the fundamental skills involved in reading, writing, arithmetic and communication" (House of Commons, 1976, p. 139).

The overall concerns can perhaps be expressed in the following fashion. In recent years, employers of recent school-leavers have found that both the applicants for posts and those they actually employ, have levels of ability and/or performance in certain basic areas, such as reading, writing and computation, which fall below the levels the employers consider necessary for adequate job performance and/or are less than levels held by similar job incumbents in earlier years. On perceiving this mismatch, one or both of two causes are usually put forward. The first is that schools are failing to produce school leavers of the same 'quality' as in previous years. Evidence for this is often advanced in terms of declining scores on selection tests used by a very large number of employers, or in terms of less measurable variables such as ability to complete an application form adequately. The second cause is usually that, whatever the overall standard of ability of school-leavers, insufficient time is given over to the study of basic skill areas.

A number of refutations of detail are possible at this point. One reason for the perception of decline in standards is that the employers are not comparing the skills of similar school-leavers. Increased access to further and higher education has meant that a large section of the above average ability range now remains in the education system for further study. The labour market of school-leavers, at 16+, therefore comprises a group with a depressed average ability. As employers have more less able young people from whom to select, it is obvious that the perception of decline should occur, as they are expecting the less able to be of as good 'quality' as the more able were in the past.

In dealing with 'harder' evidence for decline, such as the selection test scores, it may well be the case, in language at least, that the tests are invalid or inappropriate. Dixon (1979) reported that 'so far as actual use of language on the job is concerned, these tests may have serious defects. In fact, we have yet to meet one constructed specifically for that purpose" (p.6). Even so, one would expect pupils of similar ability to perform similarly on such tests unless particularly obselete and anachronistic items were included. One may therefore conclude that either the previous explanation of changing groups is true, or that there has been an actual decline in ability in undertaking these tests. A similar conclusion may be reached for such assessments as form filling, etc.

So it seems to be the case that there are qualitative rather than quantitative differences between past and present school-leavers. The cause for concern remains, however, and Dixon again puts it well: "we have met well-based and constructive criticism of the difficulties that young workers in a given firm, local industry, or organisation have faced in coping with the language demands of the job" (1979, p. 5). Perhaps the final cap can be placed on the 'standards' complaint by citing the Department of Education and Science who maintain that "it is simply untrue that there has been a general decline in educational standards ... Recent studies have shown clearly that today's school-children read better than those of thirty years ago" (1977).

Given that many complaints are likely to have been founded in the qualitative differences in the groups available for employment, another explanation is also possible. This is that the requirements of jobs have changed over time. This is a theme to which we shall return.

The Educational Climate

A number of employers have attempted some remediation amongst their recruits (H of C, 1977, p. 140) but, in general, they consider that it is the function of schools to bring pupils up to the required level and to undertake in schools any necessary remediation. That is, the schools should provide a suitable finished product. It will be useful to consider, therefore, whether some or all of the problem of mismatching school-leaver ability and job requirements is rooted in the practices of schools. One area of school life around which all else seems to revolve at 16+ level is that of public examinations. The Lockwood Committee (1964) expressed serious doubts about the educational efficacy of the English language O-level, claiming its narrow concentration on particular forms of language had its effect on the future reading and writing practices of pupils. A more recent survey by Her Majesty's Inspectorate (DES, 1979) into aspects of secondary education discovered a number of disturbing factors. "Exclusive concentration on the requirements of public examinations appeared to be an important factor in the impoverishment of reading in at least one-fifth of the schools (surveyed)". They reported a 'narrowing in the scope and quality of reading by most fourth and fifth form pupils' and that 'it was frequently taken for granted that pupils develop without help their skills of reading for different purposes' (p. 73). So it would appear that in a large number of cases, schools are paying more attention to examination syllabuses than to the educational needs of their pupils. Of course, this is not something to which schools have aspired nor come to of their own volition. The pressure by both parents and employers to render a measurable account of their practices, have forced schools to hold up the number (rather than the quality or appropriateness) of examination passes as their own certificate of credibility. It is unsurprising, therefore, that the secondary survey also found that "in at least a quarter of the schools it was the least able who suffered particularly from a failure to extend their reading or to provide an appropriate range of material" and that "although remedial help was often programmed in the lower forms, provision for poor readers in years four and five was much rarer" (p.74).

In a way, teachers have been placed in a cleft stick, forced on the one hand to concentrate on public examinations and have then become the subject of attack for the results of that concentration. The changing nature of society, particularly economically and politically, has meant that, over the last few years, schools have been increasingly required to do both: obtain high rates of examination passes and produce pupils with high levels of competency in basic skill areas. This movement may be seen as a part of, or a direct result of, the call for accountability in education. Accountability may be taken literally, to mean being called to account for money spent, or inferentially, to mean being called upon to demonstrate the adequacy of the education that pupils receive. In the United States, the movement has been established over the last decade with both meanings being used. Typically, this has involved teachers specifying in some detail the objectives of given courses or curricula, teaching the subject matter to their pupils and assessing the outcomes i.e. a form of input-output model (see e.g. Sockett (1980); Burstall (1978)).

The introduction of a similar system might have obvious attractions to employers as it uses an 'economic' model of education: input-output. Similarly, in a time of public expenditure cuts, the more measurable the end-product of a process, the more able one is to assess its worth in terms of the Budget. In this country, however, there has remained an acknowledgement, so far, that education is rather more than using resources to obtain a certain product. What has occurred, however, are moves towards a centrally controlled core curriculum, which can be directly related to the public concern of employers (and also a desire by the Department of Education and Science to reassert its involvement in the content of education (MacClure, 1979)). There has been the introduction of a national monitoring system, specifically recommended by the Bullock Committee in 1975. Further, there has been a definite policy change towards an increased emphasis on the role of schools in vocational preparation. This was particularly heralded by James Callaghan's speech at Ruskin College in 1976, when he put forward the view that "there is no virtue in producing socially well adjusted members of society who are unemployed because they do not have the skills. Nor at the other extreme must they be technically efficient robots" (T.E.S. 22nd October, 1976). Thus, there have been growing expectations that changes will take place and that emphasis be given to basic skills, as required by industry and commerce.

It is to be noted that the emphasis has now been officially placed on learning for work (despite record unemployment amongst school-leavers) and less on the personal development of the individual. Indeed, Sockett (1980) went as far as to suggest Callaghan's speech was an attempt to lay the blame for the economic ills of the country at the schools' doors. Nor has the pressure from such authors as those of Black Paper fame been lifted, in calling for various returns to previous styles of education. Their main pressure, here, of course, has been to make schools keep examination successes as high as possible.

The view that schools should increase their efforts in the area of basic skills work has more or less been accepted by teachers. For, as Sockett points out, "direct vocational preparation, whatever the rhetoric, has always been a major feature of the (education) system". The Manpower Services Commission was happy to report that "there is wide acceptance throughout the education field of the opinion that education should have some relevance to a pupil's subsequent working life" (MSC, 1979). A large number of schemes of work-experience for pupils and employer-education liaison links have grown up quite recently.

The Effect of Technological Change

The discussion so far has centred upon the dissatisfaction of employers with the level of ability in basic skill areas of school-leavers, and the demonstration by the survey of the Inspectorate that things are not as they should be in many schools as far as reading is concerned. Earlier, a further explanation of the mismatch perceived by employers was advanced: a change in the requirements of jobs.

Any changes must have been quite small to date or they would have received greater emphasis in the complaints. It is true to say, however, that changes in technology have had, and will continue to have, radical effects in the workplace. Britain has emerged to a 'post-industrial era, a period of economic and social change, adjustment to new economic bases in Western society and the emergence of the growing economic forces of the developing nations. In particular, the technological innovations related to microelectronics and semi-conductors have had and will have noticeable effects on the labour market. "The use of electronic controls in manufacturing has resulted in both loss of employment and changes in skill distribution. New, high-level electronics and software skills have been required, whilst some craft skills have been made largely redundant. Overall there has been a great deal of craft deskilling, with an increased requirement for on-the-job training" (Rothwell & Zegweld (1979)). The role of literacy in employment has therefore taken on an increased importance due to the need for initial training and subsequent retraining. It has become necessary economically that pupils, as products of the education system, possess a range of skills, not least of which is reading. This is a far cry from the days of early liberal reform when John Stuart Mill wrote of his father's belief that "all would be gained if the whole population were taught to read, if all sorts of opinion were allowed to be addressed to them by word and in writing and if, by means of the suffrage they could nominate a legislature to give effect to the opinions they adopted" (1924, p. 74). Nowadays, reading is largely a *sine qua non* and the rate of technical change means that the mismatch between the "abilities" of pupils and the requirements of jobs becomes increasingly large.

The net result of these discussions must be to conclude that, for economically imperative reasons and to allow pupils to assume a full adult role in society, an attempt must be made to minimise the mismatch in skills considered here. Of course, there are other things to be done, also of fundamental importance, such as promoting computer studies, studies in life skills for the 16 to 19 age range, etc. In the area of reading, employers are looking to the education system to keep abreast of change. There exists a political climate within education which emphasises vocational preparation.

In considering what action may be taken, one may return to the Inspectorate report. It reported that a number of schools (about one in ten) were found where the provision for reading was memorable" (p. 72). They found the provision in the majority of schools less encouraging, however, as has been described. One of the problems outlined was the pressure of examinations, resulting in their statement that "pupils were often reading little or more than they were directed to read, and that only in so far as they could see it to be necessary for examination purposes" (p. 74). They did not claim that all schools or all pupils suffered from this problem. It is clear, however, that the forms of language used in schools, particularly at fifth form level, are unlikely to be similar to those used at work. Reading for learning would seem to be out of place on the shop floor where reading to perform some other, non-reading, task would be appropriate. The higher ability pupil, of course, may be expected to cope - but, as we have seen, high flyers tend not to leave school at 16.

Whatever is the case, neither employer, teacher or politician can undertake to ameliorate the situation without discovering in detail what the mismatches specifically are. A spirit of co-operation towards the solution to this question was commended by the House of Commons Expenditure Committee (H of C, 1977) and the CBI have stated that "Employers accept that they have an obligation to state more clearly the degree of proficiency and level of competence which they require in the basic school subjects" (H of C, 1977, p. 139). On the mathematics side of basic skills, some large steps forward have been made (e.g. Knox, (1977), Fitzgerald (1979)) but moves in considering reading requirements have been much fewer, although a number of local and individual initiatives have been made, particularly between certain schools and their local factories or offices. In 1977, Coventry Local Education Authority arranged for two teachers to make regular visits to local factories or other places of work to investigate language and number work requirements (T.E.S. 4th March 1977). Somerset has set up an Education/

Industry/Commerce Working group to study aspects of working life amongst school-leavers, including a detailed survey of basic skills requirements of some jobs. The recommendation that "some analysis should take place of the nature and degree of literacy required in various jobs" (1977, p. 10), by the Expenditure Committee could be said to describe part of the SOFRP, although it was not prompted directly by it.

Local Conditions

There were several local factors which are important in terms of perspective. The Project was, of course, sited in Sheffield as it was based at the Polytechnic, but there are a number of factors associated with the city which make it a particularly appropriate location. Sheffield is a major manufacturing area of Britain, not least for its steel and cutlery industries but also is becoming increasinly important as a commercial centre. Thus the range of employers and types of job available to the Project was not only broad but also likely to be representative of other major industrial and commercial cities. Of course, London might be the exception to this but there is little reason to suppose that the reading requirements in, for instance, the construction industry in Leeds, Birmingham and Bristol are radically different from those in Sheffield.

Sheffield also has a fairly progressive Local Education Authority, with its own officers for research and evaluation. Support for the Project was welcome from this quarter and assisted in facilitating contacts in schools. The Careers Service has a Special Measures Team to cope with large-scale unemployment amongst school-leavers, also willing to assist in contacting employers and training organisations.

There may also be said to have been a degree of disquiet over the large-scale use of selection tests by employers. A number of employer-education liaison meetings had already taken place prior to the beginning of the Project. It was the case, therefore, that a climate of general welcome existed for any investigation of local reading requirements.

Certain moves by the Polytechnic itself are also relevant. An associated study, into the perceptions and attitudes of school-leavers to school, work and unemployment (Fleming & Lavercombe, 1979) was established, in line with the desire to continue research in the field of 16+ education. The basic remit of a Polytechnic also includes the forging of links with local industry, commerce and schools.

The Value of a Test

The question of the definition and methods of assessing functional reading ability and the nature of criterion-referenced measurement will be discussed in detail in subsequent chapters. It will, however, be useful to discuss here the reasons for developing a test and the likely characteristics that such a test should have.

Since the above discussion implies that the weight of basic skills education should lie on the shoulders of teachers, it follows that the research should be oriented towards the needs of the education system to fulfill this task. A number of different investigations could be made into the nature and degree of reading required in various jobs. One could, for instance, undertake a large scale survey and publish detailed pro forma of the reading requirements of innumerable jobs. Or one might attempt an analysis of the linguistic factors of reading at work and attempt to produce some synthesis applicable to all jobs, and so on. The development of a test, however, has a number of positive features to recommend it as an appropriate method. A test effectively focuses the attention of the tester on the topic he is testing. Hence, in job-related reading, the user's attention is drawn both to the related careers guidance function of all secondary teachers, and to a scrutiny of his present curriculum in relation to the content of the test and the performance of his pupils on it. The whole Project is aimed at school-based assessment and remediation, hence it is appropriate to devise a test.

Further, an investigation of job-related reading requirements also brings the nature of the reading materials involved to the attention of a number of employers. This in itself is likely to promote links with education and/or a reassessment of the quality of the employee needed for the job and/or a reappraisal of the reading materials themselves.

Of course, the technological advances may easily make any test outmoded very quickly, but once procedures for assessing job-related reading requirements have been established, reassessment and redevelopment of a test is quite simple, and has the advantage of keeping the finger of awareness on the pulse of change. Moreover, the original development provides a base line against which to measure change.

Many of these functions could be performed by a 'non-test' study, but a test has other advantages. Few teachers are prepared to wade through long reports on different reading requirements or research papers, simply because of the pressure of their work does not allow such activity except by the most dedicated or inexhaustible. A test can be a straightforward, focussing device of some simplicity in administration and interpretation. Also, both teachers and pupils are familiar with tests and possibly a reading test related to work may provoke more enthusiasm from the pupils than other school-based tests.

There are certain <u>a priori</u> conditions within the remit of SOFRP. To fall within the scope of being easy to administer, work and interpret, any test designed should last no longer than a double school period (about 70 minutes), it should use objective items (with one and only one way of answering correctly) and should be a 'paper-and-pencil' test with little or no writing.

Further, the interpretive load is lessened if the test items reflect, insofar as is possible, the reality of the reading situation, whatever that may be. For there may well be underlying factors to the processes involved, but a teacher will be more interested in a test with clear face validity as well as construct validity.

The study of reading tasks would concentrate on the first six weeks after obtaining a job, and periods of seeking employment. This was on the basis that, beyond that period, a young person has become 'socialised' into the language forms of the particular job in which he was employed. One aim behind the Project was to minimise the transition period between school and work by promoting the reading skills required for successful job performance. Of course, other factors contribute to success or failure. To minimise the problems in one area was seen as worthwhile.

Conclusion

The SOFRP was established at Sheffield City Polytechnic in supportive local conditions, to study in what had become a major area of controversy: the reading abilities of school-leavers in relation to employment. Many accusations and complaints had been levelled against the education system for "failing to provide the goods" and, whereas a number could be refuted on logical or factual grounds, the concern over a mismatch between the abilities of school leavers and the requirements of work persisted. Increasing technological advance was seen as adding to that mismatch and this economic imperative, combined with a period of retrenchment and a trend towards an accountability movement in education, were seen as promoting the general social impetus behind research of this kind. It was felt that a test had advantages over other forms of reporting, in that it would usefully focus attention on the salient points of language in a minimally disruptive fashion.

CHAPTER 2

THE DEFINITION OF FUNCTIONAL READING ABILITY

"Literacy is a term which is beginning to take on less concrete meaning. This is especially so as more interest and emphasis is being given to it" (Powell, 1977). (Functional reading is one aspect of functional literacy, and the definitions given below will be assumed to include reading unless a distinction can be usefully made). Without a stable meaning to a term, it is impossible to attempt any assessment related to it. So in this chapter, the various definitions assigned to literacy, functional literacy and functional reading ability will be discussed and then an appropriate selection made.

A number of definitions have been made which may be considered 'global', in that they deal with man in society and the literacy skills he is called upon to use. Thus, literacy has been defined such that "a person is literate when he has acquired the essential knowledge and skills which enable him to engage in all those activities in which literacy is required for effective functioning in his group and community, and whose attainments in reading, writing and arithmetic make it possible for him to continue to use these skills towards his own and the community's development and for active participation in the life of his country" (UNESCO, Stanley, 1972 p. 382). Similarly, Bormuth defines literacy as "the ability to exhibit all of the behaviours that a person needs in order to respond appropriately to all possible reading tasks" (1973-4, p. 22). The Right to Read campaign used the definition that "A literate person is one who has acquired the essential knowledge and skills in reading, writing and computation required for effective functioning in society, and whose attainment in such skills makes it possible for him to develop new aptitudes and to participate actively in the life of his times" (Ahmann, 1975). Other similar definitions are available (e.g. Hillerich (1976), Stanley (1972)).

These global definitions have certain common features: they imply a set of stable skills, used effectively in response to some set of situational requirements, coupled with capacity for learning new skills or behaviours.

In contrast, a second set of definitions include those which may be said to involve minimal behavioural criteria. The United States Office of Education has defined functional literacy as 'the ability to hold a decent job, to support self and family, to lead a life of dignity and pride' (Harman, 1970, p. 227); whilst the Bureau of Census uses the definition of 'able to read or write a simple message in English or some other language' (Kirsch & Guthrie, 1977-8). The U.S. Army, during the last World War, used functional literacy to mean "the capability of understanding written instructions necessary for conducting basic military functions and tasks" (Ganopole, 1978).

It is unlikely that many people would concur that the ability to read or write a simple message is an adequate definition of literacy. If nothing else, it fails to indicate what, if anything, else the person can do. In the same way, the realisation of culturally valued goals suggested by the Office of Education, tells one nothing about the actual literacy component in achieving the goals. It assumes if one is literate, these things will come and in a period of high unemployment, that is a very dangerous assumption. The Army definition leans toward a more useful definition and may be linked with the global definitions, save that it does not imply a learning capacity.

A very large number of definitions are based on the amount of time spent in school. UNESCO suggest that four years of primary schooling implies literacy. The National Health Survey in the United States stated that "literacy is that level of achievement which is attained by the average child in the U.S. at the beginning of the fourth grade (Kirsch & Guthrie, 1977-8). The Bureau of Census used, in addition to its definition above, a level of fifth grade (Bormuth, 1973-4). Boulmetis (1973) suggested the eighth grade.

Despite the attraction of such definitions in terms of their ease of assessment, they can have little, if any, validity (Hillerich, 1976). They suggest that a line can be drawn to separate, in some way, the 'illiterate' from the 'literate' - a false dichotomy, considering that it suggests a person could be illiterate one day and not the next. Further, there is no reason to suppose that a person completing four or five years at school has gained anything from those years except age. Hence, "if a specific grade level is designated in the belief that it assures adult reading success, it is based on an unproven assumption" (Ganopole, 1978, p. 13). A variation on the grade level definition is the achievement grade level; here, a person is assigned an achievement grade level according to his performance on a reading test. His level represents his performance in relation to the average ability of a grade group as we have seen in the National Health Survey. It remains unclear, however, how such achievement grade levels can be associated with other literacy behaviours, and at other ages. Prediction of success is notoriously difficult (see e.g. Fisher & Brown (1971); Spache (1970)).

Having looked at a number of definitions (by no means all that are available) it is perhaps time to draw some threads together. The global definitions of literacy imply a potential to undertake required reading tasks, whilst the others stress cultural goals ('a decent job', etc.) or crude 'rules-of-thumb'. The definitions of functional literacy, however, have all implied, quite rightly, the actual performance of reading and writing, not merely the potential to do so. The Chambers Everyday Dictionary (1979) gives 'functional' to mean 'designed or undertaken with special, or exclusive, regard to the purpose which it is to serve'. Hence, functional reading is literally, reading which assists in some task or purpose, 'a continuous process of applying specified skills to specified tasks' (Kirsch & Guthrie, 1977).

Global definitions are, in fact, general statements about a very large set of potential, task-related literacy behaviours and skills. In themselves they give no pointers for possible assessment procedures, which is, after all, the crux of the present study. It is necessary to operationalise a definition with which to work, and this is clearly done through an examination of definitions based on tasks and purposes. The U.S. Army definition given above is a partial answer; it defined the need to perform given reading tasks. Indeed, the common factors derived from the global definitions tend to complete the picture: a set of stable skills, used effectively in response to some set of situational requirements, coupled with the capacity for learning new skills or behaviours.

The definition closest to this synthesis comes, perhaps unsurprisingly, from another U.S. Army study. Sticht (1975) defines functional literacy as "possession of those literacy skills by an external agent between the reader and a goal the reader wishes to obtain" (p. 4). Thus, it is not simply the reader, but also the task which needs to be considered. Sticht's work is particularly relevant to the present study. Project REALISTIC (1972) investigated the reading demands of a number of military occupations and concluded that different jobs required different reading levels. This is especially important when thinking of the rather sweeping comments from employers in Chapter 1, where it was tended to be assumed that some form of basic level of reading ability was all that was needed for most jobs. Sticht found differences of up to two grade-levels between jobs.

In the context of the present research, one can use Sticht's definition to produce a useful, operational definition of occupational functional reading ability as "the level to which an individual possesses those reading skills needed to perform successfully those reading tasks required of him in seeking a job, at work and in related training".

There are a number of implications contained in this definition. A continuum of skill level is suggested, from the non-existent to the complete, and this seems appropriate when there are many tasks imposed on the reader. He may cope with some but not all. It also implies that to please completely his employer by performing his complete job successfully, he must also successfully perform all his reading tasks. (It may be, however, that the employer, despite the protestations noted in Chapter 1, may not require 100% performance levels.) Although it does not clearly say so, the definition also implies a development or learning associated with the performance of reading tasks, for one does not expect a person to remain at one and only one level of skill.

Having derived a useful definition of occupational functional reading ability, it is appropriate to turn to a consideration of how levels of this ability may be assessed.

CHAPTER 3

THE ASSESSMENT OF FUNCTIONAL READING ABILITY

The original remit of the SOFRP calls for the design of a group test to assess levels of occupational functional reading ability. It may seem strange, therefore, to spend time in consideration of how to assess functional reading ability. In fact, to specify that the assessment shall be by testing gives no clue to how that test shall be developed, what format it shall take or its relationship to the concept of functional reading. For instance, one could see how long it took pupils to answer a mathematics test and then use this as a measure of functional reading. It would hardly be valid, however, as the processes of design and the content of the test are unrelated to functional reading. Nor is it possible to ignore the fact that many tests of reading ability already exist.

Perhaps it will be useful to deal with that last point first: why existing tests of reading ability cannot be used to assess levels of occupational functional reading ability. Firstly it must be said that one would be much happier about using such a test than the mathematics test mentioned above. It can be assumed, with a fair degree of certainty, that functional and general reading abilities will be related, in that one is a specific exercise of the other. Standardised reading tests, however, are designed to evaluate the abilities of an individual in relation to the abilities of his peers (actually, in relation to the assumed distribution of those abilities) and are therefore inappropriate to the assessment of the ability to perform a specific range of tasks. Of course, the testee has performed specific tasks in undertaking the test and one could argue that this is a sufficient demonstration. Yet the tasks he has undertaken are designed to produce a particular (normal) distribution of scores <u>across all testees</u>, rather than produce a score for <u>that</u> individual regardless of the performance of others or the distribution of scores across those others. This difference in types of test and their implication for item and test design will be discussed in detail in Chapter 4. It is clear, however, that a test which is designed to rate testees relative to their fellows would not suit the remit of the Project, in that it lacks evidence that its tasks are representative of job-related reading tasks, and that it does not give a performance measure demonstrating the functional reading ability of the individual, only his performance in relation to others.

Moving on to consider what may be done, one is especially fortunate in the wealth of work with the 16+ and adult age-ranges in the United States. Many methods used may be instructive, errors avoided, theories useful.

The accountability movement in the United States has meant that basic skills testing is extremely common. At least 31 of the 50 states have laws or board of education rulings which mandate that high school students display mastery of basic skills (i.e. 'minimal competency') before being granted a diploma (Ganopole, 1978). The 'display of mastery' generally takes the form of adequate performance on several tests, amongst which there is always a reading test. A large number of the bodies concerned decided to formulate their own tests or to commission them from major testing authorities. The procedures involved in such formulation tended to be similar. In the first instance, objectives were developed, usually stated in the form of "To achieve competency, a person must display x skill". These objectives were in the main developed by panels of local teachers, school board representatives, community representatives and politicians. For example, in Phoenix, Arizona, 'the first step ... was the formulation of objective performance levels of reading proficiency. The committee chose to develop objectives in basic reading skills which an adult in our society needs in order to function without handicaps" (McDonald & Moorman, 1974).

Following on from the specification of objectives, the next step in the procedure was usually the design of test items. The Maryland assessment work was not unrepresentative in generating "a complete and comprehensive bank of over 500 test items to correspond to the approved series of behavioural objectives ... all of the original items generated were multiple choice, true-false, or matching items" (Petre & Major, 1976). Multiple choice items were by far the most common (see e.g. McDonald & Moorman, 1974; Ganopole, 1978; Fillbrandt & Merz, 1977, etc.). Where there was more variation, however, was in the use of reading materials upon which items were based. Some studies looked for 'real-life' materials to suit their objectives (MacDonald & Moorman, 1974) whilst others invented their own (Petre & Major, 1976; Ayrer, 1977).

Pilot testing and item reviews were most commonly the next move made. Items were given to large numbers of the relevant target population to see how the items performed and how testees performed on them. Tests were then assembled with controlled difficulty ranges, often with readability indices available (e.g. Ganopole, 1978). Measures of reliability were made and the final product used in determining the level of competency of each student in the target population. The use of a certain percentage correct as a 'cut-off' or 'criterion score' has also been almost universal.

Those familiar with the development of any form of educational or psychological test will recognise the essential stages covered. They will also notice the glaring omission of formal validation procedures. In a number of cases, procedures were simply not reported. In the course of developkng the Wisconsin Test of Adult Basic Education, for instance, the only report of validity measurement was to indicate that field trials had been conducted amongst 120 high school students and 37 adults, resulting in unreliable items being eliminated. No mention was made of what constituted an unreliable item (Nafziger et al, 1974). In other cases, it is made clear that content validity was assumed to be assured because of the procedures of item construction (e.g. Petre & Majer, 1976) or item reviewing. Other forms of validation were much rarer, however. This is unsurprising in view of the fact that other forms would involve more expenditure - the very thing the accountability movement in general attempts to control. One or two studies have come to light, however. An attractive study in the validation of a cut-off score was undertaken in California, where 'entry level job holders' were tested to see how well they performed on competency tests. Similar scores were then used to judge the competency of students (Fillbrandt & Merz, 1977). The Basic Reading Inventory was found to have a degree of concurrent validity by comparing its placements with teacher judgements and with other reading tests (Burnett, 1966). The author has been unable to discover any state study that has validated its test by following its testees up and assessing their 'success' compared with the test's predictions.

There have also been a number of national, large-scale studies in the United States. Perhaps the most famous of these, presumably for its being undertaken by an opinion pollster as well as its widely publicised results, was the 'Survival Literacy Study' (Harris, 1970). In this survey, Harris and his associates used facsimile (but altered) versions of four commonly used forms and asked respondents to fill them in. They reported results to the effect that 18.5 million Americans were functionally illiterate (more than 10% of answers in the forms were incorrect). This study can be said to have started the ball rolling in much the same way as James Callaghan's speech did, for it received great publicity. Unfortunately, a scathing refutation of virtually all its claims and an accurate attack on its methodology did not receive equal media coverage (Caughran & Lindlof, 1972).

The National Assessment of Educational Progress (NAEP) (Gadway & Wilson, 1976) can be seen as a monitoring exercise similar to that of the Assessment of Performance Unit in Britain, and there are a number of similarities, as will be discussed below. The NAEP essentially has been a larger-scale version of state programs, except that it occurs at the group rather than individual level by testing a sample of a given age range across the nation, and that there is no question of an individual's competency, but that of standards in basic skills nationwide. The NAEP has concentrated on defining objectives, developing test items, piloting, review, assessment of reliability, etc. Despite being criticised for 'many serious conceptual, technical and procedural deficiencies' (Greenbaum, 1977), a number of its activities have proved of interest. In science assessment, exercises have involved 'hands-on' practical testing. Group discussion exercises have been studied. Such innovation must be seen as a valuable step away from conventional practice, if sufficiently objective criteria can be established for assessment. Interesting also is the concept of 'directionality' in repeated testing, where 'any detectable change in performance in a 'desireable' direction may be taken as an indication of progress' (Burstall, 1978).

Lichtman has developed the R/EAL (Reading/Everyday Activities in Life) test (1974), again by the use of real-life materials conforming to previously determined 'essential' objectives. Her test is self-administered by the testee, using tape-recorded instructions and questions. Non-reading variables are therefore minimised. Validity was considered in two ways, here; content validity relied, as in many state programs, on the construction of items directly related to task analyses of basic reading skills; criterion-related validity (or 'concurrent' validity) was estimated by parallel assessment using the Stanford Achievement Test.

The Adult Performance Level (APL) Study (Northcutt, 1975) has developed literacy objectives and accompanying test items. This study is interesting here because of its consideration of the cultural content of literacy and, in particular, that it is technology-bound. "The implication is that literacy must be redefined as technology changes" (p. 44), a conclusion to which allusion has already been made.

All of the studies, at local and national level in the United States, have a common feature: that of the prior development of test objectives before commencing further development, especially item construction. One may conclude that the reasons for doing so, rather than an investigation of actual reading requirements *in situ* (e.g. at work, leisure or as a consumer or citizen), was that the brief - to consider minimum competency in a range of subjects - meant that such investigations were beyond their resources; or that they had decided to be somewhat prescriptive about what *should* be the skills required. Considering the definitions of functional reading previously discussed, this latter position is largely indefensible as either minimal competency or functional skill, for it assumes requirements that have not been shown to exist beyond the opinions of those involved in test design.

In contrast to the previous works, a few studies stand out as employing different methods. Perhaps the most detailed of these are the various researches of the Human Resources Research Organisation (HumRRO) into the literacy requirements of certain occupations in the U.S. Army (Sticht & Caylor (1972); Sticht et al (1972); Sticht (1973); Sticht & McFann (1975); Sticht (1975)). It was Sticht's definition that proved most useful earlier. In the pieces of research undertaken by HumRRO for the Army, detailed investigations were made into the reading requirements of jobs by discovering what was read in the course of work and in what circumstances. This yielded valuable information about how much and how often the job incumbents had to read. They also developed a readability index for use with the materials and so assess the discrepancy between the measured general reading ability of the incumbent and the difficulty of the texts with which he had to deal. Quite large discrepancies were discovered, seeming to demonstrate that functional reading does differ from general reading ability. Job-related reading task tests were designed, using actual materials and simulating actual conditions of use. Further, field observation was used as another method of assessing functional reading ability.

A second major study was that undertaken by the Educational Testing Service for the United States Department of Health, Education and Welfare (Murphy (1973 and 1975); Jackson (1972)). This study was concerned with adult reading tasks and their initial remit was as follows. They were to undertake "(1) a national survey of adult reading activities; (2) the construction of a set of reading tasks intended to sample the universe of reading tasks encountered by the ordinary American adult in our culture; (3) a national survey of actual adult performance on the set of reading tasks constructed; and (4) a study to determine the resource ceilings within which instructional systems of reading would have to operate". (Murphy, 1973, p. 4). Their initial approach was to interview a representative sample of American adults (over 16's) and over 5000 interviews were conducted. Each respondent was asked to identify what he had read in the twenty-four hours previous to the interview, to estimate how long he or she had spent on each reading activity, how important they felt each activity had been and how difficult, and, if it was meaningful to ask, why or in what circumstances they undertook each reading activity.

This study is especially interesting in that they also attempted to identify the 'benefit' to individual and society. They used large panels of judges to consider what tasks should be used; as Murphy puts it, "to identify a set of adult criterion reading tasks which adequately sample tasks for which highly favourable returns to the individual and to society can be demonstrated". Also important is the rejection of the assessment techniques common to the state school testing programmes, for the purposes were entirely different. "The tasks to be developed were not subject to the usual constraints of conventional test items.

There was no requirement that it be possible to administer all tasks in a specified length of time. Tasks did not have to be fashioned in a format for group administration. It was not necessary to fashion tasks in a multiple choice format with clever distractors for options. Developers were encouraged to be imaginative and creative ... The only constraints were that the tasks be life-like, that they be beneficial to individuals who can perform them; and that a successful response to a task be observable and scorable" (Murphy, 1973 p. 49).

Items were again validated by panel studies and the tasks used to assess, not individual performance, but national levels of competency on a range of the tasks. As this assessment was for descriptive purposes and used to evaluate relative performance between different subgroups of the American population, no validation was attempted to see whether higher 'success' rates on the criterion tasks were related to higher 'success' rates in life. Murphy (1975) suggests, however, that it may be possible to set reasonable functional reading standards for school-leavers.

In the United Kingdom, far less work has been undertaken in the area of functional skill assessment. As part of the Adult Literacy Programme, tests were to be developed within a framework "intended to relate to tasks necessarily undertaken or tasks encountered by the majority of participants in the adult literacy programme" (Gorman, 1977). Here, the tasks were defined by the discovered needs of those who acknowledged they had problems in literacy, rather than being either imposed by experts or surveyed generally.

The Assessment of Performance Unit was established in 1975, following the White Paper on Educational Disadvantage (DES, 1974), and was "to promote the development of methods of assessing and monitoring the achievement of children at school, and to seek to identify the incidence of under-achievement" (DES, 1978). The activities of the Unit can be shown to be very similar to those of the NAEP, with the involvement of people from different interested areas, assisting a core of professional researchers to assess performance nationally. Also no individual results are reported and the principles of 'light sampling' are employed (multiple-matrix sampling, see Shoemaker (1973) or Shoemaker & Hornke (1976)). In the assessment of reading, they "broadly classify reading in terms of the general purposes for which pupils of different ages are required or expected to read ... The intention is that the variety and the level of difficulty of the materials used for assessment should reflect the range of reading matter currently encountered by pupils aged 11 and 15, insofar as this can be established" (APU, 1978). Pupils are also required to demonstrate searching skills in relation to written materials.

Although the Unit has not yet published either detailed procedures, results or many example items, one may discover that items are based on collected, rather than devised, materials, such as passages from encyclopaedias, and that paper-and-pencil, objective testing instruments are being used (Gorman, 1980). Further, the large scale banking of test items by computer is being undertaken (see e.g. Wood & Skurnik (1969); Byrne (1975)), using unidimensional item scaling on the Rasch model (see e.g. Rasch (1960, 1961, 1966); Choppin (1977); Willmott & Fowles (1974) or Wright (1967, 1977)).

From this review of recent American and British practices in the field of basic skills assessment, it is possible to extract some useful techniques and considerations which are of value to the Project.

It is proposed to consider first the question of what should constitute test materials. On the one hand, a great number of local studies in the United States have decided that test materials shall be the product of a prior specification of objectives. On the other hand, a number of larger scale projects have adopted the view that 'real-world' tasks should be identified and then comprise test materials. If one bears in mind the conditions prevailing at the time of the SOFRP's inception, it may be recalled that it was particularly because the specific requirements of jobs and of job-seeking had not been identified, that the perceptions of mismatch by employers arose. Here, one is confronted with two approaches to a solution. The former suggests that requirement can be identified and specified by panels of experts (to use the term fairly loosely), whilst the latter approach suggests that requirements are best identified by observation, illumination and investigation. In the end, the balance must go towards the investigative approach, as holding more weight via objective evidence, than the panel approach where opinion is too much involved. Perhaps this is unjust to potential panelists who might well be of the highest quality in their field, but to do their job well, they must assure themselves that they know what are the requirements of their area. Hence, they themselves are forced into an investigative role. Further, panels in America have tended to prescribe levels of ability, where this Project is concerned with establishing necessary lower limits of performance by describing the current state of affairs. It would seem, then, that the test materials developed in the SOFRP should be the product of investigation and be constituted of materials derived from 'real-life', rather than be the products of construction to suit some pre-specified objective.

Several studies offer a chance to consider item types and administration and response modes rather different than those commonly used. Here, the remit of the Project is rather more restrictive, however. Lichtman's use of tape recorders to enable pupils to pace themselves and to minimise non-reading variables is particularly attractive. However, it would be true to say that young workers are rarely allowed to pace themselves at work to anything like the degree Lichtman suggests. Also, the resources for group testing by this method are extremely limited, and completely rule out any testing of a group above, say, class size within a period acceptable to most teachers. In the same way, much of the Adult Functional Reading Study (Murphy, 1973) work on novel assessment procedures may not be used. Their use of what were basically interview techniques - slow, individual testings - are completely inappropriate. This does not mean that one is restricted to familiar modes, however. Murphy's attempts to mirror 'real-life' tasks must serve as a pointer to a wider consideration of items beyond the multiple-choice. In all then, it seems that the a priori limitations on test development do not allow some of the more interesting innovations used with individuals or small groups. One is not restricted to standard item types, however.

Without wishing to pre-empt later, more detailed discussions of particular issues, one notes the very limited attention paid to questions of validity outside of the processes of test development itself. Sticht's work stands out as virtually unique in considering how well men did perform functional reading tasks compared with how well they performed on assessment instruments. Such predictive validity will be an important feature of any test developed during SOFRP, for the minimising of leaver-ability and job requirement mismatch implies that those who would fall short of requirements have been identified and remediation made available. The rejection of the prior specification of objectives paradigm also leaves open the possibility of content validation by means other than attempting to judge if an item matches an objective, or concluding that it surely must, 'all other things being equal'.

The specification of a 'cut-off' score needs very careful attention. The relationship of this concept to measurement theory is discussed in the next Chapter. Here, the question is whether or not a 'cut-off' score can be determined and, if so, if it is desireable to do so. It is, of course, statistically possible to determine a level of performance below which testees could be termed 'unsuccessful' and above which they could be termed 'successful'. A certain amount of error would be associated with such a level, however, such that errors of misclassification were possible. The use of a single level (or 'score') must therefore be arrived at following a consideration of the relative importance of misclassifications. One could specify a range of levels or scores, within which one might recommend that a tester 'look again' at the ability of the testee; for situational factors, such as illness, can cause such anomalies that a rigid 'cut-off' would ignore. Some authors have suggested that the testee's level of ability be used to describe what he can and cannot do, rather than simply classify him (Ayrer, 1977). Whilst this has many positive features, they are outweighed by considerations of teacher usage and interpretation in the first instance. Of course, the practice of 'looking again' implies scrutiny of actual skills and deficits after classification. Novel methods have been used to assess levels of acceptable performance (e.g. Fillbrand & Merz (1977); Murphy (1973)) and should receive attention. A much stronger attack on the prior setting of a standard has come from Glass & Smith (1978) who write: "We have read the writings of those who claim the ability to make the determination of mastery or competence in statistical or psychological ways ... We regard the language of performance standards to be pseudo-quantification, a meaningless application of numbers to a problem ill-suited for quantitative analysis" (p. 12). It would seem, anyway, to be the case that, in using an investigative approach, it is not possible for the SOFRP to pre-specify an acceptable standard and that empirical means must be considered.

Although few of the studies mentioned here make specific reference to the use of computer techniques, it does seem that a consequence of an investigative approach combined with a consideration of novel item types, predictive validation and an investigation of 'cut-off' scores, is likely to generate a great deal of information. Whether numerical or not, it would appear that the assistance of a computer might be required.

It is to be expected that most readers will have difficulty in dealing with reading passages that are too complex for them to understand. It follows that if some measure of the degree of congruence between the complexity of the test and the reading ability of the reader could be found, then ways might be found to maximise that congruence and minimise reading difficulty, errors and failure. With job-related reading materials, there are two approaches to this problem: the measurement of the complexity of texts in terms of the reading ability required to deal with them; and the measurement of reading ability in relation to texts the readers are required to use. The former approach is the assessment of readability; the latter, of course, is the measurement of functional reading ability.

The prime reason that measures of readability are used is to select texts appropriate to the level of reading ability of the pupil. The texts under consideration here are fixed rather than available for selection, as they are those reading materials collected during fieldwork, and so the use of readability measures for this purpose would be irrelevant. Nonetheless, some indication of the complexity of the various texts involved would be useful in the construction of test difficulty. Indeed, Sticht (1975) suggests that "The major usefulness of an appropriate readability index is that it permits an immediate estimation of the reading ability level required to understand a passage." (p. 16). Despite the fact that Sticht and his colleagues were investigating in detail the reading requirements of four jobs in the U.S. Army, a number of factors militate against the use of measures of readability in this context.

One may deduce from Sticht's contention that, for a given passage, its readability index roughly indicates a certain level of reading ability required and that comprehension of the passage by persons with less than that level is unlikely. From this, one may conclude that a test of general reading ability will serve as a sufficient measure of functional reading ability, providing the readability of the materials can be established. Sticht himself, however, provides evidence that this is not so. One would expect successful incumbents of jobs (i.e. those successfully performing all tasks required, both reading and non-reading) to have reading abilities related to the difficulty of texts they must use. Yet, Sticht found mismatches of five or six grade levels in a comparison of reading test score and readability indices and stated that "... Since the readability formula provides a roughly accurate index of the difficulty of the materials .... (the data suggests) that both high and low aptitude personnel would experience difficulty in reading and comprehending" the materials for these two jobs (Supply Clerk and Motor Repairman) (1975, p. 52). As this was not the case, for the incumbents were performing at least adequately in their jobs, one must conclude that, even where the readability of texts can be established, an indication of general reading ability is an insufficient measure of functional reading ability, or that the measure of readability was at fault. Indeed, this latter seems the better contention for general and functional reading abilities are highly likely to have some strong relationship with one another (and this is investigated empirically below, Chapter 14) and the one a good estimator of the other.

Sticht used the FORCAST formula (1973) to assess readability and, indeed, there are a good many formulae available. Dunlop (1954) identified 56 with Klare (1974-5) adding a further 23 in his review. Based as they are, largely on the structural properties of text, they fail to take into account a number of factors which take on crucial importance in the functional use of reading materials, rather than their use for general reading. "Readability formulas (sic) do not generally consider such variables as levels of abstraction, complexity of concepts, figurative and poetic language, multiple meanings, technical and scientific vocabulary, variations in format and organisation, and a host of other things related to the comprehensibility of subject area reading materials ... Neither do they take into account the variability of reading difficulty within text material except in the averaging process. More important, readability formulas do not measure the interest, the motivation, the language competence, or the experiential background of the reader in relation to the specific content of the text" (Nelson, 1978, p. 621-622). Harrison (1979) places some emphasis on motivation as a factor in going beyond one's normal reading competencies to higher levels, because of the particular importance or interest of the reading materials. It may well be that Sticht's results are a product of high motivation due to the necessities of job performance.

It is also important to note that the common pattern of readability formulae involves the use of continuous prose in a way that may be inappropriate to job-related reading materials. Such materials as job cards, price lists, etc. clearly do not have the common structural properties upon which the formulae are based.

In conclusion the, success on job-related reading tasks is a function of the difficulty of the task and the functional reading ability of the reader. As the reading texts upon which the tasks are based are being sampled and collected, there is no question of matching the texts to the general reading ability of the pupil, and the use of readability measures would be inappropriate for that purpose. In fact, performance on job-related reading tasks seems unrelated to the readability of the materials involved, within the context discussed. The nature of readability formulae make them inappropriate for use with some of the expected forms of reading materials, and irrelevant to any more detailed investigation of textual complexity beyond structural properties. Indeed, it seems that the best measure of the readability of job-related reading materials is how well readers deal with them in terms of the tasks involved. As this is the prime concern of this study, it would appear contradictory to use some other measure.

The usefulness of any educational instrument is a product of the quality of the instrument qua instrument and the strength of its relationship to the area under consideration. Here some factors in the developemnt of a test of occupational functional reading ability have been discussed, and may be seen as contributing to the strength of the relationship between a test and functional reading ability. Having determined methods which one hopes will enhance the process of development, it is necessary to turn to the second factor and consider the processes by which the quality of the instrument may be ensured.

CHAPTER 4

CRITERION-REFERENCED MEASUREMENT

Norm-referenced and Criterion-referenced Measurement

In the previous Chapter, the use of a standardised test of general reading ability was rejected for two reasons: the lack of any demonstration that the content of such a test was related to functional reading tasks, and that it was founded upon a measurement model inappropriate to use in the present circumstances. In the present Chapter, a more appropriate model is advanced and compared with the standard model. The various approaches and implementations of the new model are considered and those facets most useful to the SOFRP are selected. The statistical implications and problems in relation to this model are discussed.

It is, perhaps, most useful to go into a little more detail about the standard or classical testing model. Most tests of general reading ability are concerned with assigning a score to a testee such that his ranking in ability relative to his peers can be assessed. Such a test is known as a norm-referenced test (NRT), due to the underlying assumption of the measurement model - norm-referenced measurement (NRM) - that the distribution of scores of testees can be fitted to the normal (Gaussian) distribution. Hence, testees can be identified as below, at or above the average level for his group (a group typically being all pupils of a given age, or school year, etc.). Now, the content of such tests is devised to reflect those reading tasks that the group should be able to undertake and therefore the test score can be said also to identify what a given testee can and cannot do. The construction of a NRT, however, involves the manipulation of the difficulty levels of the items and the content of items (including distractors) in order to increase the variability of the scores achieved by testees. This means that items appear in the test less for their representative nature vis-a-vis the content of the area being tested but because they are the ones which possess the appropriate statistical characteristics from the available stock of items. It is the contribution to the variability of test scores that is all important in NRM as the test is to evaluated in terms of the ability of items to discriminate among testees (Gronlund, 1977).

Whilst it is often important to discriminate between testees, particularly in schools and colleges, it is not always the appropriate measurement to make. A NRT will allow a teacher to see who are the brightest and who the slowest in the class. What it will not usually do is to indicate what each of the pupils can and cannot do. That is, it gives little or no clues on 'subject matter proficiency' (Glaser, 1963). A test that will do this may be termed a criterion-referenced test (CRT), as it provides information on the testee's achievement in relation to 'an absolute standard of quality' (Glaser, 1963) and 'the meaningfulness of an individual score is not dependent on comparison with other testees' (Popham & Husek, 1969). A much used and excellent example of a CRT is the driving-test, in which the testee has to demonstrate proficiency in a number of specific ways in relation to a set, absolute standard: acceptable driving behaviour. How other testees perform on the test is irrelevant; though how other drivers perform in the real-world task is important in relation to 'acceptable driving behaviour', the criterion to which the test is referenced.

Definitions and Approaches

Such tests are clearly linked to decision-making e.g. whether to give remedial help to a less proficient testee, to promote a child to a more difficult reading book, or whether a certain course of teaching has been effective in getting some subject matter over to the students. "Criterion-referenced tests are devised to make decisions both about individuals and treatments" (Popham & Husek, 1969). Two major approaches may be discovered, directly resulting from how different authors have seen decision-making.

By far the most common strand of CRM testing has been linked with instructional technology. Educators previously used to considering NRT scores and referencing the scores of individual pupils to the mean, standard deviation, stanines and percentile ranks, etc., looked for similar single scores on which to base their decisions. Hence the concept of 'mastery' testing has grown up, where the testee is assigned to one of two or more groups (e.g. 'pass-fail'; 'advance-retain-remediate') on the basis of his score in relation to some 'cut-off' point(s) or 'criterion-score(s)'. Such a conception may be seen as arising from definitions of CRM based on narrow, tightly-defined objectives. Glaser & Nitko (1971) have defined a CRT as 'one that is deliberately constructed so as to yield measurements that are directly interpretable in terms of specific performance standards' (p. 653). Ivens (1970) defined a CRT as being composed of 'items keyed to a set of behavioural objectives', whilst Kriewall (1972) suggested that it should contain items which are homogeneous in difficulty for each examinee. Livingston (1972) used 'criterion-referenced' to refer to 'any test for which a criterion-score is specified <u>without</u> reference to the distribution of scores of a group of examinees' (p. 13), such as the mean.

Within mastery testing, Meskauskas (1976) identified two sorts of model. 'State' models see mastery as an 'all-or-none' dichotomy and such models have been advanced by Emrick (1971), Roudabush (1974), backed up by considerable technical discussion using Bayesian decision statistics (e.g. Hambleton & Novick (1973); Berk (1976)). On the other hand, some authors have advanced 'continuum' models, in which "mastery is viewed as a continuously-distributed ability ... (and) ... an area is identified at the upper end of this continuum, and if an individual equals or exceeds the lower bound of this area, he is termed a <u>master</u>" (Meskauskas, 1976 p. 134). Such models have been put forward by Nedelsky (1954), Ebel (1972) and Kriewall (1972).

The second, and much less common, approach to criterion-referenced measurement may be related to the original conception by Glaser. In his early work (1963), 'Competence is conceived as being a continuum characteristic. There are, at most, ambiguous suggestions that a single point exists at which competence becomes incompetence' (Glass & Smith, 1978, p. 13). Popham (1975) admits that his use of the term 'performance standard' (Popham & Husek, 1969) contributed to what Glass and Smith go on to call 'a case study in confusion and corruption of meaning' (loc. cit.). The essential differences here are the disinclination to use a cutting score and the narrowness of the definition of objectives. Popham defines CRT as a test that "is used to ascertain an individual's status with respect to a well-defined behavior (sic) domain" (1975, p. 130), the stress being on 'well-defined' rather than 'narrow'. This is much nearer to the original thinking in the area, replacing NRM with measurement of what a testee can and cannot do. Ayrer (1977) has adopted this approach, using his test to 'describe instead of certify', where 'its function was to show where a student was having problems' (p. 704).

Criterion-referenced Measurement and SOFRP

The use of CRM by the SOFRP is clearly dictated. "Since a primary goal of functional literacy measures is to assess achievement with respect to specified reading tasks, domain referenced tests which require generating a representative sample from a well-defined population (domain) of tasks seems most appropriate" (Kirsch & Guthrie, 1977-8). In terms of the SOFRP, both the two approaches to CRM have something to offer. It will be useful, first, to consider what is the domain with which the Project is concerned. It follows from the definition advanced in Chapter 2 above (occupational functional reading ability is 'the level to which an individual possesses those reading skills needed to perform successfully those reading tasks required of him in seeking a job, at work and in related training'), that the domain consists of the skills and knowledge needed by 16+ year old school leavers to perform the reading tasks encountered in their first six weeks in the above circumstances in the City of Sheffield. It is not the tasks which one is concerned to assess but the level of skills and knowledge needed to deal with them. This is obviously a broad domain, although well-defined in the context of the Project. It is highly unlikely that narrow, tightly-defined objectives would have any place in the development of the functional reading tests under discussion here, there would quite simply be far too many of them. For instance, to analyse each reading task for each job in each industry in the city is not only an enormous task but would generate an enormous number of highly specific objectives. This being far beyond the resources available and outside the level of acceptability to test users (who would be confronted either with a small set of huge tests or a huge set of small tests), it would be necessary to use less specific, more general objectives, encompassing, say, certain classes of job, or broadly defined types of reading task. This being the case, one is basically using the Popham definition.

In fact, the need for formally stated objectives becomes an irrelevance. If one is concerned with the identification of reading tasks undertaken by school-leavers, and the subsequent development of a test containing items based on materials representative of those tasks, the specification of objectives comes at the time of test assembly and not at the time of item construction. This being so, the objectives are only descriptions arrived at <u>a posteriori</u> and are irrelevant if the processes of development have been adequately reported.

Yet the mastery approach has something to offer from the continuum model. Whilst it has already been made clear (Chapter 2) that to call someone literate or illiterate is impossible or invalid on the basis of a single score, the use of a range of acceptable scores, or identifying a single score for further action by the tester rather than 'failure' by the testee, offers a number of useful features. Teachers will more often welcome a test which can identify both who is having problems and the area into which some of those problems fall. The emphasis in a test based on a broad domain is 'looking more closely' at those with lower scores, rather than 'fail'. Further, relating some score to high performance ratings on actual job reading tasks offers an opportunity to reduce the arbitrary nature of a priori standard setting. Also, 'above' and 'below' a cut-off score allows a dichotomy to be introduced for certain statistical measures, which will be particularly useful as will be demonstrated below.

To summarize so far then, norm-referenced measurement is inappropriate to assessing levels of occupational functional reading ability for, whilst it may order testees relative to one another, it gives no clue to what each testee can and cannot do. Criterion-referenced measurement offers this type of information by referencing performance on a test to an external criterion behaviour, in this case, job-related reading tasks. The range of skills and knowledge required in the performance of such tasks is too broad to allow the use of specific, objective-based mastery testing, although some features of that paradigm are useful. The type of CRT under development for the SOFRP conforms more closely to the view advanced by Popham.

Statistical Considerations

CRM differs in another way from NRM and that is in the statistical formulation of development and analysis procedures. As has been previously indicated, NRT are designed to yield wide variations in scores in order to discriminate between testees. In CRM, however, the variability in scores is irrelevant. "The meaning of the score is not dependent on comparison with other scores; it flows directly from the connection between the items and the criterion. It is, of course, true that one almost always gets variant scores on any psychological test: but that variability is not a necessary condition for a good criterion-referenced test" (Popham & Husek, 1969, p. 3). Items which all testees get right or all testees get wrong are invariably deleted in NRTs because they do not discriminate. In CRM, such an item might well be retained, for it provides information about each testee in relation to the domain under consideration. It is conceivable in CRM to have a test in which all testees get all items correct. It is indeed desireable in the context of a post-instruction test. The effect of the irrelevance of variability is to alter the whole statistical pattern associated with NRM.

Classical testing theory, as it has come to be called, uses certain common statistical procedures to describe and evaluate tests and the performance of testees upon them. To describe the scores of testees, it is usual to report the arithmetic mean score and the standard deviation, implicitly accepting that such measures are given meaning by reference to the performance of others on the test. Various significance tests are also based on that normal distribution of scores. In test evaluation, correlational techniques are used to relate two sets of measures, particularly in the study of the validity and reliability of the test, and to assess item discrimination.

If the test is not constructed to yield a widely spread set of scores, however, and can in fact yield a set of scores with little or no spread and still be a good test, it follows that the classical measures may be invalid and inappropriate. The mean and standard deviation may be highly misleading with a multimodal or highly skewed distribution - the score reported may not have been achieved by anyone at all. Other descriptors such as the median, mode and range may give a more accurate picture. Significance tests based on approximations to normality will have their basic assumptions violated and recourse to distribution-free statistics required.

It is in correlational techniques that variability becomes of crucial importance, however. The effect of reducing the variability of scores on one or both of the measures being correlated is to reduce the size of the correlation coefficient. Lord & Novick (1968) have demonstrated that the correlation of a set of data is always smaller if its standard deviation is smaller (p. 129-131). This is not to say that the interrelationship that the coefficient is attempting to measure is smaller if variability is low, but that the measure may underestimate the size of the relationship. Hence, the coefficient may be low and underestimate or it may be low because the relationship is low. There appears no way round this problem except to say that a high correlation will be no lower but that a low correlation may be higher.

Clearly, these statistical implications pose quite a number of problems in relation to the applicability of classical techniques in test construction. The use of a CRM model suggests that either modifications are made of existing techniques or that new techniques relate them. In particular, correlational methods need careful attention in the fields of item analysis, scrutiny and the estimation of reliability and validity. For one still needs such tools in CRM, to assess the quality and usefulness of the test developed. The specific problems and proposed solutions are dealt with in ensuing chapters, in order to present them in the contexts in which they occur. Two other points should be raised, however. The question of unidimensional scaling was mentioned in the last chapter and needs to be settled here. The model (Rasch, 1960) is based on the assumption that all items are testing along the same dimension. It is very unlikely that this will be the case in SOFRP except at a very general level. It is more likely that a functional reading test will be multidimensional at the level required by latent-trait models. Further, current implementations of the Rasch model (e.g. Choppin, 1974) require the deletion of items answered correctly or incorrectly by all testees. This is not seen as appropriate to SOFRP where the performance of pupils in relation to job-related reading materials is to be assessed independent to one another. For these reasons, the adoption of a latent-trait, item analysis model is not considered correct.

Finally, the whole construction of a criterion-referenced test hangs upon the demonstration of an adequate relationship between the criterion domain (job-related reading tasks) and the items making up the test (Dahl (1971); Rovinelli & Hambleton (1976)). It is essential, therefore, that the test has content validity and this is discussed in Chapter 9. Further, the relationship of test scores to the criterion domain, predictive validity, is also essential to ensure a useful product and this forms the substance of Chapters 13 to 15.

In conclusion, one must point out that the development of criterion-referenced tests has once again been almost exclusively an American affair, and as such is a product of the needs and solutions of that country's educational system. The wholesale testing of pupils for graduation minimal competency is inextricably linked with the design of dichotomous, state model, mastery tests in all but a few cases. The emphasis of inquiry has been upon the ways for making as few incorrect decisions as possible and the design of test instruments of controlled content, objectives and desired outcomes.

Much of the American work contrasts strongly with the needs and purposes of SOFRP, whilst still containing the germs of useful ideas. The content of any test produced is controlled, not by curriculum assessment needs or the prescriptive commands of a committee, but by the nature of discovered reading tasks. It is unlikely that any useful purpose can be served by a rigidly imposed dichotomy of 'acceptable-unacceptable'. It may, however, be the case that certain scores or ranges of scores can be empirically identified with levels of acceptable performance by job incumbents or with other indicators of criterion behaviour. As such a consideration is not linked with gaining a graduation diploma but with a teacher looking more or less closely at the reading skills of the testee, there is less emphasis needed on correct or incorrect decisions. It is surely better for the weight to be on the side of giving more help than is strictly necessary. Hence, scores used in a 'cut-off' sense need to be aimed higher than might be suggested by the empirical identification.

Concluding Remarks to the Introduction

The Sheffield Occupational Functional Reading Project was established to develop a criterion-referenced test to assess levels of occupational functional reading ability amongst pupils, leaving at 16+ years of age, in their last year in Sheffield schools. The Project grew out of a complex interaction of the increasingly vocal demands of employers in both industry and commerce for an improvement in the level of basic skills possessed by school-leavers, the changing nature of education itself towards a more explicitly outwardly oriented system being seen to be accountable for its practices, the economic pressures of a recession, and the growth of new technological methods. It is certain, also, that local needs played a part in the actual location and fruition of the Project.

The task to be undertaken by the SOFRP was to seek out those reading tasks encountered by the young people involved in the first six weeks of their employment, and then to construct a test to measure how well school pupils could perform these tasks. In considering a suitable definition of the term 'occupational functional reading ability', an operational definition was adopted which allowed a clear definition of the domain of tasks upon which testing was to take place. From the very large number of studies undertaken in adjacent areas of assessment, certain methods could be identified which would be of value in pursuing the research.

It is to these that this work now turns. Ensuing parts are concerned, firstly with the procedures for constructing a test, secondly with evaluating it and thirdly in considering some of the outcomes of its use. Detailed descriptions are given of each step for, in an area both politically fraught and technically complex, the clearer one's explanation, the better.

CHAPTER 5

SAMPLING METHOD

Introduction

It has been indicated, above, that the reading tasks faced by school-leavers in their initial employment must be identified, as a necessary precursor to any test development. Further, it is the aim of the Project, insofar as it is possible, to formulate the reading test or tests from actual copies of job-related reading materials. It follows that samples of such materials must be collected and that it is necessary to initiate fieldwork to obtain such reading materials and information about their use by school-leavers (i.e. the reading tasks and their context at work).

The fieldwork might proceed in at least two ways: by locating specific school-leavers and investigating their situation, or by locating specific employments and investigating the situation(s) of the school-leaver(s) within them. The former method would suggest that a random sample of all the school-leavers in one year be taken, waiting until they had found employment, contacting their employers and then visiting them *in situ* to identify their reading tasks and to collect materials. Whilst this seems a straightforward proposition, it has a number of overwhelming disadvantages. The sample would be taken from the Sheffield Metropolitan District, where the Project is located and by whose local Council it is funded. In the Metropolitan District, there are a large number of commercial enterprises and types of industry, and within each of these, a range of jobs. To ensure adequate representation across the area, a large sample of school-leavers, say 500+, would need to be followed up, during their initial period at work. This is simply beyond the resources of the Project. Further, there is little guarantee that either employer or school-leaver would agree to the procedure which would be disruptive of the latter's first few weeks at work. It might, in fact, jeopardise the young person's continued employment prospects. Also, it is within the limits of sampling theory to suggest that, quite by change, some major employing industry might be missed out, or substantially under-represented, across the industry or in one particular type of job.

To counter these disadvantages, one may approach the question from "the other end", as it were, by locating the employments into which school-leavers typically go and sampling within each one. Sample size can be substantially reduced as one would usually be sampling firms and not individuals, and the number of firms taking on school-leavers is likely to be less than the number of leavers. It is not necessary, either, for a firm to have taken on a leaver in the year of the investigation, merely to employ school-leavers, in general. The investigation can be conducted with minimal disruption, with most information collected from the firm, rather than the individual leaver. Of course, some supporting information would be needed, but immediate supervisors or personnel officers are at least as likely to know the requirements of each job, if not more so. Firms are increasingly moving towards written job-definitions and analyses. The other major advantage of using the employer as the sampling unit, rather than employee, is that a firm may well take on school-leavers into more than one job, and the investigation would thus capture many birds with one interview.

It was an integral part of SOFRP, therefore, that job-related reading tasks would be identified, and materials collected, by the use of a random sample of employers in the Sheffield Metropolitan District, supplemented by data collection from recent school-leavers, and others directly involved in job-seeking and initial employment and training.

Sample Size

The time available for fieldwork was February to September, 1978 about eight months in all. This time limitation was indicated by the desire to start piloting test items amongst 5th form pupils from the start of the academic year 1978/9. It was estimated that the maximum number of visits to firms and other bodies would be 150. Each of these visits represented at least an interview within the firm and possibly tours of the work-place, discussions with other persons, etc. The number of firms would have to be less than 150, in order to visit such bodies as Industry Training Boards, Group Training Association, Careers Service and Trade Unions, who are all involved in the job-seeking process and the initial period of employment. The maximum number of firms to be visited was, therefore, estimated at 120.

Industries and Jobs

Throughout this work, categorisation of firms will be by the British Standard Industrial Classification (BSIC), according to the major headings and subcategories generally used by the Sheffield Careers Service in their Annual Reports. An extract from the SIC is given as Appendix II. Categorisation of jobs will be by set of categories defined below, also used by the Sheffield Careers Service.

Apprenticeship:	in which articles of apprenticeship are signed and agreed, with national regulations, block and/or day release training;
Professional:	recognised training in one of the professions, e.g. articled clerk in accountancy;
Clerical:	general office and clerical duties which may or may not include any training;
Operative Training:	specific, on-the-job training for a minimum period of two months.
Others:	no formal training longer than 2 months, or none at all; non-clerical.

Figure 5.1: Definition of job-types

These categorisations have the advantage of according with national classifications and allow potential test users to evaluate the work for their own purposes. Any test designed for one set of jobs would clearly be inappropriate for a different set.

Statistics showing the occupational destinations of school-leavers during the period 1st October 1973 to 30th September 1977 (i.e. the period from the raising of the school-leaving age to the start of SOFRP), were kindly provided by Sheffield Careers Service. These are reproduced in Tables 5.1 to 5.8, and have been used to prepare Tables 5.9 to 5.16, showing trends within the period given and the distribution of employment between and within categories. These tables include 17 and 18+ year old leavers as well, but the proportion involved is sufficiently small as to make little difference. (It is to be noted that, within the firms eventually interviewed as little as 1% of all school-leaver employees were 17 and older.)

Certain categories appeared to grow or decline in proportion of leavers employed very rapidly year by year, but this was due to the low numbers involved. For the larger categories, however, certain trends were clear. The metal-working and engineering industries combined were gaining a larger proportion of leavers over time, whilst certain manufacturing industries took fewer leavers every year. The distributive trades continued to take the largest number in any one category. There was some change within groups, however, "Engineering" and "Engineer's Small Tools" had taken on fewer and fewer leavers, whilst "Other Metal Manufacture" had increased greatly. "Electrical Goods" had shown a steady, though small increase.

There were obvious and wide differences in the opportunities available to boys and girls. These differences made no real difference to the sample, however, particularly with the rising proportions of Operative Training and Others, into which the sexes go more or less equally. It was clear that, over time, the number of apprenticeships available had fallen, as had the proportion of leavers able to obtain one (N=2074 in 1973/4 and 1298 in 1976/7; 36% in 1973/4 and 23.6% in 1976/7). Clerical positions had also declined in much the same way. Professional posts stayed fairly constant at about 3.7% of all jobs, and it was the two remaining categories which gained in proportion. With the increased level of youth unemployment, however, actual numbers in these two job-types had not risen quite as high as the proportions suggest.

<u>Limitation of Categories</u>

The Sheffield Careers Service data used nearly thirty industrial or commercial categories. Some of the categories, however, employed very few school-leavers in any year and it seemed justifiable to attempt to eliminate low-employing categories. This would allow greater time to be given to the higher-employing categories.

Categories were eliminated which contained less than 1% each of school-leavers in 1976/7 as:

(a) this excluded categories which had previously only accounted for 4.71% of all leavers in 1976/7;

(b) the number of categories was reduced to nineteen, which would save time and effort reasonably employed elsewhere;

(c) no excluded category represented more than 1.5% of any job-type (Table 5.16) and, therefore, specific job-related reading materials from these categories would be of limited value. Admittedly, this is true of other categories, but their overall numbers overule this consideration.

The categories thus excluded were:-

		N	%
(i)	Agriculture	32	0.58
(ii)	Mining	41	0.74
(iii)	Chemical and Allied	9	0.16
(iv)	Electrical Goods	54	0.98
(v)	Vehicles	8	0.15
(vi)	Textiles	14	0.25
(vii)	Bricks etc.	27	0.49
(viii)	Other Manufacturing	28	0.51
(ix)	Gas, Electricity and Water	47	0.85
	Total	260	4.71

A further problem is that of heterogeneity of industries within a category. With a limited number of firms to be sampled from each category, it may well be that not all industries in a highly heterogenous category would be sampled and, conversely, that low employing industries within a high-employing but heterogeneous category will be sampled at the expense of the more important. Whilst for "Timber and Furniture" and "Paper, Printing and Publishing", this problem did not occur because their contribution to the sample (three firms each, see below) allowed adequate coverage, this was not the case for "Miscellaneous Services". Two sections of this category ("Hairdressing" and "Motor Repairs") had already been singled out for separate reporting by the Careers Service. Only the remaining significant grouping, "Hotels, Restaurants and Caterers", was taken as a separate category, and the other headings were eliminated from consideration. (It will be noted that several of these eliminated headings are only able to employ those over the age of eighteen, due to the licensing laws.)

The category "Professional and Scientific Services" suffered from a similar problem. The major part of heading 872, "Educational Services", is, in fact, a part of Sheffield Metropolitan District

Council, and therefore should be considered under "Public Administration" (see below). Further, it was decided not to include "Medical and Dental Services", as the complexity of the Health Service's organisation would be an unwarranted drain on the project's resources in relation to its level of employing school-leavers. The remaining categories were mainly professional, required fairly high entrance requirements and, collectively, accounted for very few school-leavers. The entire category was eliminated, except "Educational Services", which was dealt with as indicated.

"Public Administration" needed special consideration. It is, again, a fairly heterogenous category. It was felt that access to brances of the Civil Service and to the Armed Forces would be difficult and the effort not commensurate with the rewards. Also, in Local Government Service, the Police and the Fire Services recruit very few school-leavers at 16+ years of age. South Yorkshire County Council is located at Barnsley and so not of immediate importance. to Sheffield leavers. Therefore, it was proposed that, in its study of Public Administration, the Project confined itself to Sheffield Metropolitan District Council, its collaborating institution; the nature and scope of that study to be determined in conjunction with the Council.

Size of Firm

It was felt that sampling within each category might well result in the 'swamping' of the sample with small firms. This was considered undesirable as it was thought that the reading requirements of jobs in larger firms might be greater than smaller firms, where there is greater oral communication. Further, small firms, per se, were less likely to take on school-leavers in any significant numbers and stratification into 'Large', 'Medium' and 'Small' firms was needed. (Of course these terms need to be defined, but this will be done below.). It was suggested that firms employing less than a certain number of employees (all employees) should be eliminated from consideration in the sample. In the study "Numeracy and School

Leavers" (Knox 1977), in taking an 8% probability sample of the Sheffield area, "Firms of size less than five employees were not used, as they were very numerous and would be highly unlikely to take on school leavers" (p.2). It was proposed to use a similar lower limit, subject to upward revision as necessary.

The definitions of 'Large', 'Medium' and 'Small' in relation to size of firm presented a problem: heavy industrial firms may be extremely large (e.g. British Steel Corporation) whereas firms in other categories such as Distribution may never achieve such a size. Stratification by size may thus lead to the firms in a category clustering in one stratum or appearing in only two of the three. It was necessary, therefore, to use different definitions for different categories.

Training Bodies

The role of Industry Training Boards (ITB's) and Group Training Associations (GTA's) was a key factor in the sampling method. ITB's all exist to provide standards within their industry, but many also provide basic training, as do the GTA's. Such courses may be on a full-time, sandwich block-release or a day-release basis, and are usually limited to apprentices and operative trainees. It was therefore useful to take those categories with ITB/GTA's and consider them separately.

From the sample of firms in categories with ITB/GTA's, firms were divided into those that use the ITB/GTA training schemes (or related schemes) and those that do their own training. Of the former group, information was also gained directly from the ITB/GTA, whilst job-types for which they provided no schemes will require speárate investigation. In short, there were four routes to the gaining of reading materials for test design: via ITB/GTA's; via firms using these schemes; via other firms in the same categories; via firms in other categories.

Fourteen categories were covered by ITB's and/or GTA's:

(i) Clothing
(ii) Construction
(iii) Distribution
(iv) Engineering
(v) Metal Manufacture (B.S.I.C. 311 to 323 inclusive)
(vi) Other Metal Manufacture (B.S.I.C. 393 to 399 inclusive)
(vii) Cutlery
(viii) Engineer's Small Tools
(ix) Hand Tools
(x) Transport
(xi) Paper, Printing and Publishing
(xii) Timber and Furniture
(xiii) Motor Trades
(xiv) Hotels, Restaurants and Catering

Five are not:

(i) Food, Drink and Tobacco
(ii) Communications
(iii) Insurance, Banking and Finance
(iv) Hairdressing
(v) Sheffield Metropolitan District Council

(It was further necessary to divide the "Transport and Communications" category into its two components, as some aspects of "Transport" are covered by an ITB.)

The Sample

With the above discussion in mind, a proportional, random sample of firms was taken stratified by S.I.C. category and by size. Data on which this sample was based was kindly supplied by the Employment Services Agency.

The proportion of the sample allotted to each category was determined by the proportion (of the total) of school-leavers entering that category in one year. The data from Sheffield Careers Service was used to arrive at the number of firms to be sampled from each category and these are given in Table 5.17 (Figures are, in some cases, approximate as the reorganisation of categories, discussed above, did not allow exact comparison with the data. The names of firms supplied by the Employment Services Agency were not available for firms with less than 30 employees. This figure of 30 was accepted as the lower limit of the 'Small' category, after consultation with the Sheffield Careers Service. The following size categories were used:-

Small:	30 to 100 employees
Medium:	101 to 500 employees
Medium:	101 to 1000 employees (Metal Manufacture only)
Large:	501+ employees
Large:	1001+ employees (Metal Manufacture only)

It was also decided that first weighting be given to the larger size categories, e.g. if the number of firms to be sampled was 10 for a given industry, then four would be large and three would be Medium and Small. Every size category was to have a 'representation' however, unless there were no members (e.g. Wholesale Food and Drink, small).

A total of 117 firms and Sheffield Metropolitan District Council comprised this part of the sample. To this were added ten ITB's and GTA's, representing all of the fourteen industries given in the list above, each of which were to be contacted for more general information about their industry. It was also proposed to contact relevant trade unions and associations to discover what, if any, materials they provided for school-leavers. This, of course, waited upon the contacts with the firms, to see which were the relevant unions.

It was not, at first, felt that continuing, full-time further or higher education fell within the SOFRP remit, and colleges of further or higher education were not to be considered. During the course of the fieldwork, however, it became increasingly clear that, for some jobs in some industries, day-release or block-release to college formed the substantial basis for job-related reading. A full description of the sample of organisations contacted must include the four colleges of further education in Sheffield.

The contribution of each aspect of the sample to the collection of reading materials is shown in Figure 5.2, the Sampling Model.

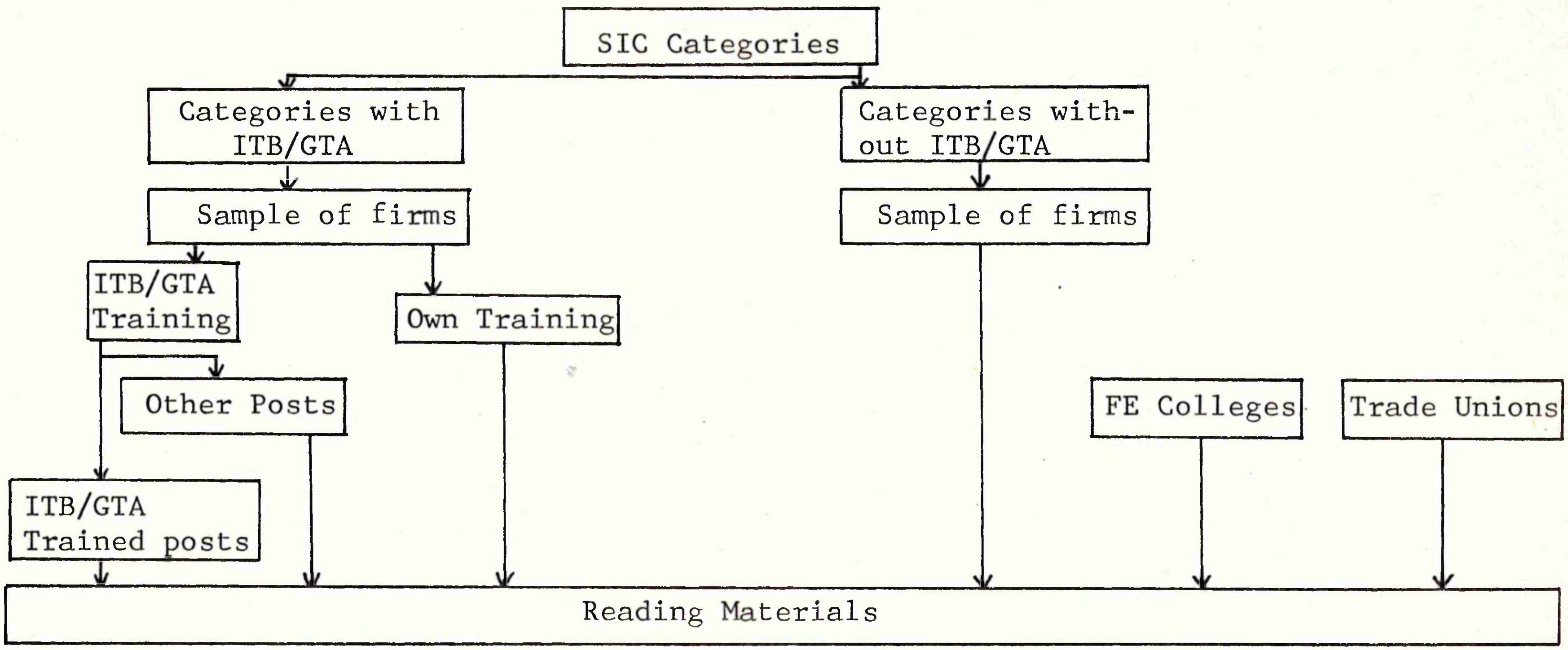


Figure 5.2 Sampling Model

CHAPTER 6

INTERVIEWS

Introduction

Having arrived at a set of organisations to contact, it is necessary to consider the specific information that is required and to select the most appropriate method for obtaining that information. These two factors constitute the prime determinants of the format of the data collection fieldwork. It would be without value to go to each firm or organisation and request a very general set of information, without defining for the firm the terms of the enquiry and dictating the limits of response. Otherwise, one might then receive every written sheet given to a school-leaver from his first interview to the end of his first year with no information about the context of its use, or one might receive a company handbook with oral details of its use, and nothing else. It cannot be expected that employers will remember, without prompting, the details of minor jobs, nor that the responses they may make will fit into neat categories.

Here, we are concerned with the identification of job-related reading tasks, the collection of relevant materials and information about their uses in the initial period (six to eight weeks) of employment. It is therefore straightforward to define, at least, an aspect of the fieldwork as being to ask: "What reading materials does a school-leaver encounter in the first six to eight weeks with you?". This, however, misses out a step, because there is no reason to expect uniformity or even progressive development within that period. Better, perhaps, to ask "What sorts of things does a school-leaver do in the first six to eight weeks? Which involve reading and what are the specific sorts of circumstances of that reading?" Hence, one receives a history of the leaver's initial employment, prompts responses along the way and delineates the limits of the responses as sufficiently broad but not admitting digression.

There is more, however, to a school-leaver's position than its history and reading requirements. Positions are filled by young people on the basis of personal and academic qualities and this information will be relevant in comparing different jobs and different industries. Moreover, some firms will have training schemes to be considered, others may have had problems with reading in the past, and such information may also be useful.

It must be said that the author's experience of the shop-floor was limited and more weighted to service work than manufacturing and commercial enterprise. The opportunity presented by informal contacts to visit a tool-making firm was, therefore, very welcome. This confirmed what had been previously felt, that the investigation of job-related reading tasks best proceeded by semi-structured interviewing, rather than any other method. Formal questionnaire methods were ruled out as too inflexible for the range of industries and jobs and requiring either a high level of prior knowledge or time for extensive pilot work - neither of which were available. A personal interview is much more likely to discover the details of individual situations than any questionnaire. Some structure was felt necessary, however, to ensure some comparability between and within industries and jobs, and some 'face-sheet' data was though valuable. Further, personal presence - rather than, say, a postal approach - allows requests for specimens of materials to be made with a much greater expectation of success.

This semi-structured interviewing was used with all three categories of informant: the firms in the sample; other organisations and recent school-leavers. The previous chapter mentions the need for supportive evidence to back up the statements of employers and this was to be obtained via these latter interviews.

Industry Training Boards and Group Training Associations

Prior to contacting any firms in their employment category, the relevant ITB or GTA was contacted and a representative interviewed. There were two main areas forming the basis of questioning: whether the firms sampled were representative of their industry in Sheffield, and the nature of typical school-leaver employment in their industry.

In eight cases, a firm was deleted from the sample for being a wholly-owned subsidiary of another firm in the sample, bankruptcy or changing business to another category, not mentioned in the original data from the Employment Services Agency.

Details were obtained from each representative about the sort of jobs into which a school-leaver might go, the likelihood of training and its organisation, ranges of qualifications required and an estimation of the amount of reading in the first few weeks. If the organisation ran its own 'off-the-job' training scheme to which firms in the sample sent employees, details were noted and an interview with the relevant personnel conducted at a later date.

These interviews proved very useful, not only in the provision of overviews of the various industries and in the materials collected, but in giving background which enabled the author to deal more knowledgeably with employers. At no other time than these interviews, were persons not working on the SOFRP allowed to learn the names of firms in the sample. During these interviews, only the names of firms in the relevant industry were made known to the interviewee.

Interviews conducted with Firms

Firms were contacted one B.S.I.C. category at a time, taking the largest employers first. This was to ensure that, if circumstances should alter, or the fieldwork take longer than expected, a very large majority of the data would have been collected.

The Sheffield Careers Service was again most helpful in allowing names of contacts to be extracted from their files. These were usually Personnel or Training Officers. Firms were contacted by letter (see Appendix III, Letter 3) initially, and then by telephone to arrange an interview, where possible.

Initially, four areas of co-operation were requested:

(i) an interview between the firm's representative and the Researcher;

(ii) the inspection of reading materials actually used on the job by school leavers;

(iii) discussion with some of the firm's recent school-leaver employees about the situations on the job in which reading was required;

(iv) observation of some of the leavers at work in situations which require reading.

It was found, however, that the latter two areas were inappropriate at that time of year (February to September), as most school leavers are taken on in the August-September period and, hence, the material being used in the early stages of fieldwork was not that of the initial six weeks of employment. It was decided, therefore, that a second stage of interviewing of a sub-sample of school leavers, later in the year, would fulfil the requirements of the discussion and observation stages.

An interview schedule was used by the author to guide the interview and to gain statistical data. This schedule is given as Appendix III , Schedule 1. All notes were handwritten. Each interview proceeded with the author giving background details to SOFRP, then requesting information about the firm, the number of employees, training procedures etc., as shown on the Schedule. The first three firms interviewed, with whom informal links had already been established, were used as pilot firms. They showed up certain areas not appearing on the early version of the Schedule and these were inserted. (These areas were probed in the pilot interviews but the prompting word did not appear on the Schedule.)

Following on from the statistical and other data, the firm's representative was taken through each school-leaver job and asked to give details of training and work in the first six weeks, including all procedures for inducting the young person into the firm, and to describe the circumstances in which reading was required. Copies of all reading materials were requested for inspection, and, where possible, specimens were collected. All firms were assured that their responses and materials would remain confidential and that specific requests would be made for the right to reproduce materials for any purpose other than internal uses within the Polytechnic. Information was also requested as to the role of any supervisor or trainer in the school-leaver's reading tasks (e.g. whether the material was also covered orally, or the trainer reinterprets the materials in such a way that reading them is never required).

Whenever time was available, the author also asked for a short tour of the work-place, in order to examine posters, notices, etc. which might have been overlooked by the firm's representative, and, again where possible, to speak briefly with supervisors about any reading problems or materials particularly needing oral reinterpretation.

The analysis of response rates of the firms is shown in Table 6.1 (see Appendix I) and a further breakdown into B.S.I.C. category and size of firm is given in Table 6.2.

The overall response rate of 70.6% is seen as very encouraging, indicating that the fieldwork had suceeded in collecting materials and information from a great range of jobs and industries. Although the low response rates for small firms in less cheering, it must be pointed out that all thirteen companies not taking on school-leavers were categorised as 'Small', in line with the assumption made in the previous chapter that smaller firms are much less likely to take on school-leavers. Collectively, all the firms visited employed more than 52,000 persons, including over 1,200 school-leavers (i.e. who left school in the calendar year prior to the interview). This represents a sample of 22.4% of all school-leavers finding work in 1976/7, far beyond the size of sample which might have been covered by using the school-leaver as the sampling unit.

The responses left no B.S.I.C. category without at least one interviewed company, though not always one of each size. Table 6.3 shows the number of each type of job into which a school-leaver went in the firms interviewed. As it can be seen, only a tiny proportion entered jobs categorised as 'Professional', and this category was therefore deleted from further consideration. The interviews showed that school-leavers entering this line of work are post-A-level leavers, not 16+ years of age. There also appeared to be a shortfall in the numbers of 'Others' employed by the sample firms. This was because it was found to be very difficult to separate them from 'Operatives', due to mixtures of definitions used by firms. It was, therefore, proposed to combine the two categories into 'Operative/Others'. Further, the Distribution industry appeared to be so very different in reading requirements from similar level jobs in other categories, that it was decided to split further the job-categories into the following:

(i) Apprentices

(ii) Clerical workers

(iii) Distribution Operative/Other

(iv) Operative/Others: non-distribution

Interviews conducted with School-leavers

In September, 1978, seventy-five recent school-leavers were interviewed at their place of work. The leavers were chosen by type of job rather than firm, and seven firms plus an ITB were asked to assist. This selection was judgemental, and the firms selected were from respondents to the previous round of interviewing. As the leavers were providing corroborative, rather than substantive, evidence, the arbitrary nature of their selection was not seen as invalidating their contributions. In fact, it had the advantage of allowing the author to deal in known quantities when comparing their information with that of the companies by which they were employed.

An analysis of the specific types of employment of these leavers is given in Table 6.4. Each interview was conducted on an informal basis, lasting about fifteen minutes. The leavers were asked about what reading they had to do, using which materials, in what context and with what degree of understanding. The relevant interview schedule is given as Appendix III, Schedule 2.

It was quite clear from the responses of the school-leavers that they were using the materials described by the employers, and in the manner given, except that there was a tendency merely to skim and discard materials provided for information. This was particularly true of handouts to do with conditions of service and other legalistic documents. The more important documents, such as Health and Safety Rules, Company manuals, did tend to be read or, at least, referred to, but when circumstances dictated it, rather than when received, as the Company tended to believe. In all, however, it can be said that these interviews confirmed the earlier, employer interviews.

Interviews with members of Colleges of Further Education

As was mentioned in the previous chapter, it had not originally been planned to collect materials from Colleges of Further Education.

In fact, many types of employment, particularly in manual work, were geared to attendance at a local College of Further Education. It was established quite early in the fieldwork that for certain jobs, there was very limited, or no contact with the employer until a substantial period of off-the-job training at a college had taken place. Further, for other jobs, attendance on a day-release course provided the major reading requirements of the initial period of employment. It was decided that these courses could not be neglected, or the research would seriously under-estimate the reading requirements of a large range of employments. To this end, all Further Education Colleges within the District Council boundaries were contacted, in a similar manner to the firms, and asked to allow investigation of the reading requirements of the initial sessions of certain courses (ascertained from the fieldwork).

Response was, in general, favourable and all relevant materials and information collected. Unfortunately, one college declined to assist. This meant that no reading materials were collected for apprentices in the Construction industry, as all initial job-related reading tasks occured on block-release. Sadly, this industry had to be deleted from further consideration, except for its contribution to clerical workers.

Details of the courses investigated are given in Figure 6.1.

Interview for Job-seeking

Part of the SOFRP remit is to consider the reading requirements of job-seeking. This task was considerably simplified by the fact that all school-leavers are normally obliged to conduct their employment seeking under the aegis of the Careers Service. A single interview sufficed to obtain the pattern of passage through this system, of an unemployed school leaver, and suitable reading materials were collected.

Sheffield Metropolitan District Council

Discussion took place with the Senior Advisor (Research and Evaluation) of the Education Department of the Council. It was decided that the complexity of the Council, with its large organisation and many functions, justified separate study and it was not included in the fieldwork. An analysis of the employment trends in the various departments of the Council are given in Table 6.5.

Trade Unions and Associations

It emerged from the fieldwork that there was considerable variability in the practices of trade unions with regard to school leavers. Some unions sought out members, others obliged them to join, others allowed them to join if they desired, some did not recruit at this age. This being the case, it was decided to approach all unions by post, and ask for materials, if they do provide them for school-leavers. This was done via Letter 6, Appendix III. Although materials were received from some unions, low response rate meant that they have not yet been the basis of item construction.

Non-Sample Data Collection

A number of respondents from the Retail sections of the Distribution Industry indicated that a large number of their sales operatives were required to familiarise themselves with merchandise from wholesalers' or manufacturers' information. They were usually unwilling to release such material as it was in constant use. A number of manufacturers and wholesalers, at both national and local level, were contacted and asked for specimens of materials that they typically provide to retailers.

Other data was collected from the local library, where technical manuals used by motor mechanics were available. Again, the firms involved were not willing to release materials in constant use.

Discussion

Whilst it has only limited bearing on the development of a functional reading test, it is interesting to note some of the different sorts of information arising from the various interviews. In particular, from the employers interviews a great deal of information was gained concerning their selection techniques, the range of academic qualifications required, the different training procedures and, of course, their encounters with reading problems amongst school-leavers. On the school-leaver side of things, a brief consideration of the young person in situ may serve to bring greater understanding of some of the later Project decisions. An appreciation of what sort of employments send school-leavers for further education may also prove of value for the reader.

Perhaps one of the most striking things to come out of the employers interviews was the extent to which some form of industrial or commercial personnel selection test was used to screen applicants. This practice was particularly prevalent in the manufacturing and engineering industries, but some posts in the insurance and banking fields were also increasingly subject to such screening. The reasons given for the testing were generally similar. As public examination results were not available until the August after the young people have left school, they could be of no assistance in selecting from applicants. Coupled with the high number of applicants for a limited number of posts, and the fairly fierce competition to get the best of the leavers by the various firms involved, selection tests were used to cut down the amount of interviewing by eliminating those not conforming to minimum performance levels on the tests. In this sense, the firms were using mastery tests. A wide range of tests were used, of varying levels of quality, including general knowledge quizzes, clerical accuracy tests, mechanical reasoning, general intelligence, english and mathematics tests. In fact, a number of local firms had combined to administer a common test to applicants, to avoid unnecessary duplication of testing within the city. A number of tests had been professionally developed for their purpose, but the majority were either tests of some general quality (e.g. intelligence) being applied in these particular circumstances, or were 'homemade' tests reflecting some facets of the employment or of general knowledge thought relevant by the firm. Manuals of administration and details of relevant test statistics were not available in the majority of cases. The testing of english or general reading ability was seen as a problem by most firms, who, whilst fairly happy in their own assessment of what arithmetic skills were required for the various jobs, were far less happy about their ability to assess reading or language fluency. It was necessary on a number of occasions for the author to reassert that the proposed test (or tests) under development was primarily intended for use in schools prior to leaving, rather than as an aid in job selection. Tests were used largely for apprentices and operatives, rather than other groups.

The use of tests, however, was, as previously indicated, an initial screening process and much more weight was placed on non-academic criteria when it came to interviewing. This finding is very similar to one of the selection strategies used by employers discussed by Ashton (1979). In particular, attitude and personality counted strongly: a young person with an interested and positive attitude to the job and with a lively disposition was rated as much more likely to be taken on than an uncommunicative, listless school-leaver. This clearly has many implications for careers teachers when advising school-leavers on how to conduct themselves at interviews. All employers regarded the interview as their most important method of selection. In certain industries, a practical test was used, but this was less common.

Linked to selection techniques is the matter of academic requirements. The majority of companies visited had no set level of requirements for any job and were often vague as to what some of the qualifications meant, particularly to do with grades and modes of the C.S.E. and its comparability with G.C.E. O-levels. As previously indicated, many firms found the timing of the issuing of results a drawback in selecting applicants. Such firms tended to make unconditional job offers, therefore, rather than dependent on the achievement of certain grades or subjects. This had obvious benefits for both employer and school-leaver in terms of guaranteeing a work-force for the former and a job for the latter. For technician levels, however, a minimum of four O-levels (or equivalents) was required, due to the minimum entry level set for the relevant further education courses implicit in technician training. Other firms, and generally not those using selection tests, imposed a set minimum requirement (again, usually four O-levels) for entry. Typically, these were larger department stores and large insurance companies and banks. It is to be noted that entry into such companies was usually delayed until results were available, rather than soon after leaving school as in other industries.

Training organisation varied greatly between jobs and industries, although actual practices tended to be common across industries. All apprentices had full-time, on-the-job training at the company or special training centre, interspersed with block or day release to a further education college, where they undertook specific courses. For other categories of job, the picture is less clear. The majority of large companies in all industries were committed to training their young staff, either by set procedures at work or in concert with day-release courses. Exceptions were from industries such as the Distribution trades, where such courses were still in their infancy. The smaller the company, the less commitment to training there was. A common attitude was that, for 16 to 18 year olds, if a course could be shown to be relevant and beneficial to the company in the short term, then paid release was available. Otherwise, evening classes were recommended. In a number of cases, financial rather than temporal help was available.

Where company training, rather than day-release, took place, it tended to be well-organised, and a large proportion of firms visited had obtained exemption from levy by their Industry Training Board for the excellence of their methods. Even when this was not the case, firms were often proud of their systems and keen to provide information.

The incidence of reported reading problems amongst school-leaver employees was negligable (two cases in over 1,200 jobs). Most firms attributed their success in avoiding this to their selection procedures. It was the applicants for jobs who were the subjects of such reports. In fact, few respondents failed to make some mention of poorly completed application forms, badly composed letters and poor command of spoken language. It was not unknown for an applicant to be asked to complete another form just before an interview, to ensure that he had completed it himself rather than a relative. Spelling was frequently a cause for complaint. Most complaints, however, were directed against the school-leavers' shortcomings in basic arithmetic, especially percentages and fractions.

School-leavers taking part in the interviews were, of course, fairly shy and reticent, but the use of specific questions early on in the interview, rather than expecting them to talk at length, seemed to overcome this. Almost all were very happy about their work and had encountered very few problems. One or two admitted to having had reading difficulties whilst at school but claimed to have no such problems at work. Most leavers were able to identify the materials provided by the firm as the ones they had received, and to select those they had had occasion to use or refer to (rather than receive and file away). At the operative level, the leavers tended to rely more upon asking fellow workmates about any work-related difficulties whilst other grades seemed happier about asking their supervisors. Materials on health and safety were rarely the subject of detailed study, and legal-type induction documents even less so. General induction documents, however, the ones giving details of rates of pay, hours of work etc., were seen as important, especially in reassuring themselves after oral explanation.

Much of the use of written materials was in repetitive tasks, particularly in the clerical field and so reading was a constant factor in working life. At higher level jobs, clerical staff did little but reading and writing and their swift introduction to computer-based work was noted. Apprentices had high reading loads also, especially in their initial periods of training. They reported an expectation, by their employers, of early self-sufficiency in finding their way to and about the relevant documents. Distribution operative/others fell into two categories: those for whom reading was an everyday activity and those for whom job-related reading requirements seemed to occur only in induction. Some sales assistants did report initial difficulties in finding their way around lists and forms.

School-leaver employees who were undertaking some job-related further education tended to report that the reading load was much higher in these courses than at work. None reported any problems, however.

Conclusion

Semi-structured interview techniques were used to obtain information about job-related reading tasks, the materials which were used and the context in which these occurred. This information was gained from training organisations, the sample of firms, recent school-leavers, the Careers Service, most Colleges of Further Education in the area and some local trade unions. Information was not collected from the Metropolitan District Council after consultation with one of the advisers to SOFRP.

The response-rates for the types of informants were favourable and a great deal of materials and data was collected. The materials can be classed as induction (e.g. Health and Safety Rules, handbooks, conditions of service, job description) and as job-functional (i.e. materials relevant to the actual performance of the job, rather than to being an employee, and include manuals, job-cards, stock-lists, all training materials, etc.).

Certain changes had to be made on the basis of the fieldwork: there were no materials available for apprentices in the Construction industry and this will affect the applicability of any test to that industry. Very few school-leavers were recruited at 16+ years into

'Professional' jobs, and that category was deleted. The materials for the Distributive trades were sufficiently unique to warrant a separate category. The difficulty involved in distinguishing between 'Operative' and 'Other' jobs meant that they were amalgamated to 'Operative/Other'.

CHAPTER 7

ANALYSIS AND CLASSIFICATION

Introduction

The collection of literally hundreds of passages, forming the basis of job-related reading tasks, would, in itself, be of interest, but of little practical value without some criteria for organising the material for interpretation and use. This is as true for these materials as for census data: one cannot make national policy decisions on the basis of cases, but on the basis of groups or categories. Similarly, one cannot design individual reading tests for each job in each firm (apart from the time involved, the restriction of target population would make validation impossible). One may use, however, the case as representative of a group or category. Our concern here is to establish the groups or categories which will be useful in the construction of the reading test, basing this upon the materials collected and the information about who uses them and in what circumstances.

Common Materials

Sheffield industry is dominated by the traditional metal manufacturing, tool making and cutlery trades. These deal in large machines, specialised equipment, hot and dirty conditions, and hazardous work. A new entrant to such a trade cannot, except for the most restricted tasks, go into the work situation untrained. He must not only learn in order to contribute to the complex tasks involved, but also to survive without causing danger to himself and to others. Recent legislation (in particular, the Health & Safety at Work Act, 1974) imposes the requirement upon the employer to inform or adequately train the entrant in the dangers of his trade. Just because the statistics are worse for heavy industry, however, does not make the light industry, shop or office any the less dangerous, and the legislation goes across the board.

Further, other legislation has imposed contractual obligations upon employers in terms of dismissal, redundancy, etc. Although this does not apply within the first six months of any job, most employers take on entrants with continuous employment in mind, and do provide written statements concerning the employment.

All this has been discussed to point out that there are no jobs whatsoever, except those where an employer is unsure about the continuous employment of an entrant or is prepared to undertake a large degree of oral discussion, in which there are no reading requirements. The provision of terms of employment documents, Health & Safety Rules, contracts, etc. mean that every school-leaver will be given something to read upon starting work. Also, because of the requirement to make reasonable effort to ensure that employees are aware of the content of safety rules, etc, the school-leaver is likely to be taken through the materials, or be asked to familiarise himself with them. The point to be made is that materials used to induct a new entrant into the company are likely to be common across jobs and industries in that they contain these basic elements, of employment terms and safety rules. Such common types of materials tend, also, to have common formats, laying down "do"'s and "don't"'s or setting out rates of pay, hours of work, holiday and sickness arrangements etc.

Industries and Jobs

Moving on from materials used across all jobs and industries, it was fairly clear from the fieldwork that materials used by a metal-working apprentice in one firm were likely to be very similar - if not identical - with materials used in another firm. This can also be established for other jobs within an industry: that the same job across firms tended to have similar reading requirements. Hence, one can be fairly confident when talking about clerical workers in insurance companies, operative/ others in the cutlery industry, shop assistants in the retail trade, etc., that they constitute a fairly uniform group in terms of reading requirements.

This leaves a very large number of categories, however: every job in every industry, and it would be desireable to cut this number down, if possible, to make any reading test developed less extensive whilst maintaining its representative nature. That is, if one could consider the reading requirements of all apprentice jobs as similar, then one could use a range of differing materials to represent apprentice tasks across industries.

To do this, one must consider two areas: technical vocabulary and the types of content of the reading passages. The former is necessary in that the more abstruse or esoteric the jargon of a job, the less comparable it is. The latter is necessary in that it offers the possibility of common classification of materials regardless of their derivation.

Technical Vocabulary

The materials collected from firms and other organisations were full of words or phrases with specific technical applications. For instance, words to describe a lathe or part of a steel rolling process, phrases of 'legalese' in the documents an insurance company uses are all terms of a specific technical nature. Also, quite common words, such as 'receipt' or 'register' take on quite different meanings in some trades.

Technical vocabulary will tend to increase the complexity of job-related reading materials and, hence, the difficulty of the associated reading task. Technical vocabulary will also tend to increase the specificity of an individual passage or text. An insurance clerk is not expected to know the intricate jargon of a construction apprentice, and vice versa. It was clear from the fieldwork, however, that the extent to which a school-leaver was expected to know any of the job-specific jargon was extremely limited. An apprentice motor mechanic would usually be expected to be able to say something about a car engine, but not necessarily to discuss its intricacies. A shop assistant might be expected to know that her till could also be called a cash register but not the meaning of 'dry goods' or 'provisions', and so on. Technical vocabulary was something to be learnt during and after induction and the *a priori* requirements in this area were few.

This being the case, it would be clearly invalid for a test of occupational functional reading ability to assess levels of 'jargon acquisition' relevant to any employment. Not only would the range of vocabulary items be enormous, but the testee would be facing a task not mirrored in real-life. If one could measure his potential for coping with jargon, this would be a different matter from testing his present knowledge. Such a task would no doubt form a fascinating study in linguistics, but this is not within the remit of SOFRP. Still, one cannot ignore technical vocabulary, because of its contribution to complexity mentioned above, but nor can it be taken out of its context. It was felt that job-related reading tasks were more to do with the structure of passages and the type of content of materials, rather than specific vocabulary items. A testee must be able to approach materials despite their jargon, and only ask for help when there is a lack of prior knowledge about a word or phrase. That is, test items are still possible based on type of content rather than on specific meanings.

Content Classification

Sticht et al (1972) developed a classification system for the content of job-related reading materials, most of which is relevant here. The classification is reproduced as Figure 7.1. Using this system, samples of reading passages from each job and industry were classified according to the majority of their content.

It was already clear from the fieldwork that clerical workers were presented with the greatest volume of reading to undertake, closely followed, of course, by apprentices. It is the nature of the former job to be involved with written materials for the vast majority of their work, although the number of different types is smaller. Apprentices in training have to deal with all types of content under the Sticht system.

1. Tables of content and indexes:

Content designating the location of information with a publication.

2. Standards and specifications:

Content setting forth specific rules or tolerances which task procedures or the completed product must conform.

3. Identification and physical description:

Content attempting to symbolically represent an object via an identifying code (stock #, nomenclature) and/or by itemizing its distinguishing physical attributes.

4. Procedural directions:

Content which presents a step-by-step description of how to carry out a specific job activity. Essential elements are equipment/materials/ingredients to be used, and how they are to be used, with presentation organized in a sequential step-wise fashion.

5. Procedural check points:

Content which presents a key word or highly summarized version of what should be done in carrying out a task rather than how it should be done. This content differs from the content classified under Procedural Directions in that it assumes the user knows how to carry out the steps once reminded that the step exists and/or reminded of the decision factors which determine whether the step is required.

6. Functional description:

Content which presents an operating (cause and effect, dependency relationships) description of some existing physical system or subsystem, or an existing administrative system or subsystem.

(By kind permission: Human Resources Research Organisation)

Figure 7.1 Definition of content-type categories

They must look things up (category 1), following instructions (4 & 5), learn how something works (6), read specifications for a piece of work (2), and read descriptions, etc. (3). In these terms, it was fairly clear that apprentices would be treated equally across industries, despite the different specific applications of the materials. For instance, a labelled diagram of a lathe is - in reading terms - much the same as a labelled diagram of an automobile engine, except for differing vocabulary. Also, a passage describing how steel is made is - in reading terms - the same as the operating principles of chemical works.

Similar statements can be made for the Distribution industry, which is being considered separately in this work (see Chapter 6 above). The reading tasks of operatives and others are common across the different trades within the industry. In the main, clerical workers face similar materials to one another across industries. At the 'upper' end, however (banking and insurance, in particular) they tend to more complexity and, whilst establishing clerical reading tasks as common, this must be borne in mind in constructing test items (i.e. some of the more complex materials should be used also).

Operative/others is also a more heterogeneous group when it comes to content. They largely only use induction materials, but those with training also may receive and use materials of other content-types. The general lack of other materials, however, leads one to consider them also as a uniform group.

The Sticht classification system proved its worth in enabling this type of consideration of reading tasks across job-types. The system was developed for use with job-related materials such as technical manuals, stock lists etc. and its applicability to the present study was very valuable. By considering the different types of content of the materials used by different jobs and industries, it became clear that the six job categories (apprentice, professional, clerical, distribution operative/other, operative/others and induction) were sufficiently homogeneous across industries to enable them to be used in test construction. That is, items could be constructed using materials from different industries, but it would be valid to suggest that they also represented reading tasks from other industries. That validity was ultimately assessed by panels of experts (see Chapter 9). Item construction proceeded upon the above argument.

CHAPTER 8

ITEM CONSTRUCTION AND ITEM ANALYSIS

Introduction

Popham & Husek (1969) suggest that item construction in criterion-referenced measurement is basically different from that in norm-referenced measurement, in that the item writer in the latter case is designing a test to give a range of scores, whilst in the former case the item writer seeks to reflect a given area of tasks and is unconcerned with variability. "Since the meaningfulness of a norm-referenced score is basically dependent on the relative position of the score in comparison with other scores, the more variability the better" (p. 3), whilst for criterion-referenced measurement, "variability is irrelevant. The meaning of the score is not dependent on comparison with other scores; it flows directly from the connection between the items and the criteria".

So the NRM item writer will seek to produce items that about half the testees get wrong (i.e. neither too hard, nor too easy) and will construct the 'distractor' answers to lure sufficient of the testees away from the correct answer to achieve that 50% level of correct response. For the CRM item writer, if his items reflect some valid aspect of the criterion skill or behaviour or area, then it is of little concern how many get it wrong or right, or how good the other possible answers are at luring testees to mark them. In fact, 'distractors' in CRM are not really 'distractors' at all. They perform three functions: firstly to provide a 'non-master' with somewhere to put his mark rather than omitting the item; secondly, to lessen the chance of getting the item correct by guesswork; and, thirdly, to provide information to the tester as to possible shortfalls in the knowledge or skill of the testee.

Clearly, then the groundrules for item construction and analysis are different for the two types of measurement. It is necessary to consider what are valid item types (and which are useful with SOFRP) in terms of their relationships to job-related reading tasks; what methods of item analysis might be used and their implications, if any, for item construction; and what methods are available for handling the data in order to assist in evaluation of items and in test construction.

Item Types

Common item types include multiple-choice, true-false, essay, short answer, completion, matching and cloze procedure (though whether this last can be called an item rather than a test type is a matter for discussion). All of these types have their advantages and disadvantages within given situations, and one may best select the types for use here by looking at the specific situation.

The remit of SOFRP called for a test consisting of a number of items which might be administered by a school teacher with reasonable ease and which might be marked objectively. That is, each item would have one and only one way of representing a correct answer. Further, the very nature of the Project itself suggested that items should, as closely as possible, reflect the types of reading task encountered by a school-leaver. Typically, a young person at work uses reading materials to find information, whether it be a specific item of knowledge, the steps in a procedure or the standards to which his task must conform; or he reads in actually performing some task, such as in filing a letter. The problem of technical vocabulary also must be raised. As discussed in the last chapter, a testee cannot, before starting work, be expected to have a wide vocabulary at his fingertips relating to his possible employment. Items must, therefore, be aimed at the overall structure of the reading and its type of content, rather than the content itself: that is, one may not test on questions of prior knowledge, but on what the passage itself says. A given passage may be untrue or erroneous, but the pupil must be able to deal with passages of the general type or category into which the false or erroneous passage falls.

These considerations make it clear that certain item types were inappropriate for the Project. Essay questions are obviously unlike job-related reading tasks and can only be objectively marked with difficulty. The cloze procedure, whilst it may measure comprehension, cannot be said to reflect the reading requirements of a school-leaver. It is one step away from those requirements: one may infer adequate performance from the measure of comprehension but other measures may give more direct indications of performance by specific questioning. Cloze procedures also run the risk of deleting technical vocabulary that the testee cannot be expected to know. Avoiding this would require a way of dividing text up into technical, less technical, etc. categories along the lines attempted by Davies & Vincent (1976) - an attempt abandoned for want of good enough criteria to make the division anything less than arbitrary (1978, personal communication).

Other item types also failed to meet this requirement for reflecting job-related reading tasks: matching and completion exercises. Short written answers to questions were rejected as they cannot usually be marked objectively without great effort in piloting, and that any writing calls for skills other than reading. It was thought from the start that an appropriate answer-mode would be making an unambiguous mark (a tick in a box, underlining etc.) rather than writing even a word or sentence.

Essentially, one is left with different forms of multiple-choice items. Although arrived at through elimination of others, these types have much to recommend them in general. The modern examination systems make much of various multiple-choice types and so pupils tend to be familiar with them. Further, job-related reading tasks often require a school-leaver to find the correct datum from amongst a choice of options. This type of task may be reflected by a straightforward multiple-choice item. Five-answer items were chosen for this, rather than true-false, as the probability of being correct by guesswork is substantially reduced. They are also easy for testers to mark. Other types were also necessary, of a less familiar kind. Often, the task faced by a school-leaver indicates that several data must be discovered or used from the reading text. It seemed appropriate, therefore, that a form of multiple-choice be used where a testee might, typically, be asked to select up to four responses from up to ten possibilities, using the passage to read as the criterion for selection. This type of task was seen as very important; in the work situation, failure to notice one of several safety rules may result in as bad an accident as not noticing any; for this reason, all correct response options had to be marked to get the item correct. One other type was used for specific types of task: ordering or filing exercises were constructed (e.g. a list of names and addresses which required ordering by placing '1' in the box next to the name ocurring earliest alphabetically, etc).

So, then, three specific types of item were constructed, to reflect as closely as possible the job-related reading tasks faced by school-leavers, insofar as this was possible within an objective, paper-and-pencil, group test for use in schools. These item types were

(i) 5-answer multiple choice
(ii) 'Action' items (partial)
(iii) 'Action' items (completion)

as described above. (The term 'action' was used as most of the initial items of this type were concerned with 'what would you do in X circumstances?' and the label has stuck.) Certain aspects of the types of item selected above are clearly arbitrary. One may ask "why four out of ten and not six out of thirteen?" or "why 5-answer multiple choice and not 4-answer?". This type of question has no real response. The selection was arbitrary, but based upon judgemental assessments of how much of the page should be covered with text - too many options may increase the visual complexity of the task - and what might take the 'sting' out of the questions. The passages to read are novel and, if one is also to introduce novel item-types, one has a responsibility to ensure the pupil has as much chance to answer the question as possible without extraneous distractions. One runs the risk, otherwise, of asking questions more difficult than the text.

Item Analysis

Even before constructing test items, it is necessary to know how they are to be analysed and what restrictions, if any, the form of analysis might impose on item construction. One must select from the techniques available those best suited to the task: although a number of classical test statistics are not appropriate for use in criterion-referenced measurement, some undoubtedly can contribute, and a number of other techniques can also be used or devised.

Classical procedures involve, in particular, the investigation of the difficulty and the discriminative ability of each item and each distractor, typically involving point - biserial correlations. "In practical test construction, the variability of test scores is increased by manipulating the difficulty levels and content of the test items " (Glaser, 1963). As suggested above, non-discriminating items are those that fail to contribute to variability in the total test scores of testees, and this is usually due to the item being too easy or too hard or to it being ambiguous in some way. The question of ambiguity will still persist in CRM - although testee responses to ambiguous questions might perform useful diagnostic functions - but the difficulty of an item is far less important. If the item reflects some important aspect of the criterion, then it is a 'good' item, regardless of its ease or difficulty. For practical purposes, however, the test constructor will be less than happy if testees always get all of his items correct. It removes the point of testing in the first place: he must either shift his criterion or give up testing.

What, however, if an item is correctly answered more often by those with lower total scores than by those with higher total scores, i.e. a negative discriminator? Clearly, this item will be unacceptable in CRM also: in effect it is contributing to the probability of false positives and to false negatives - some of those not up to standard are 'passing' and some of those up to standard are 'failing' more often, due to this item and others like it. All tests, then, need to eliminate negative discriminators.(see e.g. Popham & Husek (1968) p. 6;

Wedman (1974) p. 113; Smith (1974) p. 144).

Despite problems in the calculation of correlational values as associated with low variability (see Chapter 4 above, or Lord & Novick, (1968) p. 129), discrimination indices to be used in the sense above, may still be calculated in CRM. Low variability will tend to decrease the value, toward zero, but it will not change the sign, and it is that in which one is most interested. Of course, ambiguity can also be indicated by non-discrimination and one must combine a measure of score variability with discrimination index when considering this aspect of item analysis.

The question of the item 'distractors' must also be raised. As was suggested above, the role of other answer possibilities in CRM is not to lure away a certain proportion of testees each from the correct answer possibility but to provide somewhere for the 'non-master' to choose, to lessen guessing and to provide diagnostic information. Hence, an underused possibility is not a poor possibility, merely an aspect of quite an easy question. The selection of one answer possibility rather more often than the others and especially, more often than the correct one, may be an indicator of ambiguity - particularly in a non-discriminating item. So, whilst the calculation of discrimination indices is not called for for each answer possibility, an indication of the numbers choosing each may be valuable. Further, action-type items present very large problems in terms of answer possibilities, as each has 2^n possible combinations, where n is the number of answer boxes - the calculation of discrimination indices for each - for little purpose - would not be worthwhile.

High rates of omission will also be of interest, both as an indicator of possible ambiguity and, when compared to position in a test, information on the timing and administration of the test. Item analysis for items in SOFRP then took the form of reporting as much organised data as possible for each item: discrimination indices, response patterns, frequencies of response for each answer possibility, omission rates, etc. Clearly the amount of data manipulation involved was very large and the facilities for producing the relevant information by the analysis of test data was written into a computerised item banking system, along with other facilities, described in detail in Chapter 17, below.

Item construction proceeded more on the basis of content referenced to the criterion of job-related reading task than on the basis of either analysis or computer system. In fact, the programming of that system involved a number of tortuous innovations to fit the system to the required item-types rather than the other way around. Analysis played its part, however, and items were constructed to the item-types considered above, with each to be either correct or incorrect (one mark for each), with no partial credits for some, but not all, answer possibilities in action items. Each action item had no more than ten answer boxes and no less than three, and no more than four comprising the correct pattern. Deliberate ambiguity was avoided. Completion items were exceptions to the analysis procedure: available only to the marker to pronounce correct or incorrect (there appeared to be no gain in coding a response of some complexity for the computer system, when in so doing the marker could then pronounce on the answer). These items were marked as omit, incorrect or correct for the purposes of analysis.

Selection or Construction of Reading Passages

Each item was to be based upon one reading passage, be it continuous prose, prose plus diagram, a list, labelled diagram, a form or whatever. A very large amount of material was available from the companies and organisations visited. A certain amount of particular types of material was not available, however, in the main being the internal and confidential documentation of company offices, but also material either expensive to obtain or provided by other sources. Where this was the case, materials were either borrowed briefly or analagous materials constructed incorporating the main features. Apart from these, all items were based on passages collected from the sample. Changes to materials were made only to remove identifying marks or to alter numbering systems or to correct typographical errors or particularly ambiguous phraseology. Passages were excerpted in the majority of cases, rather than complete documents for reasons of space and timing per item.

Specific types of passages were sought for from the materials. For example, an apprentice deals with sets of procedural directions and apprentice material was searched for sets of such directions. The most representative of these was usually chosen; that is, the one containing most of the features common to all the sets available, such as numbered steps, separate headings, diagrams etc. Where no pattern was clear, a number of separate items were constructed based on several passages.

This process was repeated for all jobs and types of content. Of course, this aspect of the project started almost as soon as the collection of materials did, and a number of revisions were necessary, whilst types of passage and items were duplicated. With the development of the computerised item-banking system mentioned above, it was decided that a large number of items be produced for use in future tests or in parallel versions of the one being developed, rather than a number of items purely for one test.

There were few restrictions on the use of collected materials and most documents were considered for use, whatever their format: legal documents, handbooks, manuals of instructions, forms, stocklists, coding forms, etc. A number, however, clearly were not, strictly speaking, reading tasks but writing tasks which involved filling in numbers with no reading associated with them. Further, other materials were rejected as being those requiring complete oral explication in the context in which they were used, by a supervisor or trainer. Other materials, though collected, had been placed under a restriction to be used for internal purposes only. Their use in item construction was therefore limited to the comparison with other, similar, passages to select common features as discussed above.

As size varies in such materials and as photocopying is often inadequate for further reproduction, most passages were retyped onto A4, and diagrams transferred. Where possible, however, good originals were kept as received.

Item Construction

There will always be a tendency in item construction to use a particular passage because if is easy to write items about it, or to use a particular item type because a passage offers a nice set of answer possibilities for that type rather than another. It is a tendency to be avoided if at all possible, to avoid items with a basic mismatch in terms of content validity. Trivial items are also unclear.

Certain item types fit certain passages, according to the principles of item construction discussed above: for example, an action-type item fits a 'standards and specifications' passage, as the sort of job-related reading task with such materials often requires more than one aspect to be considered in order to be 'right'. Further, the 'what would you do next?' reading task suggests multiple-choice with 'procedural directions' or 'checkpoints' passages. Completion items lend themselves to forms, job cards, etc, where the reading tasks are quite short yet self-contained.

Using these guidelines, a large number of items were constructed. Action and completion items used one reading passage each. There were three multiple-choice items per passage. The latter were much easier to write and more were constructed than the other types. This was not the only reason, however, as most content categories and passages can best be tested via this question-type. A number of passages were not amenable to an item of the appropriate type being written about them - due to brevity or format - and rather than exclude them, one of the other types was used. At the stage of item construction, this was seen as legitimate as invalid items would be excluded at the next step, content validation (see Chapter 9).

The real, underlying, guidelines are best explained, however, by reference to Murphy (1973), who gives the following instructions to his item writers in the Adult Functional Reading Study: "Each task must look real. If the stimulus is to be a medecine label, wherever possible obtain an actual label rather than merely typing the text onto a separate piece of paper. The same holds true for pamphlets, forms, contracts, newspaper ads etc. The task must copy faithfully as possible the real world of reading. As a task writer, one of your major concerns will be the face validity of the materials you produce. They must 'look real', have some evident benefit to the respondent and be directly related to the kinds of reading most people do ... Remember that the difficulty of the task is to be a function of the stimulus material, not the questions we ask about it" (p. 52 - 53).

Examples of questions and reading passages (an item is one question and its associated passage) can be seen in the Functional Reading Test, Form A, given as an Appendix to this work (Appendix VII). Questions 1 to 6 inclusive are action-type; question 7 is a completion item; and questions 8 to 31 are multiple-choice.

CHAPTER 9

CONTENT VALIDATION

Introduction

The validity of any test is the extent to which it actually does measure what it purports to measure. Whilst this seems very straightforward, the more complex the test and the more it is based on complex, interrelated items and concepts, the more difficult it is to measure its validity. Does one go back to as many of the underlying traits and validate for each, or does one accept the high-order skills and attempt validation for them? Developers of criterion-referenced tests have used many forms of validation, usually within the four categories used in norm-referenced measurement: construct, concurrent, content and predictive validity. Some authors have used different labels for validity but these seem essentially subsumed in one of the above (e.g. Ganopole's 'descriptive validity' and 'domain-selection validity' (1978)).

Construct validation is the attempt to relate a testee's performance on the test under development to his performance on measures of traits or skills underlying that test. Koos and Chan (1972) related pupil's scores on a CR biology test to scores on tests of verbal reasoning, reading and critical thinking and obtained quite high correlations. Ewen, Gipps & Sumner (1975) confirmed the construct validity of some of the aspects of the 'Proficiency in English Tests' (NFER, 1973).

Concurrent validity is established by relating performance on the test under development to other tests measuring in the same or a similar area. Again, Koos & Chan (1972) related pupils' scores to a behavioural checklist and with scores on a 'Processes in Science' test (obtaining low correlations). Young, Knapp & Michael (1970) correlated the scores on their 'Tests of Achievement in Basic Skills' with pupils' course marks in the same areas (although the time difference between the measures also suggests predictive validation).

Fremer (1973) also suggests that a criterion-referenced test be validated against a properly constructed performance test of the same domain.

Use of the above two types of validity is less common in criterion referenced measurement than in norm-referenced measurement, however. Testers are usually more concerned with performance or attainment in some skill than in how this is underpinned by some theoretical construct, or the location of a testee on that construct's continuum. In SOFRP, one is concerned with the specific application of reading skills, in functional reading, and the performance of pupils in relation to those applications, rather than the underlying skills - more properly tested by a general reading test. Concurrent validity is also infrequently used in CRM, as development is often into new areas of assessment, usually where norm-referenced testing fails to give a sound basis for decision-making. Hence, there are no concurrent measures for comparison. Restricted variability in scores can also undercut the value of the correlational techniques used in these methods. However, it seems likely that general reading ability should be linked to functional reading performance. In particular, those with low general reading ability will also be expected to have poor functional reading scores. A measure of concurrent validity for a functional reading test will be its correlation with a test of general reading ability. That correlation would not be expected to be high, however, as the two tests are of different aspects of the reading process.

It is content validity that is most generally accepted as the type for use in criterion-referenced measurement. This is the assessment, by various means, of the extent to which an item represents some aspect of the criterion or domain under consideration. Thus, a test of reading performance should contain reading tasks and not spatial relations tasks; the latter is a proximate measure, of an underlying construct, but not a measure of reading performance itself. One could, of course, change the objectives of the test to include it, but it has no content validity for the first test. Measurement of content validity has tended to go one of four ways. Mastery test developers often justify the inclusion of items on the basis that, having been constructed from set objectives, they have, a priori, content validity (see e.g. Cox (1970), Osburn (1968)). Others seek some empirical assessment of this validation, using pre- and post- test evaluations to assess the 'sensitivity' of individual items (e.g. Roudabush, 1973, Ozenne, 1971). Most common, however, are the two types of judgemental assessment by experts. Here, experts (those presumed by virtue of their position or performance to have special knowledge of the criterion or domain) assess either the relevance of a specific item to the general domain from which it has been derived, or the relevance of an item to the specific objective for which it was written. Martuza describes this latter as "having two or more content specialists judge the relevance of each item to the objective it is intended to measure and ... , using some index of interjudge agreement as the measure of item content validity" (1977, p. 283). Rovinelli & Hambleton (1976) used their own version of the Hemphill-Westie index of homogeneity of placement (1950), the Index amongst experts on their placement of a given item for a specific objective (see also Hambleton et al. (1975)). The fourth type of content validation is the use of panels of experts without sophisticated interjudge congruence assessments. Such studies typically involve a broad domain, unsuited to the development of objectives, such as Murphy (1973). The task of such panels has been variously described as the decision as to 'which items to use based on their knowledge and experience of the field' (Klein, 1974), the judgement of 'the congruence between the items and the domain specifications" (Berk & DeGagni (1979)) and a judgement 'based on the test's apparent relevance to the behaviours legitimately inferable from those delimited by the criterion' (Popham & Husek (1969)). Murphy (1973) used a panel of twenty-four persons, representing industry, commerce, education, government, journalism and consumer groups to assess the content validity of items for the Adult Functional Reading Study. The development of the Progressive Achievement Tests involved hundreds of teachers sitting as curriculum/content experts (Elley & Reid (1969)). Teacher judgements were also used in the Michigan Assessment Program (Royal, (1974)).

The validation task before the SOFRP at this stage of the research is much the same as that of the Adult Functional Reading Study (Murphy, 1973): to ensure that the items constructed are representative of tasks encountered by school-leavers on starting work. In another context, Gatewood & Schoenfeldt express the point thus: a test is "construct valid if the knowledge, skills or abilities being measured are actually essential and critical for successful job performance" (1977). Hence, for SOFRP, to be content valid, the passages selected must be common or very important and the reading tasks tested via the question must be essential and critical, or frequent. The frequency of tasks is dictated from the situation in which a school-leaver finds himself: his tasks may be fairly trivial but he must be able to cope with them. On the other hand, the infrequent but highly important task cannot be neglected. The assessment of such questions is best decided by pushing the whole question back to those who first provided the data. Having manipulated the data, the researcher must take care he has not changed it unrecognisably, nor misused it to the point of uselessness.

Predictive validation of criterion-referenced tests occurs less often, presumably for lack of an associated future measure. Hambleton, Roberts & Traub (1970) compared pupil scores on a criterion-referenced test with a conventional test given at a later date as a measure of predictive validity. Sticht (1975) and his co-workers used various predictive variables, including job-related reading task tests, to investigate the test to which they were related to job performance. Part of the SOFRP is the predictive validation of a test of occupational functional reading ability by relating test scores to job performance ratings, and this is discussed in detail below (Chapter 13). A discussion of the content validation of the items constructed in the course of the Project forms the remainder of this chapter.

Panels of Experts

In considering the content validity of items for a test of occupational functional reading ability, it can be seen that two panels of experts are needed: the "employers" - personnel and training officers, etc. involved in the survey to collect data - and experts in language itself. The former may judge the relevance and importance of a given item but not the linguistic aspects of the way in which the task was posed. The latter are more appropriate persons to undertake such a scrutiny, to assess, for instance, whether the question is more complex than warranted, asked in the passive voice when the reading passage is active, etc. As Murphy (1973) wrote: "the difficulty of the task is to be a function of the stimulus material, not the questions we ask about it" (p. 53, previously cited). The memberships and functions of these two panels are discussed below.

Employers Panels

Well over 250 items had been constructed according to the procedures discussed in the previous chapter. A number of these were based on materials used to induct a young person into the company, the remainder were job-specific. It was decided that, rather than have a large and unwieldy panel of employer experts, a number of panels would be formed, one for each job-type, with every panel scrutinising induction materials. It was felt that a range of interests should be reflected in the panels and a number of employers or persons from related organisations were invited to attend one or other of the panels. These invitations were issued to persons involved in companies taking on a large number of school-leavers, of different jobs. The panel on which they were invited to serve was determined by the category into which they recruited most leavers. Other persons were invited to serve on panels where it was felt their contribution would be most valuable.

Three panels were eventually convened (apprentices, clerical workers and distribution operative/others). It proved impossible to co-ordinate a panel for operative/others and this panel was conducted by post, letter and telephone, with the members providing written comments on items and clarification being obtained by telephone.

Membership of the panels is listed below:

Director (with responsibility for personnel and training), large Caterers
Manager, major branch of a supermarket chain
Personnel Officer of large steel manufacturers
Personnel Officer of major engineering company
Employment Officer of large tool manufacturer
Senior member of Iron and Steel ITB
Senior member of Engineering ITB
Training Officer of medium sized cutler
Senior Careers Officer
Manager of branch of large insurance company
Manager of branch of major bank
Training Officer of same
Staff trainer of major distribution outlet

Although no panel member specifically requested anonymity, it is felt proper not to reveal their names, nor that of their establishments. Responsibility for the test and its development rests with the Polytechnic and not with those who so kindly gave up their time to assist.

The procedure for validation was kept as simple as possible to avoid any ambiguities or confusions allowing a suspect item through the 'net'. Panel members were sent a booklet containing all items for their panel, including all induction items, at least one week in advance of the meeting. Instructions for scrutiny were sent with these items and these are given in Appendix III. Panel members were firstly asked to go through the reading passages and to assess whether they felt they were representative of passages from their industry (or firm alone): representative meant similar in style, format and structure, typical of the sort of material common to their industry. Each item had a form attached, for completion. Members were asked to signify whether the item was irrelevant, obscure, not used in that industry or not used by a school-leaver. The members were asked to look at each question and to judge the relevance and importance of the task in the job-related reading of a school-leaver. Other written comments were welcomed and a separate page was available to list materials not represented by those in the booklet.

Each panel was convened over one morning, with members invited to stay for lunch. Each item was discussed with the researchers, along the lines outlined above. Other, more general comments tended to be elicited over lunch.

Following each panel, the researchers met to discuss each item. The majority were validated without question, whilst others were rejected for various reasons (particularly, that the task or material involved was more difficult than that actually undertaken by school-leavers, even though they did encounter it). A number of items were revised for clarity, to include a more relevant answer possibility or to exclude a trivial aspect. Some were rejected as tending too much towards a non-reading task. The remaining items were forwarded to the Linguists' Panel.

Linguists' Panel

Two linguists from the Polytechnic were invited to assist in a further scrutiny of the items, this time on the basis of the language, not the content, of items. Three areas were under consideration:

(i) the use of the Sticht classification, to judge whether the categorisation had been undertaken correctly and whether any additional categories were necessary;

(ii) the position of invented items, to judge whether the invention of items to reflect materials not otherwise available had introduced any bias or invalidity, on the basis of their own experience and compared to any relevant materials that were available;

(iii) the complexity of the questions, to judge whether any undue complexity or unnecessary adherence to formal grammatical structure had been introduced, particularly in relation to the reading passage. The questions were not to be more complex than the passage.

Again, all items were made available before the panel meeting, with notes on the requirements. The panel met for several hours on one afternoon and discussed each item in the terms given above. It was suggested by one of the linguists that a further classification system be introduced, to assist in the identification of specific difficulties: that of linguistic task. It was argued that classification had been made for industry, job, content of reading passage and for question type, but not for the 'question in relation to the reading passage' i.e. the linguistic task underlying each item. This was readily accepted, and a system adopted. This is given in Figure 9.1 below:

(i) Referential: where the question pases a task requiring locating correct information;

(ii) Regulative: where the task requires an answer in terms of a behaviour, either an act of commission or omission;

(iii) Attitudinal: where the question requires the description of a state of mind, or approach to a task;

(iv) Definitive: where the question requires a description of the meaning of a word or words.

Figure 9.1: Linguistic tasks

Further amendments and deletion of items were made. All surviving items (173) were placed on computer storage and formed the basis of the computerised item-banking system (see Chapter 17). The number of items in each category for each classification system is given in Tables 9.1 to 9.4 in Appendix I.

CHAPTER 10

PILOT TESTING AND TEST CONSTRUCTION

Introduction

It is common practice to pilot sets of test items amongst members of the target population. This piloting forms a further empirical basis for accepting or rejecting items and provides general administrative information. In particular, ambiguous items need to be identified, as do negative discriminators and to indicate the types, if any, of items which are most easy and most difficult. This latter is not for the purposes of exclusion or retention, but to give the researcher some idea of how hard or how easy any particular test he devises will be. This will be important in recommending its use. Administratively, it is necessary to determine the number of items which may be answered in the time period and "to discover weaknesses or needed improvements in the mechanics of test taking, in the directions to examiner and examinee, in the provisions for the responses, in the sample or fore-exercises, in the typographical format, and so forth" (Conrad, 1951, p. 251).

Such piloting should, of course, take place amongst the target population of the final test and consideration must be given to the selection of such a sample. Criteria must be advanced for the rejection or amendation of an item and means provided for repiloting amended items. Following on from this, it is possible to construct a functional reading test for evaluation.

Target Population

SOFRP aims to develop tests of occupational functional reading ability, for use with Fifth-Form pupils in their last year in Sheffield secondary schools. The exact target population are those Sheffield Fifth-Form pupils intending to leave at the age of 16+ years, intending to seek and gain employment. Of course, any test developed should be valid for other cities or areas with similar industries. Unfortunately, it is difficult, if not impossible to sort the leavers from the non-leavers in a Fifth-Form, as progression to Sixth-Form or Further Education is so often determined by examination results published in the summer following the leaving date. Also, from the fieldwork it was clear that those with several O-levels had no reading difficulties at work and would likely find a functional reading test very easy. It was felt, therefore, that more fruitful data might be obtained by piloting amongst the 'effective' target population: those with middle to low ability (i.e. those predominantly taking CSE examinations, or few examinations at all) rather than boring the higher ability pupils or subjecting remedial pupils to an ordeal. Further, it was reasoned that if the middle-low group answered correctly so would the higher group, if incorrectly then also the lowest group. Much information was then to be gained from the middle-low ability group.

For the purposes of the pilot, the target population was restricted to those predominatly taking CSE examinations or few examinations in the Fifth-Form of state schools in Sheffield Metropolitan District.

Sampling the Population

Piloting of test items need not involve a large sample, as it needs only a small set of responses from a typical group to indicate patterns of response, point out ambiguities etc. If one has a large number of items, however, one must test several groups, rather than have all subjects take every item: fatigue will play its part and is to be avoided.

Sheffield has thirty-eight schools (all comprehensive) with Fifth-Form pupils and, due to the pursuit of enlightened policies, cachement areas have been organised to attempt a mix of pupils in each school wherever possible. Inner city schools are thus not the blighted denizens of the working class, nor the suburban schools the privileged paradises of the middle-class. The system is not perfect, of course, but it does mean that any one of a number of schools can be selected as being fairly representative of the whole population. 'Fairly' is a judgemental term but when such a judgement is left to one familiar with every school in the area, one can be reasonably sure of some accuracy. That person was a Senior Advisor in the Local Education Authority, who assisted in the arrangements for contacting each school used in the Project at this and later stages.

Just one school was used in the initial piloting, in an area of above average S.E.S. but with pupils from other areas attending. It was a school of good reputation but not the highest flyer. There were approximately 200+ to a year group, covering the range of ability and it can be said that this was a fair sample of Sheffield Fifth formers. Only one school was used in order to minimise the amount of travelling and administration involved in visiting several schools.

The school was asked to provide, in the first instance, groups of about 10 to 15 pupils, of middle to low ability.

Assembly of Test Booklets

All items were assembled into six test booklets of twenty or twenty-one items each. All booklets had some action or completion items as a first section, preceded by an example and a page of introductory instructions. This was followed by a second section of multiple-choice items, again preceded by examples and instructions. One booklet also contained an orally administered item.

For each item, the reading passage preceded the question part of the item. The booklets were bound with title sheets with room for the testee's name and school's name.

Administration

Three members of Polytechnic staff, including the author, undertook the pilot testing and common administration instructions were agreed. These included a common introduction to the tester and the nature and reasons for the testing. The instructions then continued with the reading out of the instruction page in the text booklet and instructions for the example. Similar instructions were given for the second section. In addition, each tester had a timing schedule on which to record administration times, the time after which two-thirds had finished and when all had completed, for each section.

The testers each tested two groups on the same morning in the pilot school. Each group had about ten pupils, assigned to each group in no particular order, and each group took a unique set of items. No school staff were present at the testing.

Criteria for Item Scrutiny

In a sense, item rejection, revision or retention is undertaken on a largely judgemental basis. Following the discussion of item analysis in Chapter 8, above, the usual item statistics are only useful if they have high values (restricted variability will lower the values: therefore a high value can be trusted to be no lower, but a low value cannot be trusted not to be higher). An 'inoperative' answer possibility may not need changing, for one is testing to see how well the testee can do, not to see how well one can lure him from the correct answer.

One or two things do have value, however. An incorrect answer possibility selected very frequently - particularly more often than the correct one - indicates an ambiguity, probably in the question stem or the answer possibility itself. A uniform spread of answers across the possibilities tends to indicate guesswork: the item may be either very difficult or ambiguous. A high discrimination index would be evidence for the former, a low index might indicate the latter; a negative value certainly would indicate the latter. A high rate of ommission indicates that the testees are probably failing to understand the question part of the item. If the passage were misunderstood, one would expect some guessing or wrong answers, but not omission. The observant tester will also notice items upon which a great deal of time is spent by pupils.

Scrutiny of Items and Repiloting

All test items were subject to scrutiny on the basis of the criteria discussed above. Most defects in items were minor, having already passed through content validation procedures designed to find ambiguities or other undue complexity. A number did, however, require revision of either the question stem or a specific answer possibility. A number of typographical errors were also discovered and corrected.

On a more general note, it was decided that, to reflect more accurately job-related reading tasks, the question should precede the reading passage. It was felt that a young person at work more often tends to go to a passage with specific purposes, rather than reading through it and then answering questions.

A number of items were deleted as inappropriate materials for the test, on the grounds that to remove the ambiguity or other defect would trivialise or destroy the item per se. Others were amended and these were repiloted in the same school. This repiloting also gave the opportunity to correct the details of administrative procedure and to check on the timing of the test. Time on the practice items had been found too lengthy and also, the change to the question being placed first meant changes in procedure and explanation. Two groups of about thirty pupils each were given a new booklet each for this purpose, tested by the author and another tester in one morning.

Results of Piloting

In all, 173 items were successfully constructed, validated and piloted, although some were different versions of the same item.

It was found that, within the double school period allotted for testing (70 minutes), a maximum of 32 items, of mixed types, in two sections, could be answered. This figure might be exceeded if all items were of multiple-choice, or vice versa for action or completion items. It was decided that an absolute maximum of 35 items per double school period was likely, given time for introduction and administrative procedures.

In general, the administrative procedures were acceptable, given that the action and completion items were novel types to the pupils, who took a little time to get used to them. Although ideally it was desired to allow all pupils to finish, the need for separate administration instructions for different sections indicated that a time limit would need to be imposed on the first section, to ensure that pupils were able to continue to the second section without individual instruction. Pupils would have free license to return to the earlier section if they desired, however.

Construction of a Functional Reading Test

With a large number of test items successfully validated and piloted and with information available on the number of items which may be included in a test, the question of constructing a test of occupational functional reading ability arises. Such a test must aim to include a wide variety of items drawn from different jobs, of different types of content and joining different linguistic tasks. All of this must fit within the framework of the result of the SOFRP and the empirically determined criteria previously discussed.

Content of the Test

Of the six content-types derived from the Sticht classification, items were available for five ("Tables of Contents and Indexes" had no items constructed for it as the questions proved tended to be more complex than the reading passage, or be too ambiguous, or the reading passages available were inappropriate for item construction). Of the seven job types (including "Unemployment and Job Seeking" in category 7), there were no items available for "Professional" for reasons previously discussed (Chapter 7). Of the four linguistic tasks, no items of the "Attitudinal" category survived piloting.

In order to have every combination of these categories, it would be necessary to have at least ninety items in the test (5 x 6 x 3 = 90). This was clearly impossible within the double school period time limit. It was decided to construct two tests, therefore: Form A, which would be as complete a sample as possible of the different combinations of job, content and linguistic task; and Form B, which would provide extra items to complement Form A in decision-making. That is, Form A would be a test in its own right, but optionally, more information could be obtained by also administering Form B. The combination of these (Form A and Form B) would give a more complete picture of pupil performance if necessary. Both tests would be timed to last a double school period each and contain roughly 30 items.

Functional Reading Test, Form A

Form A consisted of 31 test items. No category, of content, job or linguistic task was omitted except "Unemployed", and these categories formed subtests for more detailed analysis. The following figures (Figures 10.1 to 10.3 show the numbers of items in each category:

Job Category	Number of Items
Apprentices	14
Clerical	5
Distribution	4
Operative/others	2
Induction	6
Total	31

Figure 10.1: Items per Job Category, Form A

Content Category	Number of Items
Standards & Specification	5
Identification & Physical Description	10
Procedural Directions	6
Procedural Checkpoints	4
Functional Description	6
Total	31

Figure 10.2: Items per Content Category, Form A

Linguistic Task	Number of Items
Referential	17
Regulative	13
Definitional	1
Total	31

Figure 10.3: Items per Linguistic Task, Form A

Form A is given as Appendix VII, below. It comprises of six action items, one completion item and twenty-four multiple choice items.

Functional Reading Test, Form B

Form B consisted of 30 items, the scores on which were to be added onto those for Form A, to comprise a long form, if desired. The following figures show the numbers of items in each category:

Job Category	Form B	Complete Long Form
Apprentices	3	17
Clerical	4	9
Distribution	10	14
Operative/others	9	11
Induction	4	10
Total	30	61

Figure 10.4: Items per Job Category, Form B and Complete Long Form

Content Category	Form B	Complete Long Form
Standards & Specification	12	17
Identification & Physical Description	14	24
Procedural Directions	3	9
Procedural Checkpoint	0	4
Functional Description	1	7
Total	30	61

Figure 10.5: Items per Content Category, Form B and Complete Long Form

Linguistic Task	Form B	Complete Long Form
Referential	14	31
Regulative	9	22
Definitional	7	8
Total	30	61

Figure 10.6: Items per Linguistic Task, Form B and Complete Long Form

CHAPTER 11

TEST CONSTRUCTION - CONCLUDING REMARKS

The processes leading up to and including the construction of a test must always be the most difficult and arduous steps in its development. At least, in test evaluation one is dealing with a finished product, whereas in construction, one has to go out and gather in a wealth of data and impose some order upon it, then painstakingly over-produce items, see how they perform, discard and change them and then select those survivors most appropriate to the purposes of the test. It is valuable, therefore, to look back at the construction process for FRT A and B and see how it might have been improved, what was particularly interesting, etc., before going on to consider how good its products, the tests, are.

In social and educational research, a study often stands or falls according to the quality of its sample and the procedures for achieving that sample. It is felt that the samples of firms, ITB/GTA's, colleges and school-leaver employees used in this study have much to recommend them, both in terms of sampling method and the rates of response achieved. There were, at the time of sampling, 624 firms employing thirty or more persons in Sheffield in the industrial categories under study. The sample of 117 firms therefore represents approximately 19% of those firms, and the response of 78 firms represents over 12%. Such a sample can compare quite favourably with others in similar fields (e.g. Knox (1977); Freedman (1979); Anderton (1979)). A number of exclusions were made before sampling, however, and despite the detailed arguments given in Chapter 5 for these, one may wish to ask whether there was any significant loss of information arising from this procedure. It would have been of great interest to study the Armed Forces, the Civil Service and the National Health

Service, of course, and, certainly, their individual uniqueness would have contributed data of a significantly different kind to the study. Yet one must continually bear in mind the local nature of the study and the restriction in resources available. The bodies mentioned have resources available to undertake their own research in this field in excess of those available to the SOFRP for its entire work. The exclusion of low employing categories is seen as entirely justifiable in view of the actual time that each interview entailed. Similarly, firms employing less than thirty people are unlikely to take on school leavers in any consistent fashion.

The response rate of nearly 71% of firms contacted is seen as quite exceptional, considering a comment in evidence to the Expenditure Committee to the effect that employers can receive up to two requests per day of a similar nature (Hof C, 1977). Very few outright refusals were received, and the high response rate must be seen as a combination of the employers' interest in the work of the Project and of the method of approach: by letter and following telephone call to arrange an appointment. Except for those who were willing to participate but with whom it was impossible to arrange an interview, it is not thought that a higher rate of response could have been achieved.

It is regrettable that the college of further education involved in training construction industry apprentices declined to assist in the SOFRP. The information from the other colleges was very full, however, and most useful. Perhaps more information might have been gained by attendance at some of the lectures and other teaching sessions, but it was not realised that the colleges provided as much materials as it turned out they did.

Apart from their nervousness at interview, the corroborative information of the school-leaver employees was direct and useful. It is still felt, however, that better data might have been collected by detailed observation of reading in situ; but as has been mentioned, this did not prove acceptable to employers. A separate study for this purpose is suggested for others to undertake.

In the analysis of the information and the reading passages collected, the use of classification systems proved exceptionally useful and led directly to a more organised approach to item construction. This field, of the analysis of job-related reading materials, and, in particular, the assessment of readability, remains open for a great deal of further investigation. The construction of novel item types posed a number of problems in terms of objective marking, partial credits and deciding which type of item best suited a particular passage. The criterion of attempting to reflect a job-related reading task is seen as particularly useful, although the restriction to paper-and-pencil testing made this especially difficult for some tasks.

Techniques for item analysis were, to a large degree, fitted to the circumstances from a range of sources, whilst others were developed specifically for the SOFRP. There is a general lack of agreement as to the appropriate techniques to be used, and very few general criteria have been advanced. Due to the use of non-standard item-types, criteria for the acceptability of items were developed for the SOFRP uniquely and may not prove useful elsewhere. Unfortunately, the weight of technology in this area has been in mastery testing of unidimensional traits.

Throughout criterion-referenced measurement, it is recognised that the content validity of test items is of crucial importance. When one is using a broad domain of tasks, the use of panels of experts seems to be the only possible way to achieve such validity. There are other advantages also, for the use of panels involves the members in the Project and with each other, to the promotion, hopefully, of more attention to the materials used in industry and commerce.

The linguistic tasks defined by the linguists' panel need closer study than has been possible in this work. It may be better to adopt another label than 'linguistic', such as 'reader approaches to text'. Further study allied to analysis and readability might do well to consider the nature of the linguistic variables inherent in the reading situation and in the test.

Pilot testing of items and administrative procedures was badly affected by atrocious weather conditions and industrial action. For the testees, the strangeness of testing was compounded by the abnormal disruption of school life in general by heavy snow. Such continual distraction may have served to depress their scores, and perhaps have affected different ability groups in different ways. There are also a number of disadvantages to testing only in one school. The pupils may well either become unusually sophisticated at responding to the items or may react against being tested so often. In the former case, the timing of the test is thrown out, in the latter the evaluation of the items is made more difficult. The number of occasions of testing were no more than three, however, and the school staff attempted not to put forward the same pupils each time. It might also have been valuable to undertake rather more pilot work on, say, different ordering of items and sections, conduct interviews with pupils on their attitudes to the novel item-types etc., but circumstances did not permit this. Whilst it is felt that the pilot work was adequate for the purposes involved, it is acknowledged that a number of other advantageous studies might have been made.

In assembling the functional reading test, it would have been of particular value to have a larger number of items - one for each subtest combination, 90 - but this was clearly impossible. The final selection of items for tests A and B are, in a sense, quota samples from the Item Bank, and their inclusion is judgemental. This certainly does not invalidate the tests but particular attention needs to be given to the evaluation of the tests.

In all, then, the construction of Forms A and B has followed the paths laid down in the basic remit of the SOFRP, to identify and classify job-related reading tasks and to use the materials involved to construct a criterion-referenced group test. The following section of this work is concerned with the detailed evaluation of the tests developed here, firstly for reliability of their measurements and secondly for the relationship of these measurements to job-related reading: predictive validity.

CHAPTER 12

TEST RELIABILITY

Introduction

The method of assessment of a test's reliability depends upon the use to which the test is to be put. If the designer is concerned with the question of whether all the items measure the same thing i.e. the homogeneity of a unidimensional test, he will use a measure of internal consistency, or a split-halves technique. He might develop an equivalent form of the test by identical processes to see if they matched up in the responses of the same pupils to them. He might be concerned with decision-making and wish to assess the reliability of the test in suggesting the same decision on more than one occasion for the same testee i.e. the stability of the test. Most criterion-referenced tests are to do with making decisions about the attainments or abilities of individuals or about the efficacy of instructional procedures. In each case, the reliability of that decision is of paramount importance, and the test must contribute to that reliability by being consistent in its recommendations. The testees' scores must tend to be the same on more than one occasion of testing; and this is true whether the test is one that classifies a testee into 'master' or 'non-master', or one that uses individual scores as part of a larger decision-making process. In fact, Swaminathan, Hambleton & Algina (1974) go as far as defining the reliability of a criterion-referenced test as "the measure of agreement between the decisions made in repeated test administrations". Whilst this is not the case for all criterion-referenced tests, it is sufficiently so for the majority of effort in the area of reliability theory to be concentrated on methods for most accurately assessing the true scores of individuals, and hence the reliability over time of the test. Almost uniformly, they involve the test-retest procedure.

Such procedures, however, expose the test materials to the testees on two occasions and are therefore prone to a practice effect, and to other testee-related effects (such as rejection of the testing 'discipline', where repeated testing causes the testee to falsify his responses deliberately in reaction to the task). Slower workers may reach items previously omitted due to lack of time; quicker workers may become careless. Small changes in scores that such effects are likely to bring about have particular importance when the range of scores is restricted. In the extreme case, where all high scorers were careless and lost one mark because of it and all low scorers gained one mark through practice, all these changes would produce a low estimate of reliability, when in fact all testees had similar marks on both testing occasions.

A problem related to this and following on directly from it, is that of homoscedasticity, where the test may have a different level of reliability for different ability ranges. For example, the slower worker, mentioned above, whose score improves on retesting contributes more to a lower estimation of reliability by changing his score, than does a top-scoring testee whose score stays the same. Further, a test with a ceiling-effect, with many scores clustering around the higher score range, may have different reliabilities for that cluster compared with a spread of scores lower down the score range. This may, of course, be a function of the variability of the two ranges: cluster with low variability and spread with high variability. Nonetheless, these aspects of the procedures for estimating test reliability must be borne in mind.

It is clearly of importance that the Functional Reading Tests, Form A and Form B (FRT A and FRT B) be reliable over time and that a test-retest assessment will be appropriate. It will not, however, be sufficient to assess the reliability of the total test score alone. One of the major difficulties in a heterogeneous test is the assignment of a meaningful interpretation to a summative score. Only if the items are hierarchical in some way can a single score exactly state what the testee can and cannot do - a part of the definition of a criterion referenced test (e.g. Nitko, 1970 p. 653). Davis suggests that the conditions for this to be exactly the case in any test are never met in actual practice (Davis, 1974 p. 45). One method of overcoming this difficulty is to report subtest scores within a heterogeneous test. Brown (1970) indicates that reliability coefficients are often reported for the summative score only, rather than subtests as well. Ganopole (1978, p.45) suggests, however, that "it is only when reliability estimates are reported for each subtest for which a score is derived, that we can have any assurance that the subtest possesses acceptable reliability". It will be worthwhile to compare and report subtest reliability also.

To pursue the argument a little further, in a heterogeneous criterion-referenced test, comprising several subtests, is it worthwhile to report a summative score, rather than subtest scores? For, if the real information about a testee's performance comes from the subtests, of what use is a single figure? Clearly, it is only of value if the subtest scores are themselves related to the total score i.e. to get a high total score, each subtest score must be high also. If this is the case, the summative score will be an indicator of overall performance across subtests. This requires a measure of internal consistency, not of the item-total type (e.g. Kuder-Richardson 20) which is based on unidimensionality, but the establishment of an acceptable relationship between subtest and total scores. With such a measure, it becomes meaningful to talk about standards and cutoff scores. These are used frequently in criterion-referenced testing, particularly in mastery tests where a common practice is to establish, by some means, a criterion score to determine 'pass' or 'fail' grades for testees. It was the intention of SOFRP to investigate criterion scores empirically, following the arguments against their arbitrary determination advanced by Glass and Smith (1978), discussed in Chapter 4 above, but, as much of the following discussion hinges on the use of criterion scores, it will be useful to bear in mind that their use remains a possibility at this stage. The empirical investigation is discussed in Chapters 14 and 15 below.

It will also be of value to assess the reliability of the estimate of a testee's performance on functional reading tasks i.e. the 'domain status' of the testee. This is distinct from the stability of the test; there we are considering the reliability of the assignment of the testee to one of two or more mastery states. Here we are considering the confidence limits of the testee's score on the test.

So, then, three measures of reliability are sought for FRT A and FRT B: a coefficient of stability, or consistency of decision-making, for summative scores and for subtest scores; an internal consistency measure of the relationship of subtest scores to summative score and confidence limits for individual scores. It is to be noted that a high coefficient for test-retest for the summative score will be of little value if there is a poor relationship between the total and the subtest scores.

Estimate of Criterion-referenced reliability

It is in the estimation of reliability that the spectre of low variability in scores looms most oppressively. Popham & Husek summarise the problem neatly: 'Stability might certainly be important for a criterion-referenced test, but in that case, a test-retest correlation coefficient, dependent as it is on variability, is not necessarily the way to assess it ... If a criterion-referenced test has a high average inter-item correlation that is fine. If the test has a high test-retest correlation, that also is fine. The point is not that these indices cannot be used to support the consistency of the test. The point is that a criterion-referenced test could be highly consistent, either internally or temporarily, and yet indices dependent on variability might not reflect that consistency (1969, p. 5-6). Popham & Husek do not rule out the use of the usual product-moment correlation, and indeed, indicate that a high value means high reliability.

Jackson (1970) contends that in actual practice, there will always be variability in scores and so correlational techniques may be used. Yet in a situation where the measure itself may be unreliable, it is necessary to have other techniques to hand, to build a corpus of information upon which to base decisions. The use of the computer brings even the most complicated technique whithin easy grasp and there can be little excuse for not considering a range of possibilities.

Wedman (1974) suggests there are three schools of thought in the area of criterion-referenced reliability estimation: those applying classical techniques; those reformulating classical techniques for criterion-referenced measurement; and those advocating the standard error of measurement as a technique. A number of techniques are concerned with pre-insturction testing versus post-instruction testing (e.g. Ivens (1970) , Cox and Graham (1966)) and, as this is not applicable to SOFRP, will not be considered here. Below, the latter two schools of thought will be discussed, along with other suggestions that have more recently arisen.

Livingston (1972a) reformulated certain norm-referenced concepts for use in criterion-referenced measurement by using deviations from the criterion score, rather than from the mean score, as the equivalent of variance. This enabled him to develop a special correlation coefficient, of product-moment type, with moments taken around the criterion score rather than the mean. Harris (1972), however, argues that Livingston's coefficient may be inconsistent with different ranges of talent: "Livingston's "bigger" coefficients (than classically derived ones) can readily be secured by implicitly extending the range of talent" (p. 28-29). Further, he shows that the standard error of measurement for both techniques is the same and concludes that "his work fails to advance reliability theory for the special case of criterion-referenced testing" (p.29). Although Livingston's reply (1972b) answers many of Harris' criticisms, Shavelson, Block and Ravitch (1972) suggest that Livingston's coefficient is a function of the criterion as well as the individuals' responses to items. They further conclude that his coefficient does not have the same meaning as the classical reliability measures and so should not be counted as such. Divgi (1978) notes, from Livingston's equation for his index:

$$\frac{\text{var}(\tau) + (\mu - C)^2}{\text{var}(X) + (\mu - C)^2} = \frac{\rho + (\mu - C)^2 / \text{var}(X)}{1 + (\mu - C)^2 / \text{var}(X)}$$

(where τ and X are true and observed scores respectively, μ is the mean score, C is the criterion (cut-off) score and ρ is the classical coefficient of reliability), that, "for any given value of ρ, Livingston's index increases with the distance between the mean and criterion scores, and therefore it can have an appreciable value even when $\rho = 0$" (p. 3).

Other coefficients have been suggested or adapted for use in the assessment of criterion-referenced test reliability. Swaminathan, Hambleton & Algina (1974) proposed the use of Cohen's κ (Cohen, 1960) as a measure of the agreement in the assignment of testees to mastery states, between two testings, corrected for chance agreement. κ is defined as

$$\kappa = (p_o - p_c) / (1 - p_c)$$

where p_o is the observed proportion of agreement and p_c is the expected chance proportion of agreement. For two test administrations to the same sample of testees, for any given cut-off score, a figure similar to Figure 12.1 is obtained.

		First Administration		
		Master (M)	Non-Master (N)	Marginal Proportion
Second Administration	Master (M)	p_{MM}	p_{NM}	p_{2M}
	Non-Master (N)	p_{NM}	p p_{NN}	p_{1N}
	Marginal Prop.	p_{1M}	p_{1N}	

Figure 12.1 : Crosstabulation of Mastery proportions

where p is the proportion in each cell.

Thus,

$$p_o = p_{MM} + p_{NN}$$

and

$$p_c = (p_{1M} \times p_{2M}) + (p_{1N} \times p_{2N})$$

Coefficient κ has much to recommend it: it is simple to calculate and is intuitively attractive, for it seems most reasonable that the proportion of agreement in classification should be in direct relation to the reliability of the test. There are a number of objections to its use, however, which should be considered. Reid & Roberts (1978) compared κ with the coefficient ϕ via a Monte Carlo procedure, for the dichotomous case, and concluded that, in practice, the difference between the coefficients is so small as to suggest that the most easily calculable (ϕ) be used instead. Cohen (1960) noted that this would tend to be the case when the marginal proportions of categories were similar for test and retest. Subkoviak (1978) used four different procedures for estimating reliability of mastery tests. He notes that "the Swaminathan procedure (κ) produces unbiased estimates; but it requires two testings and standard errors are relatively large for classroom size samples" (p. 115). His comparison is with three measures all assuming item homogeneity (the Subkoviak (1976), Huynh (1976) and Marshall-Haertel (1976) estimates) and perhaps this is invidious with a procedure unbiased by the content in its estimations. Divgi (1978) writes that all the above measures are still dependent upon the distributions of ability of the groups tested, and suggests that, therefore, there is still reference to a normative model. He raises the point that if criterion-referenced measurement is to be group-independent, the measure of reliability should also be group-independent also. His coefficient, k , is however only for homogeneous item groups and, perhaps, the probabilities of misclassification given in his Table 2 are unacceptably high. Further, Swaminathan et al themselves indicate that the coefficient is dependent upon factors that affect the decision process (1974 p. 266). Martuza (1977) writes that p_o and κ are dependent on a number of factors associated with the decision process e.g. the value of the cut score (i.e. the score used to classify examinees as masters and non-masters), test length ... the number of alternatives per item if multiple choice items are used, and the homogeneity of the examinee group in which the decisions concerning mastery and non-mastery are being made. As a result, the values of p_o and κ are interpretable only when information concerning these factors is available" (p. 280 - 281).

Despite the above criticisms and qualifications, the use of κ seems straightforward and appropriate to the circumstances, and its adoption, with an open mind, might prove worthwhile. Studies involving its use did not report any difficulties (Berk & Degagni, 1979; Panell & Laabs, 1979; Close (1978)).

Millman (1974) has suggested a method for calculating the reliability of estimates of domain status using the mean absolute difference between scores on two identically generated tests. This would be inappropriate for a test-retest situation, however, where one might reasonably expect some practice effect to be apparent, increasing the size of the mean absolute difference. It might be argued, however, that there is a difference between testees who answer incorrectly on the first occasion and change to correct on the second, compared with those who change from correct to incorrect. A practice effect, in it gives pupils a combined chance to assess the possible responses to an item, may be assumed to act to cause a group shift towards 'correctness' rather than 'incorrectness', for a reliable test. That is, with a test which is inconsistent (due, perhaps, to ambiguity in some items), one might expect any extra time to be just as likely - if not more so - to act to cause a shift to 'incorrectness' as 'correctness'. It therefore will be instructive to examine any shift to 'incorrectness' via a count for each item or over the whole test, of the number of testees changing from correct on the first administration to incorrect on the second. The lower the count, the happier one can be that strange changes are not taking place within the test (Cf also, Beggs & Lewis, (1975) p. 200).

Other estimates of domain status have used the standard error of measurement, following classical procedures. As Hambleton, Swaminathan & Algina(1976) point out: "Whereas classical approaches to reliability estimation are affected adversely by the homogeneity of scores often obtained with criterion-referenced tests, the standard error of measurement is relatively unaffected " (p. 58), and can thus be used as an estimate of the amount of error associated with the domain score. Shavelson et al (1972) and Kriewall(1972) both recommend this approach. They seem, however, to make no allowance for severely skewed distributions, where neither assumptions of approximate normality nor of normal or binominal distributions of error may hold. If one is taking, as SOFRP does, the proportion-correct score as the best estimate of a testee's level of performance, it may prove unjustifiable to employ this method if the distribution of scores is skewed.

From this discussion of estimates of reliability in criterion-referenced measurement, one may select the methods which seem most applicable to SOFRP. Firstly, there can be no harm in calculating a classical reliability coefficient, for if it is high, it still indicates a reliable test (Popham & Husek, 1969, previously quoted on this matter). As a method of looking at the amount of agreement, independent of the variability of scores, coefficient k may well prove useful, although its high rate of indeterminacy may mean no productive value is obtained. For such purposes, k will be calculated for each cutting score, for the whole test and each subtest. Thus, no a priori determination of a cutting score need be made until after empirical investigation, but the data will be available. As a measure of reliability based on individual scores, rather than proportions above and below a cutting score, an adaptation of Millman's procedure will be used. For each test item, the number of changes in response ('correct' to 'incorrect' and vice versa) will be tallied and these 'Discrepancies' reported. The change from 'correct' to 'incorrect' will be considered as most important. A high proportion of changes in this direction will indicate an unreliable test. The calculation is performed for items, rather than for testees whose score declines as this gives an indication of which items, if any, are producing a shift to 'incorrectness'.

Assessment of FRT reliability

A test-retest procedure was used to assess the reliability of the two forms of the Functional Reading Test.

Form A (FRT A) was administered to 47 pupils in the first pilot school on two occasions separated by a three week interval.

Form B (FRT B) was administered to 44 pupils in a second pilot school on two occasions separated by a three week interval. This second school was selected after consultation with the Senior Advisor for Research and Evaluation of the Local Education Authority, and was considered to be very similar to the first pilot school in terms of cachement.

Reliability of FRT A

Test-retest results are given in Tables 12.1, 12.2, 12.3, 12.4 and 12.5.

The whole-test reliability of $r = 0.86$ (Table 12.1) is certainly sufficiently high enough to indicate that FRT A is reliable in classical terms. Further, all but one of the subtests also have acceptably high coefficients, to indicate their reliability. The exception is LING - 4, but as this has only one item this result was expected. The score distribution is strongly negatively skewed, which suggests that the results may yield some misgivings. The standard error of measurement is 1.6 but this is regarded as no particular use in view of the significant difference of the distribution from normality.

The correlation of the subtest totals with the whole-test total (Table 12.2) shows a high degree of internal consistency. This indicates that the summative score may be used as a convenient indicator of performance on the subtests also.

The coefficients κ and ϕ for each score of the test (Table 12.3) show a range of values, a number of which are of interest. Values of 1.0 are recorded for scores 10 and 11, an indication of perfect agreement between test and retest. There are only two testees with scores below these, however, and forty-five with scores at or above these levels. The areas to consider in the first instance are where most scores appear, for if the test is unreliable for the test-retest scores of most testees, it is of little value. The mean score is approximately 22 out of 31 items, with a standard deviation of about 5.7, and it would seem therefore that most testees will have scored in the range 16 to 28. This is the range in which to consider the coefficients, and within it their values range from 0.46 (27 out of 31) to 0.85 (0.86 for ϕ) for 19 out of 31. Some of these values are probably unacceptably low, but it seems likely that if any particular score is used as a cutting score in the final uses of the FRT A, a reasonable level of reliability can be attached to the decisions involved.

Values of the two coefficients are also available for the individual subtests of FRT A. Here, the picture is less clear. One expects the estimates of subtest reliability to be lower than the estimates for the whole-test when using product-moment correlational techniques, as the range of scores available is more restricted than for the whole test. This problem does not arise in using coefficients of agreement, however, as one is treating numbers of subjects (to be classified) rather than the number of items they have taken. The number (47) is constant throughout. The same level of acceptability applies to subtests, therefore, as for the whole test. If one wishes to impose separate cutting-scores for each subtest, then it seems that some coefficients again will be unacceptably low, rather more often than one would wish. In some subtests, coefficients do not rise above 0.60. The use of cutting scores with individual subtests must be undertaken with caution, therefore, if at all.

The response consistency (Table 12.4) data shows signs of a practice effect, with 11.05% of responses shifting to 'correctness'. In fact, 57.4% of testees improved their scores by one or more points (Table 12.5). One testee improved his score by 12 points, but this must be put down to individual situational factors as no other testees performed in this way.

The discovery of a practice effect is quite important, particularly when dealing with novel forms of test items. It may well be that the testees' improvement is due to increased familiarity with the administrative procedures, rather than the materials used for testing. The previously lower scores may be anomalous and not directly related to functional reading ability. This has implications for the estimation of a testee's true domain score, and it suggests that the proportion-correct score is not necessarily the best estimate available, as assumed here. It was decided to investigate this question empirically and this is discussed below.

Despite this possible objection, it does appear that FRT A has satisfactorily passed the tests for reliability discussed above. The product-moment correlation is high for whole-test total and for subtest totals; the intra-test correlations are high; and high values are available for κ and ϕ, using various cutting scores within the likely range.

Reliability of FRT B

Results for the test-retest procedure for FRT B are given in Tables 12.6 to 12.10 inclusive.

It is with this test that the problem of restricted variability in scores is really made manifest. In a 30-item test, the range of scores is 9 points (20 to 29 at the test; 21 to 30 at the retest), and the distribution is not only severely, negatively skewed, but has a standard deviation as low as 2.27. It is hardly surprising that the test-retest correlation coefficients are all unacceptably low - and, in one case, negative (Table 12.6).

Further, the internal consistency correlations are also generally low, particularly in subtests with fewer items and hence a greater restriction on the range of scores possible.

The agreement coefficients (Table 12.8) are unacceptably low for the cutting scores for which a value can be obtained.

A strong practice effect is also in evidence for this test. Almost twice as many response changes were to 'correctness' compared to 'incorrectness' (Table 12.9) and a massive 72.72% of testees improved their total score by one or more points.(Table 12.10).

The measures used to estimate the reliability of FRT B give values which do not allow the retention of the test for testing. It is too unreliable, insofar as these measures can indicate; and it is that indication that may be at fault. Restricted variability will lower the calculated value of a product-moment correlation coefficient, and the test-retest coefficient obtained may not reflect the true strength of the relationship. Most of the agreement coefficients were indeterminate in value (a problem which occurs if either of the main diagonal cells has no members - see Figure 12.1), whilst it is clear that the test <u>is</u> consistent in assigning testees to mastery states for the cutting scores 1 to 20. It is the fact that there are no 'non-masters' which affects the calculations. It is little wonder, also, that the two agreement coefficients have low values when there is a strong practice effect as well as a restricted range of scores. For example, using the test mean (rounded) as 25 as the cutting score, whereas half the testees scored above this value in the test, in the retest the new mean of 26 shows that <u>more</u> than half have scored above 25. Thus, the proportion of cases in the off-diagonal cells (see Figure 12.1) has increased and the agreement decreased. This restricted range of scores may not be the same for pupils of lower ability, from whom different estimates of reliability might be obtained. It seems likely, therefore, that the test may suffer from homoscedasticity.

The measure that does indicate a high level of reliability for the test is the analysis of response consistency. In all, 87.2% of all responses to items were the same on both occasions (Table 12.9). This shows that Form B is probably a reliable test in terms of pupil-item-responses, but this is, firstly, insufficient in itself to allow its retention for use; and secondly, it is irrelevant to the question of decision-making consistency, for most testees could respond similarly to most items and yet still change mastery states by one or two points, depending on the cutting score.

FRT B seems worthy of further study, particularly its administration to groups of different ability ranges to test for homoscedasticity. Unfortunately, this study could not be undertaken at the time, due to the need to proceed to the next stages of the research. Form B was not the subject of further use, it not having been demonstrated to be of sufficient reliability.

Investigation of Practice Effect

Following earlier comments on the practice effects associated with the two functional reading tests, it was decided to undertake some empirical investigations of the effects, to see if some form of practice test was required to introduce novel item-types before testing proper. The observed form of the effect was that of increased total scores for the majority of testees and the preponderance of changes of responses being toward 'correctness'. It is the measure of changes in response (discrepancies) which is the more important, for the sum of response changes is greater than the sum of the changes in totals for both tests, indicating that, for some pupils, the same total is made up of different items on the two occasions.

There are likely to be two major factors in the change in responses in the test-retest procedure: increased familiarity with the item-types and administrative procedures, and the increased amount of time available to reflect on the reading task. If a testee is confronted with an item, he is more likely to answer incorrectly if the item is of an unfamiliar type. A test-retest procedure, in this case, may perform a teaching function, so that on the retest, the item-type is no longer unfamiliar and he is more certain how to undertake the task involved. This will presumably be true for all similar items in the test and he may therefore be expected to improve his score. On a third occasion of testing, however, the changes in response that have been observed are due, not to increasing familiarity with the item-type, but to the increase in time available in which to reflect and evaluate possible answers, then one might expect the testee to improve his score on both a retest and a third occasion of testing. For both of these factors, we may assume that other factors which contribute to changes in response from 'correct' to 'incorrect' remain constant for all three occasions of testing. (This may not be completely true, for repeated testing on the same materials may produce reaction against the test and responses on a random basis; but factors such as carelessness, ambiguity in items, etc. should hold steady whichever testing it is).

From the discussions above, some hypotheses may be formulated. It is convenient to take the second factor, time on test, as the formulation of a null hypothesis. This may be stated as

(i) there will be no difference in the amount of increase in total scores and change of responses towards correctness between testing occasions 1 and 2 and testing occasions 2 and 3.

The other major factor, increased procedural familiarity, is a disproof of the above hypothesis, that

(ii) The increase in total scores and change of responses towards correctness will be greater for testing occasions 1 and 2 than occasions 2 and 3.

In order to investigate the practice effect, a number of 4th Form pupils in another Sheffield secondary school were administered the FRT A on three separate occasions. 4th Form pupils were used as testing was undertaken in the Summer Term, 1980 and 5th Form pupils were therefore taking public examinations. The age difference was thought to make little difference, however, as the testees would be 5th Form pupils within three months. The school had offered its help previously, and as it was located in an area of no particular socio-economic or industrial singularity, it was felt that testees were unlikely to have any especially biased characteristics. Two careers guidance groups formed the basis of the sample and testing was undertaken at three to four week intervals by the Project supervisor, who has kindly made the results available. In all, 31 pupils took the test on all three occasions.

The means and standard deviations for these groups were 21.03, 22.00, 22.90 and 4.11, 4.54, 4.27 respectively and thus show little difference from the mean and standard deviation of the previous sample tested in the reliability study.

The measures under scrutiny here were the sum of scores (the sum of the number of correct responses made by the testees) and the responses changes over the three occasions. These are summarised in Tables 12.11 and 12.12. It can be seen that the increase in the sum of scores is greater between testing occasions 1 and 2 than between occasions 2 and 3, but the actual difference, 2, out of a total of nearly 700 responses is by no means significant. On the other hand, the changes in responses (Table 12.12) does seem to indicate that the expected increases have taken place. The changes from Incorrect to Correct and vice versa have declined in their relative proportions of responses for testing occasions 2 and 3 compared to occasions 1 and 2. The decline may not be significant, however, and this was investigated, using

Since some change was expected from each retest occasion, it seems reasonable to use the mean changes as the expected values in the computation, for we are not concerned with the changes themselves but between which occasions they occurred.. Using the data from Table 12.12, no significant result was obtained (χ^2 = 11.5, df = 7) and from this one must conclude that increased procedural familiarity is not a significant factor in the increase in scores between test and retest, and that no practice test is called for.

Conclusion

In conclusion, one may summarise a very complex topic by saying that restricted variability is a particular problem in the estimation of the reliability of criterion-referenced tests. The use of test-retest and internal consistency methods are called for, but in forms adapted to their present context. A number of methods have been used in this study with varying degrees of success.

The two functional reading tests, A and B were investigated for reliability and, whilst FRT A was found to have an acceptable level of reliability, this was not the case for FRT B. It was concluded that FRT B might, in fact, be reliable due to the evidence of consistent or improved response patterns, but that this evidence alone was insufficient to support the retention of the test for further use.

A strong practice effect was demonstrated and the role of administrative procedures in this was investigated. It was concluded that no practice test was needed to introduce novel item-types before the testing proper.

CHAPTER 13

PREDICTIVE VALIDITY: MEASUREMENT CONSIDERATIONS

The development of a test of occupational functional reading ability is predicated upon the assumption that, in some way, the measurements obtained from its use are positively related to adequate performance of job-related reading tasks. If no such relationship exists, there is little point in developing such a test. Further, because of the temporal difference between testing and job performance, that relationship, if existing, may be considered one of prediction: the use of test scores to predict the level of performance. One expects a high functional reading test score to be associated with, and predict, high performance measures on job-related reading tasks.

The picture is not so simple, however. School-leavers, like everyone else in the labour market, enter different sorts of jobs which have different entrance requirements and different tasks involved. Many such jobs have requirements based on academic standards or judgemental assessments of general ability, and these are largely determined by the content of the job. Thus, insurance companies tend to employ school leavers of proven academic achievement (usually only those with four or more O-level passes), on the basis that the work demands abilities associated with that level of achievement. Similarly, van boys are recruited with little consideration of general ability as the job is undemanding. Non-academic criteria, such as reliability, perseverence, appearance and manners take on much greater weight. It follows, therefore, that if different sorts of school-leaver are taken on into jobs with different requirements, it is not valid to compare, say, a highly intelligent child doing a difficult job with a moderately able child doing an easy job, or with a less able child doing a moderately difficult job. One must compare like with like; to do so suggests that the job should be held constant (e.g. all apprentices) and then any investigation of a predictive relationship should proceed using the test scores of school-leavers and measures of performance on that job. Similarly, this should be undertaken for each job-type.

This brings us to a consideration of the measures of performance on job-related reading tasks. Ideally, one requires a technique for observing and assessing the performance on such tasks of a number of school-leavers, for whom functional reading test scores are also available. This suggests that either another test be used or some set of objective criteria be developed and applied over the first six weeks of the school-leaver's employment. The former represents a circular problem, as one would need to validate such a test by the same procedures. The latter implies that every working moment of a number of school-leavers would be under observation, or that some objective timed observation schedule be used. There are many objections to this, however, not least of which is the large amount of time needed for one person to undertake the observation versus the limit of six weeks in which to do so. The major objection is the disruption such an investigation would mean. In that initial period, everyone involved in the induction, training and supervision of a school-leaver, including the leaver himself, are heavily committed and very unlikely to welcome or tolerate the prescence of an observer. (A similar proposal was discussed with one company, to provide a detailed analysis of the reading tasks of engineering apprentices as part of the early fieldwork. A courteous refusal was received, on the grounds given above. This was from a firm otherwise prepared to give every possible assistance; less co-operative firms could certainly not be expected to be any more helpful, and probably less courteous!) The same objection applies to the possibility of training supervisors to undertake the observations themselves.

As it is necessary to have some assessment of performance, other methods must be considered. The most convenient method would seem to be a rating scale of some simplicity, requiring little time to complete by a supervisor, but phrased to elicit as much useful information as possible. Super and Crites (1962) write that "Ratings of performance are a widely used criterion, probably the most common because of the relative ease of obtaining them. The history of ratings has, however, been extremely disappointing, and when they are relied upon today it should be only because of inability to find or devise a better criterion and after systematic steps have been taken to make them as reliable as possible" (p. 38). It would be unfair to suggest that the use of rating scales indicated in this study for reasons of necessity rather than choice. Although other means would be preferable, the judgemental assessment of performance by supervisors is the very assessment that a firm is most likely to make itself, and the one which is most familiar and best undertaken by the supervisors. Ghiselli reports that "Experience indicates that criteria of occupational success generally have a reasonably substantial degree of reliability, although the reliability is by no means perfect" (1966, p. 24). He mentions that test-retest correlations of ratings tend to lie in the range 0.7 to 0.8. Supervisors' ratings have been widely used and it seems at least justifiable to continue that use here. In a related area, selection testing for engineering apprentices, Frisby, Vincent and Lancashire (1959) used supervisors' ratings, although they, too, were dissatisfied that no other measure of on-the-job performance was available.

Sticht et al (1972) used a Supervisor's Questionnaire, consisting of five or six point scales on 14 attributes such as conduct, job performance, co-operativeness etc.; plus twelve items such as "Does he need more supervision on the job than most?". They report, however, negligable correlations of reading performance with these scales, for the four categories of occupation with which they were dealing. Their scales, however, were aimed at assessing job proficiency, rather than specifically tasks involving reading, and it is a reasonable assumption to expect that ratings of this latter will be related to functional reading test scores. In jobs with high reading requirements in the initial period of employment, one would expect testees with high scores to do well in the job as a whole and, conversely, for jobs with low requirements, there to be no relationship between test score and an assessment of the whole job. Information to substantiate this view would be useful to compare with Sticht's findings, and it would seem worthwhile to ask for a rating not only on performance in tasks involving reading but also on tasks not involving reading.

One expects the prediction of performance in job-related reading tasks from functional reading test scores to be linear, within job types. If the test items have been correctly developed to reflect, insofar as possible, typical job-related reading tasks, there should be a directly proportional relationship between the two. Perhaps this view is too simple, however. The use of Supervisors' Ratings allows one to suggest that the supervisors may themselves introduce some intervening or modifying variable into the relationship. For instance, a 'halo' effect may occur, where the perception of overall adequate job performance may lead the supervisor to neglect specific performance deficits in reading. The reverse effect could also occur. Time on the job may also be a factor. Although ideally, rating should be at the same time (the end of the initial six weeks of employment) for each testee, it would be beyond the control of the researcher to ensure this condition was met. Age of the testee may affect the relationship, in that older testees may be perceived as more mature and/or given higher-level tasks to undertake. Situational factors in the job market may introduce biases in the number and quality of entrants into various jobs and industries.

All in all, then, whilst a linear prediction seems to have face validity, at least in the first instance, it will be necessary to consider the possibility of increasing the quality of prediction via a multi-factor approach. Of course, such an approach has technical and operational problems of its own, not least of which include the questions of additivity and linearity of the factors, and the interpretation of test scores.

This presupposes the actual measurement of the relationship between performance and test scores, and the statistical problems involved. The recurrent problems of restricted variability and skew of scores again need consideration. Linear regression of one variable on another is given by

$$\tilde{Y} = \rho . \frac{\sigma_Y}{\sigma_X} . X + \mu_Y - \rho . \frac{\sigma_Y}{\sigma_X} . \mu_X$$

where $\tilde{Y}$ is the predicted value of variable Y

ρ is the correlation coefficient for variables X and Y

σ_Y and σ_X are the standard deviations of Y and X respectively, and

μ_Y and μ_X are the means of Y and X respectively.

Thus, restricted variability of either variable may produce valueless results. The correlation coefficient, as we have seen, may tend to be smaller than the size of the actual relationship. The ratio $\sigma_Y : \sigma_X$ will be seriously affected as either value approaches zero; for, if σ_X is very small, X will approach μ_X and $\tilde{Y}$ approach μ_Y; and if σ_Y is very small, $\tilde{Y}$ again will approach μ_Y (as σ_Y/σ_X nears zero). Whilst it is obvious that, in the light of restricted information about the nature of X, the mean of Y is the best estimate for $\tilde{Y}$, this is of very limited value in establishing the predictive validity of the test.

Severe skew or kurtosis will also affect the prediction, for the further from a normal distribution, the less acceptable is the parametric model of regression. Distribution-free methods would be called for and, as we are concerned with the relative ordering of subjects rather than the reproducibility of scores, non-parametric methods using ranking may be used in the way not acceptable in the test-retest estimate of reliability. Linear, parametric regression techniques may be used, therefore, but should variability be restricted, or approximation to a normal distribution of test scores or ratings be impossible, non-parametric methods of correlation may be substituted.

A multi-factor approach suggests a step-wise multiple-regression analysis, bringing each factor in turn into the regression equation in order to estimate its contribution in terms of variance explained. At once, it is clear that the caveats discussed above will apply in the multiple regression case. Moreover, a number of analyses are required to delineate the interrelationships of the factors themselves, to see whether the assumption of additivity can be met if necessary, to postulate (weak) causal ordering if there are intervening variables and to see if the relationships are linear or not. Non-parametric methods will not give this range of analysis, but they will certainly allow an investigation of the interrelationship of factors and their individual correlations with the predicted variable (performance on job-related reading tasks). Importantly, whatever form of analysis is undertaken, one will be able to discover which factors are the best predictors and whether test score is a sufficiently good predictor, on its own and compared with other predictors. It would not do for the test score to be correlated with supervisors' rating at about 0.3 and age at rating to be correlated 0.9! There would seem little point in testing if this were the case.

In conclusion then, the establishing of the predictive validity of the functional reading test requires there to be some outside performance criterion, which forms the predicted variable. The criterion most useful and available is the rating by the testee's immediate supervisor of his performance on job-related reading tasks and, for comparison, on tasks not involving reading. The measurement of the relationship has particular problems related to the statistical methods employed and non-parametric estimates may be required. Further, one must enter the field with one's eyes open to the reality of the matter in hand. Correlation coefficients, both parametric and non-parametric, do not appear to have had values in excess of 0.6 in studies of the predictive validity of tests, in a range of circumstances (e.g. Super & Crites (1962), Ghiselli (1966), Williams and Boreham (1972)) and there is little reason to suppose that either situational or technical factors are likely to make the measure of relationship between test scores and performance in the present study substantially in excess of this.

The validation requires a number of school-pupils to take the functional reading test and then some or all of them to be followed up on obtaining employment. Their supervisors should then be asked to provide assessments of their performance. These two aspects, testing and rating, form the next two chapters.

CHAPTER 14

SWEEP TESTING AND ANALYSIS

Introduction

Sweep testing, the administration of the functional reading test to a large sample of the target population, was undertaken primarily for the purpose of establishing the test's predictive validity, discussed in the previous chapter. There are a number of other things that sweep testing enables one to do, however, and the opportunity to undertake other valuable analyses should not be overlooked.

Perhaps the most important of these other analyses is the chance to make some measurement of the concurrent validity of the test. One expects occupational functional reading ability to be quite highly related to general reading ability, but not perfectly, as the latter lacks the specific application of the former. It is also reasonable to expect that general reading ability is a predictor of performance in job-related reading tasks, but again not as good a predictor as occupational functional reading ability. Sticht (1975) found correlations of between 0.65 to 0.80 between a standardised reading test and job-related reading task tests and it seems likely that that may be the case here. If so, it would provide evidence for the concurrent validity (i.e. its validity in relation to other similar measures). To this end, the opportunity of sweep testing may be used to administer a standardised, norm-referenced test of general reading ability as well as the functional reading test, and the total scores of the two tests correlated. The standardised test scores may also be used as a predictor variable in the predictive validation of the functional reading test, for purposes of comparison.

Measures previously established may be reinvestigated using the larger number of cases that sweep testing provides. The internal consistency of FRT A (the correlation of the subtest totals with the whole test total) is one such measure. Further, testing on a larger scale enables the developer to look at the performance of different ability ranges and to consider the distribution of scores on the test and subtests. Bearing in mind the comments throughout this work on the problems associated with restricted variability, the more pupils tested the better, for this will tend to provide greater variance in scores and hence better estimates of the test's characteristics: or, failing this, a clearer description of the score distributions will enable the test developer to use statistical tests and reporting appropriately. As the previous uses of FRT A on any large scale (in the reliability study, Chapter 12 above), have shown, the distributions are likely to be severely skewed and leptokurtic. Hence, normality may not be assumed and non-parametric tests might be called for, along with reports of modality and range.

Allied to the variability of the scores of the group tested is the question of who should be tested, for the abilities of the pupils tested will undoubtedly affect the score variability. The target population for the study has been previously defined (Chapter 10, above) as the non-remedial Fifth Form pupils in Sheffield secondary schools. It was noted that, strictly speaking, those who would not be leaving at the end of the academic year should also be excluded as they were not 16+ year old school-leavers, but that this was not feasible as the decision to stay or to leave is often delayed to the end of the year and, even then, often dependent upon examination results not available until the August after leaving. In fact, the inclusion of probable 'non-leavers' is of positive value at this stage of the work. It helps to delineate the score ranges on the two reading tests of those who do leave and those who do not. It also allows a consideration of the ability of the tests to predict leaving versus non-leaving. This can also be taken slightly further and an investigation into the predictive power of the tests of the status of testees in terms of staying on at school, entering full-time further education, gaining employment and not gaining employment.

It may also prove instructive to look at sex differences on the tests. There is some evidence to suggest that, early in life, the language ability of girls is higher than that of boys (e.g. Cameron, Livson & Bayley (1967); Kagan (1971); Moore (1967)), but little to suggest that this continues to the 16 year old age range. Should it be the case that the differences persist and are demonstrated in the test scores, this may provide explanatory evidence as to the reasons for the biased recruitment into certain occupations of girls rather than boys, and vice versa. It is to be noted, however, that other characteristics, such as careful working, are associated with clerical occupations and with higher test scores and it may be that such characteristics form an underlying factor in any sex differences both in scores and at work.

Sampling the population

As in most research, it proved impossible, for reasons of finance and time, to test the entire population and only a sample was studied. The size of this same was set at approximately 500 pupils. This number was partly determined by the cost of the production of functional reading tests and the purchase of an equal number of norm-referenced tests, and by the consideration of the following factors:

(a) the time available for testing, marking and following pupils up after their leaving school;

(b) the proportion of the sample who would actually leave school in the same academic year as the testing;

(c) the number of subjects needed to obtain as representative sample of the different areas of the city, and ability ranges as possible.

As will be discussed below, the question of available time was radically affected by situational factors, but it was considered, during the planning stage, that complete testing, marking, analysis and follow-up would require at least eight or nine months for a sample of 500, particularly as each testee would take two tests and therefore require two visits. This period of validation was at least as long as was available and certainly no more could be tested that the 500.

Statistics obtained from the Careers Service (Table 5.9) show that about five-eighths of the Fifth Form year group across the city who leave school, tend to find work by September of the same year. This has important implications for the predictive validation. The functional reading test was developed to assess levels of job-related reading performance and the predictive validation must therefore be primarily concerned with employment, rather than unemployment or further, full-time education. Further, a number of those finding employment may not be traceable and, in other cases, the testee or employer may not agree to participate in the research. A certain amount of non-response must be expected. There is no way of controlling, a priori, the "employability" of subjects in the sample, nor the responsivity of employers to the follow-up. The only method available to this study to ensure an acceptable - and useful - number of successful follow-ups is to take a sample of sufficient size to make allowances for the attrition discussed above. The number needed for the employment follow-up was estimated at approximately 200 subjects. A sample of 500 pupils was estimated to yield 400+ leavers, of whom 250+ might reasonably be expected to have found employment by September of the leaving year. Non-response was expected to cut this number to about the target figure of 200 subjects.

The third factor in sampling, the representative nature of the sample in terms of area and ability, was also of importance. In order to minimise any bias in the follow-up, it was necessary to start with as unbiased a sample as possible. A school year group is approximately 7500 pupils in Sheffield and a sample of less than 500 would seem to be no more than "scraping the surface" of that year group.

There are thirty-nine secondary schools with Fifth Forms within the area of the Local Education Authority. A sample of 500 pupils could be obtained by taking about thirteen pupils from each school. This would require a great deal of time in terms of test administration and it was decided to select a number of schools and obtain larger groups of pupils from within them. The main considerations here were to obtain a sample representing all geographical areas of the city contributing to the employment of school-leavers, and to ensure a range of ability.

The argument may be stated in this way: Sheffield is a large city, with a population in excess of half a million, and contains a large number of industries, and a large number of areas differentiated by socio-economic status and the type of work available locally. Despite excellent public transport facilities, the inhabitants of the city still attempt to work fairly close to home and this is also true for school-leavers (Sheffield Careers Service, personal communication). Hence, there is geographical differentiation in the type of work undertaken by school-leavers. It was therefore considered appropriate to divide the city into six major areas on an 'industrial-cachement' basis, these areas being conveniently bounded by major roads. Sampling within each area would produce a sample less biased by regional differences.

The situation was not quite so simple, however, as a number of schools were geographically remote from the city, whilst still administratively within its boundaries. These were excluded from the sampling frame, as they tended not to provide school-leavers for employment for Sheffield, but for other neighbouring areas. Further, four schools were heavily committed to involvement in an E.E.C. Project and, after consultation with the LEA Senior Advisor for Research and Evaluation, were also excluded from the sampling frame. They were considered to be unrepresentative of schools in their areas and might contribute unwaranted bias to the sample; also their staff were thought to be sufficiently involved in research, without asking them for further help. Lastly, schools used in the pilot study were not used again.

Twenty-nine schools remained to be sampled, and a sample size of eight schools thought appropriate, being one from each of the geographical areas and one extra for each of the two most populus areas. This indicated that about 50 pupils from each school would be required.

It is a researcher's first concern, ethically, to ensure his work is undertaken in such a way as to minimise any detriment to his subjects. This is particularly true when the subjects are Fifth Form pupils, most of whom are candidates for public examinations. In this case, the concern of the researcher was to minimise the disruption of the individual's school work, and also the disruption within the school. A random sample of pupils across the year group inevitably means that the entire year group is disrupted. Using the school's usual administrative or curricular groupings, however, ensures that only the actual testees are inconvenienced and, providing the timing within the school year is right, this will also be minimised. Such groupings are rarely across the ability range, however. Three ability groups were therefore defined, and a medium ability group, plus one other, decided on a random basis, were requested from each school. That is, one school might be asked for a group of high ability pupils and a group of medium ability pupils, whilst another might be asked for low and medium ability pupils.

These ability groups were:

(a) High ability: those pupils predominantly taking GCE O-level examinations in several subjects;

(b) Medium ability: those pupils predominantly taking CSE examinations in several subjects;

(c) Low ability: those taking few examinations, or only restricted grade examinations, or none at all, non-remedial.

In this way, all ability groups were represented, with middle ability tending to be heavily to the fore, as this was the group for whom the test was primarily designed.

The sample of 500 pupils was achieved, therefore, by using the school as the initial sampling unit, excluding inappropriate schools, randomly sampling schools within geographical areas and then sampling ability groups within each sample school. One school was completely mixed ability in all areas and so were asked to provide any two groups.

Periods of test administration

Headteachers of the eight sample schools were contacted by letter (Appendix III, Letter 11) and a discussion with the researcher requested. One school refused to consider involvement, however, due to the presence of testers from the Assessment of Performance Unit at that time. All remaining schools agreed to assist, but a number with reservations. These were due to the timing of the testing period. It had been hoped to complete pilot work and analysis by early Spring 1979 and then proceed straight to sweep testing. The winter of 1978-1979, however, was extremely severe in Sheffield and, combined with a continuing industrial action during the Spring Term by public employees, meant that the researcher was forced to request the administration of the tests at the beginning of the Summer Term. There were a number of well-founded objections to this, primarily that this clashed with the period of examinations and, secondly that the severe winter and school closures had cost pupils sufficient disruption without the school agreeing to even more.

A number of staggered testing periods were agreed, involving pupils from two academic years. Four schools agreed to assist in the first three weeks of the Summer Term, 1980 (and also, in one case, at the end of that Term), whilst the remaining schools agreed to participate in the following Spring Term, sampling being from their next Fifth Form.

Test administration

Each pupil tested was to take two tests: the Functional Reading Test, Form A (FRT A) and the Edinburgh Reading Test, Form 4 (ERT 4). This latter was chosen as it was standardised to 16.0 years, and, being published in November 1977, was up-to-date in its language. Also, it had a number of subtests corresponding to some aspects of the FRT A. It was recognised that a number of pupils would be over 16.0 and that the tabulated quotients for the ERT 4 would be inappropriate. A linear regression extrapolation of quotients for ages 16.1 to 17.0 was undertaken, using the tabulated quotients as the data-base. These quotients are given as Appendix VIII. Raw scores have been used for the subtests, however, as no other data was available for them.

It had been hoped to have a standard gap of three weeks between the two tests, but this proved impossible in almost all cases. Due to the timing of the periods of administration, it became very much a matter of testing whenever the pupils were made available, often before morning break and then immediately afterwards. Not all schools provided the groups all together, meaning several visits to test different pupils. Further, one school offered the entire, non-remedial Fifth Form as subjects, rather than just 50. This offer was accepted as it relieved the necessity of replacing the school refusing to participate.

A copy of the FRT A and its manual of administration are given as Appendices VII and VI. Naturally, with two tests to be taken, the groups tested were not always identically constituted on each occasion. For a number of pupils, therefore, only one set of the test results were available. Pupils who did not take the ERT 4 also gave no date of birth.

The number of testees in each school, by sex, is given in Table 14.1.

The number in each ability range, by school, is given in Table 14.2.

In all, 471 pupils were involved in the sweep testing: 428 took the FRT A, 417 pupils took the ERT 4 and 375 pupils took both. All testing, marking and reporting of results was undertaken by the author. Pupils from School 7 (upper ability group) took a version of FRT A which used a separate answer-sheet. Apart from initial administrative difficulties, their performance did not seem affected by this. In fact, they had the highest mean score (see below).

Sweep Test Analysis

As mentioned above, the data obtained from sweep testing is not only useful for the predictive validation but for several other purposes. The remainder of this chapter will be concerned with these other aspects, whilst the following chapter (Chapter 15) deals exclusively with the predictive validation based on job performance measures, and the prediction of employment/training status.

Those other aspects may be considered as firstly, the performance of the FRT A, including its subtests; the concurrent validity of the FRT A; and the performance of the testees on the two tests, including sex differences.

Performance of FRT A

Summary statistics for the FRT A are shown in Table 14.3. The severe negative skew and leptokurtosis are apparent, as expected from the distribution of scores on previous uses of the test, and the skew is also shown in Figure 14.1.

The frequency distributions of the whole test and each subtest are given in Tables 14.4 to 14.17.

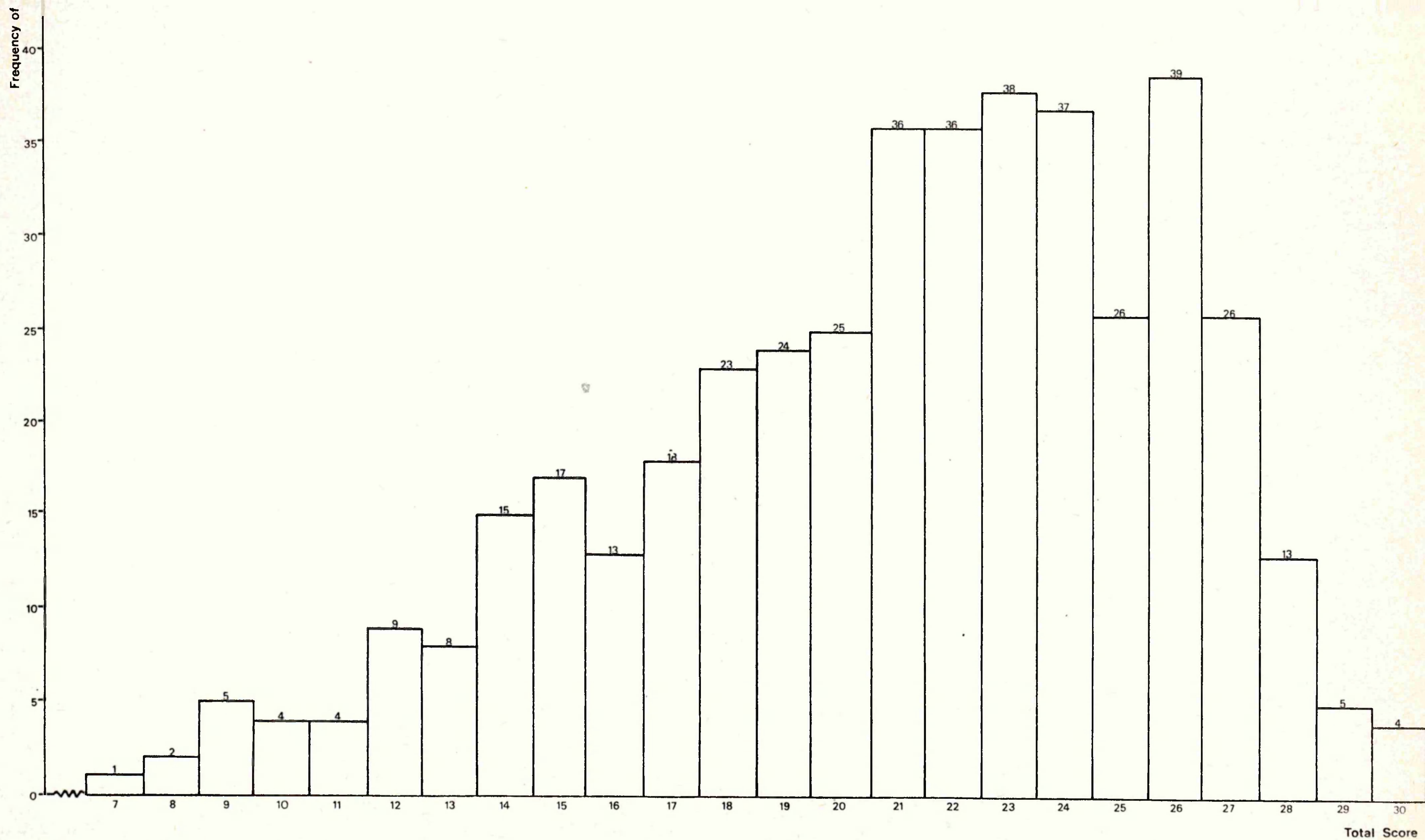


Figure 14.1: Frequency Distribution of Scores : FRT A

The measures of mean and median (21.07 and 21.8 respectively) are fairly close together for the whole test, indicating that the mean is still a useful measure of central tendency. Despite the skew, the standard deviation of 4.8 shows that there is indeed a fair degree of variability, although the top third of possible scores have two-thirds of the cases.

The majority of the subtests show similar distributions and it is therefore of little surprise that the intercorrelations of subtest totals with whole test totals are in general high, as shown in Table 14.18. This table also shows relatively low (but still significant) correlations between subtests of the same type (i.e. content, job and linguistic task), demonstrating to some degree that the individual subtests are measuring the separate categories for which they were designed.

It is not proposed to discuss individual test items nor to report individual statistics for them. Certain facts did emerge, however. The position of an item had some effect on the omission rate, in that the two or three items at the end of each section had much higher rates than other items. The level of omission was as high as 10% for lower ability groups and it may be that two or three minutes need to be added to the response time, for each section. No item had a negative discrimination index, either within the whole test or its individual subtests. Difficulty levels ranged from 18% to 93 % correct (the mean difficulty is, of course, the group test mean of 68.9%).

Concurrent validity

Summary statistics of the ERT 4 are given in Table 14.19.

The correlations of the FRT A total scores with the two total score measures of the ERT 4 (raw score and quotient) are given in

Table 14.20. These correlations are highly significant and indicate that the two tests are measuring within the same field. It is to be noted that the correlation of the FRT A with itself is not much in excess of these values (0.86, see Chapter 12), suggesting that many of the properties measured by the ERT 4 and that, in fact, the former may not be a better predictor of job performance than the latter, but only as good as a predictor. It may be concluded, however, that the FRT A has a high degree of concurrent validity.

There are a number of other interesting relationships one may investigate under the heading of concurrent validity. Just as the FRT A total score is based on the sum of subtest scores (and the assumption of a strong relationship between all subtests and the total), so too is the ERT 4 raw score and quotient based on their relations with the five subtests within the test. Table 14.21 shows the inter-relationships for the ERT 4, based on the scores obtained in sweep testing, rather than those in the test manual. It follows then, that if total measures are intercorrelated, so should relevant subtest measures also be intercorrelated. In particular, one is looking at the content types and the linguistic tasks in the FRT A (job-types are not relevant to the ERT 4) and the comprehension, vocabulary, reading for facts and skimming subtests in the ERT 4 (points of view would have been relevant had any attitudinal linguistic tasks been included in the FRT A). The following investigations were undertaken therefore:

(i) the measurement of the correlation between the content subtests and the comprehension subtest (based on the assumption that comprehension is an underlying skill in using any written material);

(ii) the measurement of the correlation between the referential linguistic tasks and the reading for facts and skimming subtest (as these all involve searching for information within reading passages);

(iii) the measurement of the relationship between the definitional linguistic task and the vocabulary subtest (as these both involve stating the meaning of a word of a phrase).

The size and significance of the correlations for the first two of these investigations are given in Tables 14.22 and 14.23 respectively. Again, the concurrent validity of the subtests has been established. The relationship between the vocabulary and definitional subtests was investigated by crosstabulation and resulted in a significant result ($\chi^2 = 56.2$; df = 32; $p < 0.005$; contingency coefficient = 0.36).

Pupil Performance on FRT A and ERT 4

Figure 14.1 has already shown the distribution of total scores on the FRT A to be severely negatively skewed and this pattern is generally true for all the groups tested. Table 14.24 gives a breakdown for each school and each sex and it can be seen that the mean scores for each group are usually quite close to the total mean score. School 5 (Lower ability) is an exception, however, as this group contained a number of pupils whose ERT 4 scores gave reading quotients of less than 75. In general, it is not proposed to consider differences between schools, as this would be invidious in the light of the deliberate requests for specific ability groups. What will be more useful is a consideration of ability-groups in relation to FRT A scores.

The ability-groups defined above were useful in ensuring that pupils of different levels of general school performance were included in the sample. A more accurate measure of general reading ability is now available, however, the ERT 4 quotients, and it therefore would be instructive to look at FRT A performance for different ranges of quotients on the ERT 4. Figure 14.2 shows the distribution of ERT 4 quotients in the sample compared with the general normal distribution. Due to the sample's bias to the middle-ability range, the standard deviation of the quotients is less than the normal curve, giving the leptokurtosis visible. The two means are, however, virtually

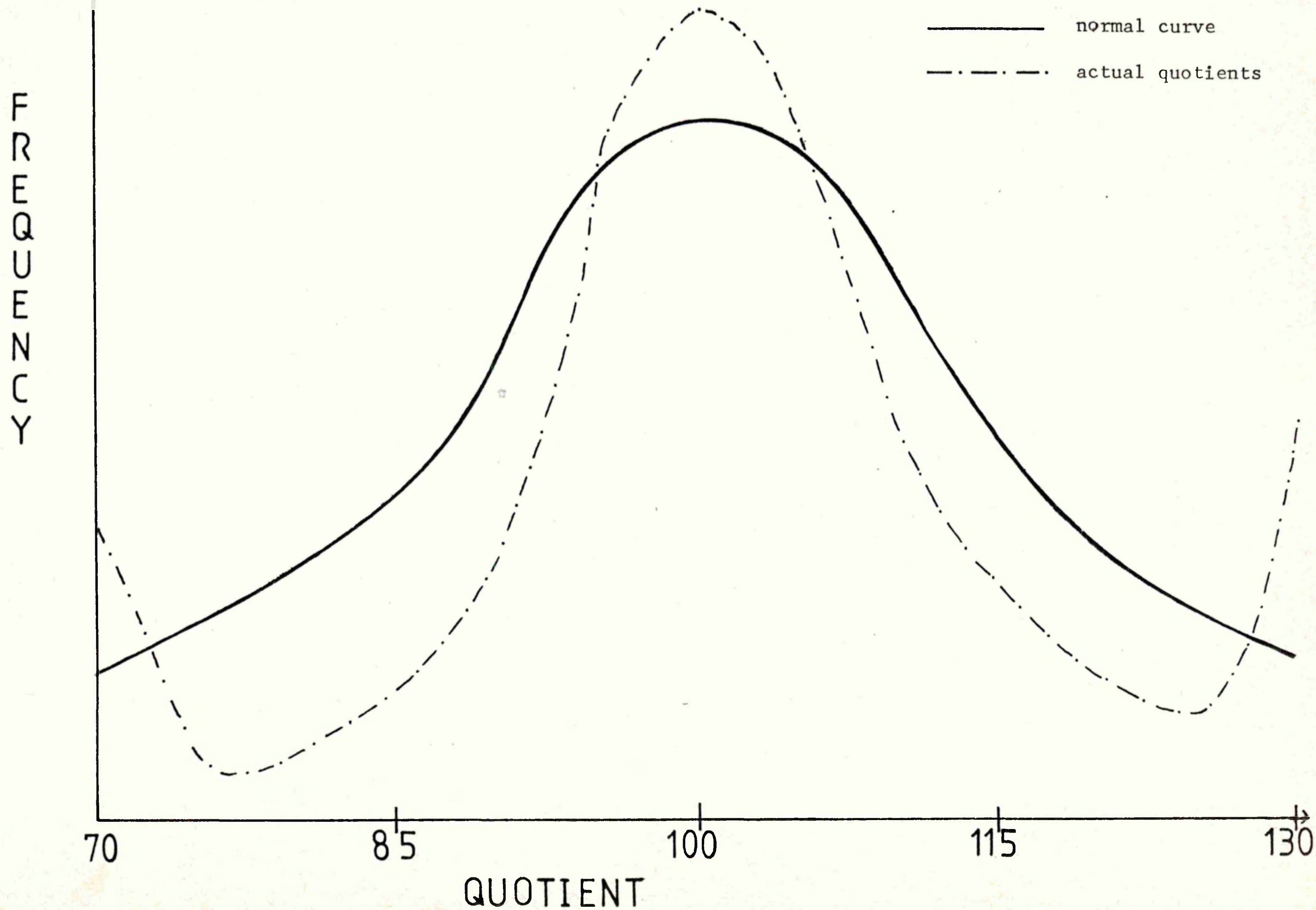


Figure 14.2: Frequency Distribution of ERT 4 quotients

identical. Floor and ceiling effects are due to the extrapolation of quotients described above, as the tabulated quotients from the Manual had no variability for certain raw scores and the mean quotient had to be used. There are several ways of defining reading ability-groups using the ERT 4. One may use standard deviation groups of either the obtained scores or the normal distribution; one may use above and below the mean, or each third of the range. It is common practice to use standard deviation groups, and that convention will be used here. Both the suggested standard deviation groupings are worth pursuing, as the extrapolation of quotients may distort one and not the other. These ability groups will prove useful in a consideration of whether any FRT A scores fall exclusively within one ERT 4 group, and also in the predictive validation.

Table 14.25 gives a breakdown of ERT 4 quotients for the different schools and each sex. Similar ranges to those for the FRT A can be observed.

ERT 4 and FRT A Performance

Table 14.26 shows the FRT A score ranges associated with the theoretical standard deviation groups of the ERT 4 and Table 14.27 those for the actual groups. In each case it is not possible to distinguish any general reading ability score such that it completely divides one FRT A range from all others. It would thus not be possible to use this general reading test in the context of functional reading ability, as the degree of overlap of the FRT A scores shows that attempting to relate a standard deviation group to a functional reading score range would be very unreliable.

Pupil Age and FRT A Scores

Table 14.28 shows the correlations of FRT A Total and subtest scores with pupil ages (in months). Most relationships were not significant, but the FRT A Total is positively related to age at test. The size of the coefficient is extremely small, however, and it is probably true to say that the age of pupils made very little difference to their scores. Spearman correlation coefficients are only fractionally larger. Table 14.29 gives a breakdown of mean ages by school and sex.

Sex differences

Results of significance tests of the difference between mean scores for each sex are reported in Tables 14.30 and 14.31. It can be seen that for the whole test and for most of the subtests in the FRT A, females scored significantly higher than males. This may be related to the only significant result for the ERT 4: the subtest on skimming. It may be suggested that females are not necessarily better at functional reading in general, but perhaps better at finding the answers in the time available.

It is surprising that a norm-referenced test designed to give a normal distribution of scores across all testees should give scores significantly higher on one ot its subtests, for females. This may be an incidental product of the extrapolation, although there were no sex factors involved, or insufficient work in development, or reflect an abnormal trend, or merely be a trend that has resulted from testing a non-normal group.

Conclusion

In conclusion, one may say that large scale use of the FRT A produced no significant deviations from the results given by pilot work. In fact, one may draw satisfaction from the fact that all groups tested, bar one of particularly low ability, obtained very similar means and standard deviations.

The FRT A was demonstrated to have concurrent validity, by correlating test scores with the quotients obtained for the same pupils using the ERT 4.

A number of sex differences in responses were demonstrated but this was linked to one particular searching skill, rather than to some basic higher ability in girls.

The ERT 4 quotients could not separate out specific score ranges on the FRT A, with whatever grouping was used.

There was a small relationship between age at test and FRT A score.

With this brief summary, one may say that the first step in predictive validation - obtaining a large number of scores - had been achieved, and that without revealing any anomalies in the FRT A. From the collection of scores, one passes on to the use of those scores in following up and assessing the performance of the testees on job-related reading tasks.

CHAPTER 15

PREDICTIVE VALIDATION

Introduction

In Chapter 13, the measurement considerations involved in predictive validation of the FRT A were discussed. Two major factors were advanced: obtaining scores on a large number of pupils, and the follow-up of these pupils as they left school and sought employment or training. The first of these has been described in detail in the last Chapter and here the follow-up is the subject.

There are three basic steps to the follow-up: the design of a suitable rating scale, the location of the tested pupils after leaving school and the completion of the rating scale, and the statistical investigation of the predictive relationship. Although conducted over two separate years, as has been explained, the procedures involved were in general identical, and such changes as were made between years will be indicated. In all, however, the data resulting from the sweep testing and follow-up in each year are combined to give a continuous set of data.

Design of a rating-scale

The prime aim of any rating scale is to provide accurate and simple information, and to this end there were two major considerations for the SOFRP. One was, of course, to receive the right information in a suitably objective form; that is, to devise the right questions and response categories. The other factor was to ensure that the scale was easy to complete, as this was to be undertaken by the leaver's immediate supervisor, who could not be expected to be familiar with such exercises.

In the first instance, it is clear that the scale was concerned with rating the subject's performance on job-related reading tasks. There are other factors in performance of such tasks, however, than just reading. Perseverence would be important, as supervisors might rate a subject down for giving up too easily. The subject would, under these circumstances, not be reaching required levels of functional reading. Similarly, the amount of help for which the subject asks would be a direct indication of the level of difficulty he is having with the task. Speed of performance will also be a factor.

Moreover, following the earlier discussion of Sticht et al.'s (1972) work, it was felt valuable to compare ratings not only on tasks involving reading but also on tasks <u>not</u> involving reading. This also allows one to investigate any 'halo' effect by correlating reading and non-reading performance. Finally, a factor of 'oral language fluency' may be introduced as a control. This is included to study the perceptions of the rater of the ability of the subject to express himself orally. One would expect this to affect the perception of other factors, especially if the subject is rated poorly on self-expression.

From the content of the rating, one moves onto the presentation and completion aspects. It was felt that simple five - or seven-point scales with set answer options would be most appropriate for their simplicity. A number of employers who had previously assisted in the SOFRP were contacted by telephone and asked about how their supervisors might respond to the task. In all cases, they expressed the opinion that such a task would cause no difficulty, as their supervisors were generally used to making such assessments about new employees. One employer sent a copy of a 11-point rating schedule used in his company.

The rating scale designed is given in Appendix III, Schedule 4. Five-point scales were used in preference to any others as firstly, they provide a middle position and secondly, they do not involve spuriously fine judgements. Five-point scales were used also because this number of points is common in Likert scaling (Moser & Kalton (1974)). No estimation of the reliability of the scale was made owing to the difficulty in setting up an adequate experimental situation, but each item was made as simple, clear and unambiguous as possible. Administration instructions were made precise and two examples given.

The scales were scored by assigning a value from 1 to 5 according to the point chosen, the higher the value the more positive the rating. Scales were then summed to give a performance total, reading performance total and a non-reading performance total. These were used in the validation along with the single scale for oral fluency.

Locating the School-leavers

Without the interest and co-operation of the Sheffield Careers Service, it would have been virtually impossible to undertake any thorough follow-up of the pupils tested. The Service is responsible both for finding young people employment and for their periods of of unemployment. They have a record of every person about to leave school and should, theoretically, be informed at every stage about their employment 'status' - be it looking for work, returning to school, in permanent or temporary employment, youth opportunities scheme, etc. In practice, the information available is not quite complete at any one time, for a proportion of pupils find their own work on leaving school and have no contact with the Service. Also, a smaller number remain unemployed for fairly short periods of time whilst not claiming any social benefits, and are temporarily 'out of the system'. Further, some school-leavers change their minds and return to school at the last moment, or say they are returning and then seek a job. Nonetheless, the Careers Service attempts and to a large extent succeeds in keeping track of the several thousands of school-leavers each year. Their files therefore represented the most ideal source of information for the follow-up.

Access to these files was kindly granted to the author, under assurance of strict confidentiality. Here, the procedures involved differed between the two years of validation. In the first year, the author visited the Careers Service offices and inspected the records of each testee for whom a file existed, at about three-weekly intervals between July and mid-September (1979). This proved quite laborious and was inconvenient to the Careers Service Officers wishing access to the same files. In the second year, therefore, it was arranged to place a card in each file, for completion by a Careers Officer in the course of normal work, when one of the subjects gained employment or registered on a youth opportunities or work experience course. These cards were then sent to a Senior Careers Officer who forwarded them to the author. A specimen card is given as Appendix III, Schedule 3. In each case, the name of a contact in the employment gained by the subject was requested, in order to make a more direct approach.

The latter arrangement for locating the school-leavers was much easier to administer than the former system and the Senior Careers Officer issued a number of periodic reminders. The author visited the offices at the end of the first week in September (1980) to locate those who were unemployed or were known to be returning to school or entering further education.

At the end of that same period, the headteacher of each school assisting in the SOFRP, in the relevant year, was contacted and asked to specify which subjects had returned to school.

Occupational Destinations

The majority of school leavers in the sample did not find work, and this is shown in Table 15.1. Only 36.8% of the sample were in employment by the close of data collection, two months after the effective school-leaving date. Almost as many stayed in full-time education as went into employment. Rather fewer girls found jobs, but rather more went to colleges of further education. Many of the leavers could not be traced, but it is a reasonable assumption that this group may be divided between the employed and unemployed in the same relative proportions as those traced, for those returning to school or entering college were more easily identified. The proportion known to have entered employment is rather less than expected (see Chapter 14) and is a direct reflection of the increasing difficulties facing school-leavers in the labour market. Of the 173 known to have jobs, specific types of job could be identified for 162 of them and an analysis of these is given in Table 15.2. The pattern was essentially similar to previous years, with boys predominating in the apprenticeships and girls in the clerical field. The preponderance of boys in the operative/others category reflected the poor chances of girls finding employment overall.

Contacting Employers

As each employed leaver was located, his or her employer was contacted after a delay of about two weeks. This delay was in order to obtain a rating nearer the middle or end of the six weeks initial period in which the SOFRP was interested, rather than right at the beginning.

Again, procedures differed in each year. In the first year, the employer received a letter asking him to participate, indicating the nature of the co-operation required and enclosing a specimen rating scale, but not specifying the name of the school leaver of whom the rating was to be made. The author would then telephone to arrange an appointment, discuss the project and if the employer agreed to assist, leave a copy of the rating scale with the name of the subject, and a pre-paid return envelope. As will be discussed below, the response rate was very low and even those employers who agreed to assist by rating their employee were highly resistant to a visit by the researcher as well. In fact, no visits were made, the employers preferring to accept the subject's name over the telephone and return the scale at their own expense. This being the case, a change in procedure was adopted in the second year and all the relevant information, including the subject's name and a return envelope, was sent to the employer without a request for a visit. Telephone contact was made if four weeks elapsed without a reply, to request the form's return.

Return Rate

The return rate of completed forms was just over 30% of employers contacted, as shown in Table 15.3. Sixty-two of the school-leavers known to be in employment at the close of the location exercise had only just started their work and it was not possible to send out forms and secure their return within the given period. Such leavers were often entering jobs conditional upon their examination results and may have constituted a particularly high ability group, for conditional job-offers were found to be commonest for the higher requirement jobs (see Chapter 6).

The actual number of forms was disappointing, as only twenty-eight were fully usable. This had serious effects upon statistical measurement of relationships between the key variables in the validation. Further, the range of responses made was quite limited, with very few expressing dissatisfaction of any sort. Out of a total possible of 45, the mean rating was 37.7 (83.7%), with a range of 22 (48.9%) to 44 (98%) and a mode of 35 (77.8%). Employers are, of course, expected to take on young people whom they think will perform adequately the tasks of their jobs. Also, a strong 'halo' effect can be demonstrated (Table 15.4), of the interrelationship of reading, non-reading and oral fluency measures.

Relationship of Performance to Reading Ability and Other Predictors

All relationships in the predictive validation were investigated non-parametrically owing to the clear non-normality of the FRT A score distribution, and the low number of cases for which complete data was available.

Table 15.5 gives details of the relationships between four predictor variables and the performance measures across job-types.

The Induction subtest total was included as a predictor as it was not a job-specific subtest. Age at test was included as a variable which may have some bearing, as it correlated with FRT A measures (see Table 14.28). All values in the Table are small. and non-significant, as was expected for the combined measures across job-types. It is interesting to note, however, the negative relationship between performance and age. One might expect an older employee to be more mature and thus be rated more highly - a positive relationship. Perhaps employers expect more of older employees. Also of note are the low measures of relationship between general reading ability (ERT 4 quotient) where one might have expected higher values.

Tables 15.6 to 15.9 show similar details for each of the four job-types. For apprentices, the number of scales available meant that no meaningful relationship could be demonstrated. For clerical workers, the values are again all not significant, although one notes that the FRT A values are all slightly higher than their ERT 4 counterparts. The Table for distribution operative/others (15.7) does contain some highly significant values, although the number of cases involved still leads one to treat them with caution. Included in these significant measures are the relationships between both reading test scores and the total performance measure and reading performance total. The strong negative relationship leaves one wondering whether poor reading performance is earning high performance ratings.

For operative/others, the same caveats must apply about the interpretation of coefficients derived from only seven cases. The highly significant and strong positive relationships between reading performance and reading test totals are encouraging, as is the relationship between reading performance and the indcution subtest of the FRT A, as most of the reading material applicable to this job-type is of the induction type.

In conclusion, one must say that the very small number of cases involved for each job-type allows only the statement that the data do not permit any firm conclusion to be made about the predictive validity of the FRT A in terms of total performance, reading, non-reading performance totals and oral fluency, despite the significant values obtained for a number of the relationships.

Employment and FRT A Scores

In lieu of a straightforward relationship between FRT A scores and performance measures, it is of interest to consider whether there exist any links between test score and employment status, and job-type. Such information would be of particular value in the interpretation of test scores. Of course, employers had no knowledge of FRT A scores when deciding to employ the various school-leavers and the relationships, if demonstrated, can only be of association and not causation.

An analysis of the ranges and means of FRT A scores is given for each employment status group in Table 15.10. It can quickly be seen that those returning to school and going on to further education have the highest mean scores, whilst the mean for the employed exceeds that of the unemployed. Differences in mean scores are of little value, however, as they only indicate that the ordering is as expected. They give no information as to whether employment status is linked with specific test performances. Such a relationship can be demonstrated to be highly significant ($\chi^2 = 155.5$; df = 92; $p < 0.00001$; N = 427) but this is insufficient in itself. One needs to know how well test scores predict employment status, and two statistics are available to do this: lambda and the uncertainty coefficient (as part of the output from the Statistical Package for the Social Sciences by which these results were analysed). Both these measures indicate the proportion of uncertainty about the value of one variable that is reduced if one knows the value of the other variable in the relationship. In this case, lambda = 0.12 and the uncertainty coefficient = 0.13, indicating that even when the FRT A score is known, one is still largely unable to predict the employment status of the testee (one remains at least 87% as uncertain as when one did <u>not</u> know the FRT A score).

A similar analysis may be made of the type of job into which testees went. Table 15.11 gives an analysis of ranges and means of FRT A scores, and again one can see an ordering in line with the difficulty of the jobs in reading terms (clerical apprentices distribution operative/others). The overall relationship is not significant on this occasion ($\chi^2 = 104.7$; df = 88; $p < 0.1$; N = 146) but the statistics for prediction are a little higher (lambda = 0.24, uncertainty coefficient = 0.22). Again, however, only a quarter of the uncertainty has gone out of the prediction of job-type.

In all then, the use of the FRT A total score to predict either employment status or job-type has not been shown to be effective. A number of factors may be advanced to explain this. Firstly, as has been suggested, good functional reading ability itself is not a sufficient condition for employment, although it is usually necessary. So, for a taxing job, high ability is important whilst for an untaxing job it is almost irrelevant. Other criteria must be met, such as mathematical ability, presentation, interest, etc. in order to gain employment. Secondly, the labour market is such that even the most employable may not find work, simply by not having been in the 'right place at the right time'. Despite all efforts, it is often the case that some pupils with the best developed skills and highest levels of ability will be missed out whilst those with fewest skills and lowest ability will by chance obtain employment. With a glut of labour compared with vacancies, a slight resemblance to a lottery is easy to perceive.

Further, just such a depression in the labour market is likely to encourage pupils to stay at school who might otherwise not have done so, and push into further education those who might have previously gone straight into employment. Such moves might have arisen from either a realisation of the need for saleable skills obtainable only after further study, the desire to be involved in some activity rather than suffer the boredom of unemployment, or simply the desire for self-improvement.

Similar factors may be seen to operate in the actual type of job entered. Certainly, non-reading criteria are applied to apprentice selection via the battery of tests many applicants must face. Also, in a time of high unemployment amongst school-leavers, no doubt many have taken whatever was offered, rather than a job commensurate with their skills or abilities.

Cutting Scores and the FRT A

With the demonstrated lack of predictive ability of the FRT A total or job-related subtest scores to performance, employment status or job-type, any investigation of the value of a cutting score is of limited application. Yet it may be that what little evidence has been found for a relationship between at least employment status and job-type may be used to see if any particular FRT A score is associated with a maximum decrease in uncertainty.

Such an investigation was carried out by assigning 0 or 1 to a variable depending upon whether an individual's score was less than the cutting score, or equal to or above it respectively, and cross-tabulating that variable with employment status and job-type. This was undertaken for each score in the obtained range of scores (7 to 30). Results are given in Tables 15.12 and 15.13 and it can be seen that the maximum decrease in uncertainty (i.e. the highest level of predictive ability) is 20.0%, for a cutting score of 11, for job-type, and 6% for a cutting score of 24, for employment status. Both of these are actually less than the decrease in uncertainty associated with the raw FRT A score, as the variability of the predictor is decreased from a range of 22 to a range of 1.

This being so, there is no case for the use of a cutting score to determine employment status or job-type. The raw score itself is a better predictor and that only of very limited value. In fact, any grouping of FRT A scores is likely to decrease the amount of prediction compared with the raw score itself.

Similar results were not calculated for any of the performance measures, as the size of relationships and the number of cases involved were too low to justify any conclusions being drawn from the measures.

Conclusion

Of the 470 pupils tested, 173 obtained work within the follow-up period. The employers of 101 of these were contacted and asked to complete a very simple rating form about their performance on reading and non-reading job-related tasks, and their oral fluency. Only 28 completely usable forms were received.

Investigations into the relationships between the predictor variables (FRT A total, ERT 4 quotient, job-specific subtests, Induction subtest and age at test) and the predicted variables (performance total, reading and non-reading totals, oral fluency rating, employment status and job-type) yielded low predictions and generally statistically insignificant results. The use of a cutting score also yielded negative results for some of these relationships.

It must be concluded that the predictive validity of the FRT A in relation to performance at work, employment seeking and in terms of job obtained has not been successfully demonstrated. The implications for the interpretation and uses of the test are discussed in the following Chapter.

CHAPTER 16

TEST EVALUATION & INTERPRETATION

In evaluating any test, the two most important factors are the reliability and validity of the measurements that can be made using the test. One must be able to place one's trust in the score or value achieved by a testee, or at least know the limits of trust associated with such a score. The test must also be testing what it is supposed to be testing. When one has evaluated the test and the measures reach acceptable standards, one can give a meaningful interpretation to any testee's performance on the test.

The main form of reliability for a criterion-referenced test is its stability, or test-retest reproducibility. Testees should achieve very similar scores on two occasions of testing, given that a number of factors may result in small changes to the actual values on the second occasion. In the case of the FRT A, this stability has been demonstrated, even though a noticeable increase in scores across testees was apparent. This 'practice-effect' was investigated and found not to be significantly related to administrative procedures. It was concluded that the increases were probably due to the extra time available to study the questions and test passages.

A second feature of reliability estimation in criterion-referenced measurement is often the measurement of a test's internal consistency - the extent to which each test item reflects the underlying dimension being measured. Here, a rather difference evaluation was made. It was realised that the range of materials and questions used were unlikely to reflect any single dimension, yet it was still desired to use a summative score in test interpretation. This being the case, it was necessary to demonstrate a strong relationship between the total score on each subtest and the total score on the whole test. In both the particular study undertaken for reliability estimation and the later, large-scale testing, such relationships were demonstrated to highly significant levels.

The nature of the distribution of scores obtained upon which to calculate the above estimates means that some caution should be used with other measures, however. The severe negative skew of scores suggests that the standard error of measurement may well be meaningless. As such, it was rather more difficult to estimate confidence limits for scores in the same way as one might for norm-referenced test scores.

Three forms of validation were made of the FRT A: content, concurrent and predictive. The first of these was undertaken with the use of several panels of experts, of both employers and linguists. In this way, no items of irrelevant or dubious content were allowed into the item pool. In fact, the virtually exclusive use of collected rather than devised reading passages was seen as an essential first step in assuring content validity. In this context, panel studies have the added advantage of having members evaluate known quantities, rather than express a rather broad opinion on invented materials. The test itself was not separately validated in this way, but devised from the item pool in such a way as to guarantee inclusion of some items of every type available. The FRT A may be considered, therefore, to be content valid.

The concurrent validity of the FRT A was assessed in relation to the ERT 4. In a sense, it could be argued that this was not strictly concurrent validity, as the two tests were not congruent in aims or objectives. However, if the FRT A was not a valid measure of <u>some</u> reading ability, it would be unlikely to be related to a test of general reading ability. In that the two tests lacked congruency, a less than perfect relationship may be assumed, and the estimation of concurrent validity would be an insufficient measure of validity on its own. The actual measurement, on 374 testees, bears this out. The FRT A may be considered, therefore, to have a degree of concurrent validity.

The previous Chapter has recorded the failure to obtain a satisfactory measure of predictive validity for the FRT A, either in terms of performance at work, employment status or the particular job obtained. Several factors may be suggested to account for this. The first of these is, of course, that there may exist <u>no</u> relationship between test performance and job performance. It may be that there are so many other variables in job performance that to attempt to separate out reading in particular is fruitless. Yet employers have consistently held up literacy skills as one of the two major areas of basic skill necessities for school-leavers, which implies that they are aware of the need for reading in job performance. This in turn implies that they should be aware of reading performance in job performance. This leads one to suggest that there <u>must</u> exist some relationship between functional reading and job performance and that the failure to establish any acceptable value for that relationship is a <u>measurement</u> problem rather than any other.

The measurement in this case has two sides: the reading test and the rating scale. It has been demonstrated that the FRT A had content validity and one may therefore argue that, if any measurement was to be valid, test scores were likely to be so. So one must consider the rating scale side of measurement to be the more likely candidate for scrutiny. There are many points at issue here, not least of which is the economic climate in which the school-leavers were seeking work. It has already been argued that leavers could not be guaranteed a job no matter what level of ability they possessed, nor, if obtaining work, that it was at a level commensurate with their ability. In order to compensate for this, as clear an overall picture of job performance as possible was needed i.e. a large number of completed rating forms.

Quite the opposite actually occurred within the follow-up period: less than 40% of the sample of testees could be traced to any specific employment and less than 30% of those employers contacted returned completely usable forms. There is justifiable concern that such returns as were received did not constitute in any way a representative sample of the employments obtained.

Further, the range of ratings made by employers was very limited, with very few unfavourable responses being made. This does call into question the use of rating scales at all, despite the reasons for their use discussed in Chapter 13, for if employers do not make unfavourable comments then little is achieved by a a scale, unless one substitutes a 'degree of favourableness' scale, rather than one dealing in the unfavourable as well as the favourable. Perhaps the roots lie again in the labour market: if an employer can select from a wide field of school-leavers, those he actually takes on may only perform favourably, having been screened by whatever process for potential for good performance. The relationship here with functional reading ability would be very limited.

In terms of test interpretation, the lack of demonstrable predictive validity has serious effects. One is unable to predict whether a given pupil will have problems with job-related reading tasks. Nor can one say that he or she is more or less likely to obtain any particular type of job, or whether they are likely to find work at all. However, the picture is not completely gloomy. As the test has good content validity, and consists of a representative sample of the reading tasks required of school-leavers, a tester can use the test to see what a pupil can and cannot do in coping with such tasks. Such a use may be employed in two areas: remediation and careers advice.

In the former of these, the teacher must make the decision as to what the pupil should be able to do, or the extent to which he or she might achieve competency with the tasks. The ideal will, of course, be 100% performance on the FRT A, and a pre-instruction, post-instruction testing might be used. From the employer's point of view and presumably also from the pupil's point of view, any improvement in functional reading skill is desireable.

In terms of a careers advisory function, the FRT A may be used to assess specific competencies in relation to known job descriptions. If detailed specifications of the reading required in the job are known - and here the SOFRP may have grounds for further development - the responses of the testee to individual items or subtests should provide a degree of specific prediction not demonstrated in the general validation.

Testing for piloting and validation has been confined to 5th form pupils, but the practice effect study used 4th form pupils at the end of their school year. Very little differences, and certainly no significant ones, could be seen between the performances of the two year groups. It would seem, therefore, that the test may be used with either 5th or 4th form pupils. This relieves the burden on an already crowded fifth year, and allows extra time for appropriate remediation.

In conclusion, the FRT A may be used with 5th form pupils, or 4th form pupils towards the end of the school year. Two functions can be shown to be appropriate: the use of the FRT A to describe the specific competencies of the pupil in order for general remediation on all fronts; and its use in careers advice in relation to specific, well-defined jobs. The use of the content and linguistic task subtests is called for, rather than the job-type subtests.

It is undoubtedly necessary for the FRT A to have predictive validity and this should be established at some time in the future. The same questions would eed to be answered: the relationship of scores to performance and to type of job, and the possible use of criterion scores. A consideration of how the validation study might be improved is given in Chapter 21 below.

CHAPTER 17

ITEM BANKING SYSTEM

Introduction

Colin Byrne suggests that "A question bank is simply a collection of test questions organized to enable easy access to particular types of questions, within the bank" (1975, p.6). In fact, question banks (or 'item banks', a frequently used alternative) can be banks of questions, banks of data associated with questions, or a combination of both. Many banks have been developed as manual, card-index systems but the commonest trend in contemporary test construction is toward computerized systems. As it has already been mentioned that SOFRP uses a computerized system, that will be the main focus of this chapter.

There have been many developments in the field of computerized item banking since Wood & Skurnik (1969) published their important work on the subject. Byrne (1975, p.1) reports, on a cautious estimate, the existence of well over 100 item banks of various kinds throughout the world. There is, of course, a wide range of characteristics pioneered by these banks: some are complete systems which take the user from question-type specification to full test print-out very quickly, other produce lists of item identity numbers related to a file of physical item masters stored elsewhere; some mark the completed tests, others give the test. It is very much the needs of those involved with the bank which determine the nature of the system produced. One bank may have a high degree of technical flexibility, another may be designed to promote superlative test construction, yet another be designed for simplicity of use. The categories 'computer oriented', 'test oriented' and 'user oriented' have been coined by Byrne (1975, p. 12-13).

Computer oriented item banks are not uncommon, particularly in the United States. The CAM Project (Gorth, Allen & Grayson, 1971) uses an immense system, storing questions and data, which selects items and constructs tests. It prints versions of the test and provides punched cards for associated marking and analysis programs. Buckley-Sharp and Harris (1970) describe a huge system for multiple-choice question banking with its own control language. They suggest that "In a data banking system, the definition of the file structure and the precise format of data records are the prime considerations" (p. 230). The IBM FIBEL system is also an extremely flexible, technically excellent computer system. "Users are able to create, maintain and use question banks incorporating both multiple choice questions and questions with numeric answers. It allows insertion of randomly generated data into question frameworks. It allows feedback to the teacher on class performance and to individual examinees on performance on individual questions, if the examiner, in placing the questions on the bank, provided comments with each of the alternative answers" (Hammond, Dean & Morgan, 1976).

In contrast are the test oriented bank designers, to whom the ability to design tests and analyse the results is more important than file structure or relating data to the architecture of the hardware. Further, manipulation of the associated data tends to outweigh any need to keep the items in computer storage. The Computerized Item Banking System developed by Massey & Newbould (1976) is essentially such a system, aimed primarily to develop multiple choice items for specific purposes, and to do so in an efficient and effective manner. Hence they are concerned with what item statistics are required, parameters for item acceptability, reports to item writers and question takers, rather than creating a huge system into which their bank can be inserted. Similarly, Hazlett (1970) has developed MEDSIRCH for use in medical education.

The third type of item bank is user oriented. Here, the bank designer is concerned to enable the unsophisticated computer user to assemble a test, and have results analysed as quickly (and painlessly) as possible. A good example is that of the Classroom Teacher Support System (Lippey, Toggerburger & Brown, 1971), developed for high school history teaching in Los Angeles.

There are, of course, many other systems than the examples considered above (see, particularly, a symposium on item banks in Educational Technology, March 1973). At this point, it is necessary to discuss the sort of item banking system that would be appropriate for use in the SOFRP. The database for the system would be, obviously, the items surviving the content validation procedures, and their associated data. Whilst this was not a large number of items, the test developer would still be able to construct a wide variety of tests, of differing contents, item-types, derivations, and test lengths. In addition to this, the criterion-referenced nature of the test items would mean a range of statistical analyses were needed, plus facilities for retest analysis etc. It is also likely that the construction of tests would be undertaken only by personnel involved in SOFRP, but that such persons would not be sophisticated users. One may conclude, therefore, that the type of bank needed for SOFRP falls between the test and user oriented types of bank. It must allow easy access by an unsophisticated computer user, but give sufficient information for a sophisticated test developer. The nature of the reading passage suggests that only the associated data be banked and the physical items stored elsewhere.

Facilities of the Item Banking System (IBS)

The IBS should support the test constructor at all stages of construction, and this bespeaks a range of facilities. These can best be seen by following the process of test construction.

Firstly, the constructor must discover if items of the type he requires exist, in general terms. A facility for listing the number of items for each classification is therefore necessary (e.g. the number of items associated with apprentice training). Linked with this may be the need to link two classifications together (e.g. the number of items which are for apprentices and are also from the engineering industry), and a crosstabulation procedure is required. Those familiar with the CROSSTABS procedure of the Statistical Package for the Social Science will recognise the value of this facility.

Having established that some items exist that he may require, the constructor will wish to know which specific items fit his requirements. He will also need to look at other characteristics of these items, to see if there are any undesireable characteristics ... or he may know in advance what sort of characteristics he does not want (e.g. he may wish for engineering apprentice items which are not multiple-choice questions). So a facility is called for by which the test constructor can call up the associated data of items conforming to certain specifications. For instance, he may require all items for engineering apprentices, which are not multiple choice, which have previously been omitted by less than five testees, and which have been taken by more than twenty-five testees. Of course, he may be less certain than this about his requirements and may wish items for engineering or apprentices, or perhaps engineering or the cutlery trade, etc. A selection facility must have a fair amount of flexibility.

With his questions, he may go on to assemble a test and administer it to pupils. As one cannot, according to the model used here, predict the test characteristics except in general, this stage is beyond the assistance of the IBS. The next step, that of marking and analysis, is not. There are devices (optical mark readers) for coding pupil-item-responses directly into computer storage, but they are less helpful in the type of mixed tests developed here than their uses in multiple choice examinations elsewhere. Hence, coding and card punching is the task of the constructor; but marking can be undertaken by the computer (coding is the transfer of responses into a machine-readable form;

marking is the interpretation of responses into, for instance, correct and incorrect categories, etc.). Typically, the test constructor will wish to know about individual and group performance; item, test and subtest performance; and, possibly, retest and reliability parameters. Further, he will wish the database kept up-to-date with his new test responses. An IBS facility is therefore required to mark and analyse pupil, item and test performance, undertake reliability estimation and update the data on each item.

Incidental to test construction, the test developer may wish to investigate interrelationships with the database, for all items or for a selected few. Such investigations are likely, in the first instance to be correlations, and facilities are needed to undertake the statistics, and the selection and statistics procedures.

From this outline of the essential facilities, the details of the IBS are given, briefly, below. Full operating details are given in the Manual of the SOFRP IBS, included as Appendix IV.

Data Stored Per Item

Twenty-nine pieces of information are stored for each test item. These are of three types:

(a) identification data,
(b) classification data and
(c) performance data.

The first of these is simply a three-digit number, and all subsequent data are stored as numerical codes or values.

The classificatory data consist of the following types, with the number of codes for each given in parentheses:

(i)	industry (B.S.I.C. categories)	(38)
(ii)	content	(6)

(iii)	job	(8)
(iv)	format of material	(10)
(v)	code number of firm supplying material	(120)
(vi)	size of that firm	(8)
(vii)	question type	(3)
(viii)	key	(four digits representing box numbers)
(ix)	position of page	(4)
(x)	code of last test	-
(xi)	question number in last test	-
(xii)	number of answer boxes	(10)
(xiii)	linguistic task	(4)
(xiv)	replacement item (if any)	-

Performance data consist of the following:

(i) the number taking the item

(ii) the number omitting the item

(iii) the number selecting each answer box

(iv) the number correctly responding

(v) the number of answer changes (discrepancies) at a test and retest.

There are no upper or lower limits on the data.

No indices of discrimination are recorded as these are not regarded as being of particular value (see Chapter 9) and will be different depending upon the ability range of the group tested.

<u>Major IBS Options</u>

IBS options are characterised by an option name related to their functions. There are six major options, conforming to the facilities described above.

Option BANKLIST provides frequency counts of items for each code of ten of the major classificatory data-types. These ten can be over-ridden by the IBS user to substitute other types.

Option CROZTABS produces a crosstabulation of any two of the classificatory data-types.

Option SELECT allows the IBS user to select items on the basis of given specifications. He may request a given number of items or a general search for all items with his specifications. He may specify that a data-type shall have a given code (an 'EQUALS' criterion); that it shall have a code or value less than that specified (a 'LESS THAN' criterion); that it shall have a code or value greater than that specified (a 'GREATER THAN' criterion); or that it shall not have a given code (a 'RESTRICTIONS' criterion). He may specify up to thirty of each type and also specify how many of each type of criteria are to be fulfilled (e.g. 3 out of 4 EQUALS criteria and 1 out of 2 LESS THAN criteria). He may specify how many of the twenty-nine data-types he wishes to have printed out for each selected item.

Option UPDATE scores test responses and performs statistical analyses. A number of suboptions are available:

Suboption	Function
UPDATE	Updates the database with the new test results.
RETEST	Causes retest data to be read and retest analyses to be output for each of the below as appropriate.
SCORES	Causes the output of individual results.
GROUP	Causes the output of group results.
GRPNAMES	Causes the output of testee names in suboptions SCORES and GROUP.
TESTCHAR	Causes the output of a list of the number of items in the test and in each subtest.
ITEMDATA	Causes the output of a table of data giving the number of responses and omissions for each item and answer box.

Suboption	Function
QUALITY	Causes statements of item quality to be output with the table of ITEMDATA.
RESPONSE	Causes a list of response patterns and frequencies for each item.
DISCRIMS	Causes the output of discrimination indices for each item for the whole test or each subtest.
MATRICES	Causes the output of a table of means and standard deviations for each subtest and intercorrelation matrices for test and subtest totals. A test-retest matrix is output if RETEST is used.
DISCREPS	Causes the output of a table of the changes of answers in a test retest.
KAPPA	Causes the output of κ and ϕ coefficients for each cutting score of the test and subtests if RETEST is used.

Option STATIST intercorrelates two data-types for all items in the bank.

Option SELSTAT combines SELECT and STATIST to select items before correlating only their values on the required data-types.

Access to the IBS

An interactive program allows the IBS user to design a data statement, giving his requests for the various options and suboptions. The IBS then uses this data statement to perform the procedures. Output is sent via the internal postal system to the IBS user. Handling of punched cards is kept to a minimum and is only encountered when providing pupil-item-responses for analysis, or introducing new items into the Bank.

IBS error checking does not allow the user to perform unreliable analyses or to produce FORTRAN execution errors. Diagnostic statements are issued and the job cancelled when the user inadvertently enters meaningless or incompatible data.

Conclusion

The SOFRP IBS is designed to allow the designer of a criterion-referenced test of occupational functional reading ability to construct a test with the maximum amount of analytic flexibility, whilst minimising the amount of prior knowledge of computing or data analysis with which he has to deal. The user has to know what sort of items he wants and what analyses and output he requires. Thereafter, he must be able to fill in a simple coding form and answer questions at a computer terminal. Everything is automatic from then on.

The IBS proved exceptionally useful in the construction of FRT A, enabling many different analyses to be undertaken, and on different groups. Having results from testing to hand within three to four days meant that changes in items could be implemented very quickly and that schools involved in piloting and sweep testing could be provided with their results within a short period of time.

The selection procedures were valuable in the design of diagnostic instruments (see Chapters 19 & 20), allowing specific items to be drawn from the bank. This arose from an analysis of characteristic errors encountered when using the FRT A, an analysis performed on the ITEMDATA tables produced by the IBS.

CHAPTER 18

VOCABULARY BANKING SYSTEM

Introduction

The matter of technical vocabulary has been raised on several occasions during this work, particularly in relation to content validity. Different jobs and different industries all have their jargons, their shorthand references, their nicknames and abbreviations, none of which the school-leaver can be expected to know before he starts work. Nonetheless, technical terms do appear in job-related reading materials and cannot be neglected in the SOFRP.

In constructing tests of occupational functional reading ability, it seems likely that some indication of the frequency of appearance of technical terms in the materials under consideration would be of use. The constructor may wish to know if texts from one industry or job carry particularly high proportions of jargon, or do not carry any technical terms. Similarly, he may want some specific vocabulary items as a focal point for discussion with teachers or employers; or he may wish to look at the contexts in which jargon occurs. Whatever his purpose, some facility whereby technical vocabulary for different jobs and industries may be listed, would be useful for the test constructor. To this end, a second computerised banking system was developed, the SOFRP Vocabulary Banking System (VBS), which is described below.

Identification of Technical Vocabulary

For the purposes of SOFRP, technical vocabulary is constituted of words or phrases with single, specific technical meanings, and words or phrases in everyday use given special meanings, within a passage of job-related reading material. From this definition, it follows that some technique must be employed to identify not only the specific technical jargon, but to select, from their contexts, those everyday terms given special meaning. There appears to be no mechanical, objective way of doing this, and one needs to take up the services of experts again, for informed, judgemental assessments of vocabulary items.

The choice of such experts requires particular consideration. At first glance, the problem of identifying technical terms seems automatically to suggest a panel of persons who use the terms, who, by understanding their meaning, can indicate their technical nature. Yet these are the persons who, by their very familiarity with the materials, will tend to overlook a number of their industry's common technical terms (thinking that everyone will know their meanings). In fact, during the interviews with employers, discussion of the job-related reading materials often elicited the comment from the respondent that much of the content was more technical than they had previously thought.

To counter this, one's experts may be selected as an "off-the-street" sample, using measures of general agreement amongst naive judges as to what constitutes technical. One could even use groups of school-leavers themselves. This, however, is too far in the opposite direction. The naive respondent could not be expected to be particularly sophisticated either in the breadth or use of his everyday vocabulary and one therefore would constantly run the risk of unfamiliar, but not technical, terms being selected. Their informed assessment of context might also be limited, allowing everyday, but technically used, terms to escape.

A panel of experts to select technical vocabulary would need, then, to be comprised of persons who may be considered generally ignorant of job-related technical materials, but be sophisticated users of language, with a wide general vocabulary. Such persons would tend to be able to identify readily specific technical terms (and not merely unfamiliar ones), whilst also be able to assess, in an informed way, the contextual contribution in making everyday terms into technical ones.

Such experts were available within the Polytechnic, as members of the Departments of English and of Communication Studies may be considered as sophisticated users (and students of) language, whilst having limited experience in the common use of job-related vocabulary in the industries under consideration. The former Department obviously concerns itself with literature, language development, linguistics, writing and oracy; whilst the latter deals with the many different forms of communication, including language.

Rather than attempt to assess the vocabulary content of all the job-related materials collected during the fieldwork interviews (a task of some enormity unlikely to earn the co-operation of the members of the Departments involved), it was decided to assess only those reading passages of items in the Item Bank. This decision had the merit not only of limiting the amount of work involved, but also of dealing with materials which had been content validated. Each member of the two Departments received five passages to assess, with a explanatory letter (see Letter 10, Appendix III). They were asked to underline each technical term (using the definition above) in the passages. It was hoped to have at least two assessments of each passage, but, although response was fair, insufficient pairs were received to allow any realistic comparison. Each term was listed, and coded onto punched cards, along with its industry, job, content and item number, using the same codes as the SOFRP IBS (Chapter 17). All vocabulary items were placed in computer storage and sorted by industry, content and job.

Facilities of VBS

The Vocabulary Banking System (VBS) is a computer program which accesses the computer storage area and performs certain operations in relation to it. There are two major facilities available: the VBS user may select a range of, or a particular, vocabulary item on the basis of given characteristics; or he may request a tabulation of the number of items for each industry, job, content-type and test item in the Vocabulary Bank.

To select vocabulary items, the user may specify a combination of several characteristics. He may wish items from a certain industry, or a certain job or of a certain content-type, or a combination of all three. Thus, if he wishes, vocabulary items appearing in functional descriptions used by apprentices in the engineering industry, then he may specify content as code 6 , job as code 1 and industry as code 3, and all the items with those characteristics will be printed out. He may, however, be unsure of, say, the job for which he wants items. Not specifying a value for type of job will result in items for the given industry and content being output for each job, not just apprentices. Similarly, one or more of these three specifications may be held constant whilst items are printed out for all values of the unspecified ones.

In combination with these three data types, the user may have a certain word or phrase he wishes to find; or wishes to know in which industries it occurs, etc. He may, therefore, specify a word or phrase of up to 40 characters in length, to be sought in the Vocabulary Bank. He may do this with one or more of the other data-types being held constant, or without them. This latter means the entire bank is searched and all occurrences of the word or phrase are printed out with their associated data.

The other major facility produces a listing of the number of vocabulary items with each code of the four associated data types: industry, job, content and item-number.

Figures 18.1 and 18.2 give examples of the printed output from accession of the Vocabulary Bank. The full operating procedures for using the VBS, including procedures for adding items to the Bank, are given in the "Manual to the Vocabulary Banking System", Appendix V Flowcharts explaining the flow of control within the searches can be found in Appendix IX.

CHAPTER 19

ERROR ANALYSIS

Introduction

As a necessary precursor to any consideration of the diagnosis of functional reading difficulties, one must first discover what errors are made by testees, and, where possible, attribute these errors to some cause or factor. It is too simple to say that each and every error that a testee makes is a result of some shortfall in his functional reading ability, for the testee might be upset or unwell at the time of testing, or suddenly distracted during one item. The test itself contains novel item-types and fairly complicated administration instructions which may produce initial confusion in the testee. The time-limit on the test may result in the omission of items at the end of a section due to the testee's slow speed of working rather than his lack of functional reading ability (although most employers would suggest that someone who cannot perform the reading tasks in the time available is indeed below the required ability level). The point is, however, that not all errors encountered are functional reading errors, and that only errors made consistently by one pupil and/or by a large proportion of pupils are likely to be distinguisable from the other factors mentioned above.

Errors are evidence, however, rather than ends in themselves. Merely to know that a certain testee answered a certain item incorrectly is insufficient for the purposes of diagnosis and remediation. At this level, one could only teach the pupil how to answer the item correctly on another occasion. One could not even be sure he could then deal with items of a similar type, although one would hope for some generalisation. It is necessary to use the evidence one has to try and fathom the intricacies of responding to the items. That is, a process of induction from error to general functional reading difficulty is called for, to create categories of difficulty to be diagnosed and for which remedial activities can be identified and put into operation. Induction is a logical process in need of later empirical investigation and that cannot be discussed here, but the evidence that is available can be used as fruitfully as possible. One cannot reach inside a testee's brain and observe the processes involved in making an error (if one could, one could presumably measure functional reading ability rather more directly than the testing described in this work); one must rely on three sources of evidence: what the test is like, what the testee does and what the testee says he does. The third of these sources is beyond the scope of this work, dealing mainly as it does with the test development, rather than further empirical work; the nature of such a study is considered in Chapter 20.

The analysis of functional reading errors here, therefore, proceeds using information about testee-responses and test items. It must be stressed that the following analysis and ensuing argument are to a very large extent judgemental processes, particularly in the choice of categories of error and difficulties, and that other interpretations are possible. The boundaries of categories may not be as clear-cut as suggested here, nor the specific errors be as important as put forward here. The actual construction of diagnostic tests must wait upon further evidence; what is given here is a list of error types likely to be important in diagnosis and remediation.

Evidence for Functional Reading Difficulties

A function of the computerised item banking system is to produce tables showing how testees have responded to each item, and this is a particularly valuable data source. These testees-item-responses give the following information:

(i) the number of testees omitting the item;

(ii) the number of testees selecting each answer box in the item;

(iii) the number of testees selecting each combination of answer-boxes, where more than one has been selected per item;

(iv) the number of testees correctly answering the item; and

(v) the discrimination index for each item.

A number of statements can immediately be made about these pieces of information. For multiple-choice items, the probability of a pupil selecting a single answer-box by guesswork is 0.2, thus for any answer-box except the key (correct answer) to be selected by more than one-fifth of the testees and/or more than the key, is indicative of some particular error associated with that answer-box. Further, if all answer-boxes received an equal number of responses, guesswork would seem to be at work, indicating some general ambiguity about how to respond. Similar lines of argument may be used in relation to action-type items (though not for the probability of different combinations, rather for the selection of individual answer-boxes). High rates of omission may in the first instance be related to the position of the item in the test. Failing any obvious relationship, one may suspect some lack of clarity in the administrative instructions, or some ambiguity in the text, or that the item is particularly difficult. Negative discriminators have been mentioned on previous occasions as indicators of poor item quality, in particular, of ambiguity.

These, then, are some of the initial statements one can make about patterns of response to criterion-referenced test items. It would not do, however, to ignore the other, very obvious factor in testee-item-responses, that is, the nature of the test item itself, both the reading passage and the question. Here we are concerned with such considerations as

(i) the presentation of the reading passage, including its visual complexity, use of illustrations, etc.;

(ii) the format of the material, e.g. as continuous prose a list of instructions, etc.

(iii) the purpose of the material e.g. to inform, instruct, guide, etc., which will be closely related to the content-type of the material;

(iv) the task the question poses;

(v) the type of response required (e.g. how many answers are required from how many possibilities);

(vi) the nature of the incorrect answer possibilities.

All of these may have different effects upon different readers. Those with poorer vision may be affected more readily by presentation factors. Other may find illustrations of great importance compared with their fellows to whom the text is all. Some may find continuous prose more familiar than sets of instructions, and vice versa. When undertaking error analysis with a group of near normative qualities (although biased to the middle-ability range, see Chapter 14), it is hoped that no abnormal proportions of any particular kind of student (e.g. the short-sighted or those with hearing difficulties) will be present, and that errors discovered are generalisable to the whole target population.

Allied to the above considerations are, of course, the content of the reading passage, including vocabulary, various grammatical forms, the use of formal or colloquial forms, and other structural and linguistic factors. Further, the concept load may be particularly important with pupils encountering new material as well as new types of material. It is here that some measure of readability would come in useful, were an appropriate index available. As was made clear in Chapter 3, however, no index provides a measure for the non-prose or short passages, nor do those available usually take into account the sorts of linguistic factors which may be important (see e.g. Nelson (1978) loc. cit.).

The Functional Reading Test, Form A is given in Appendix VII and the various sets of data on testee-responses during the sweep testing are given in Table 19.1.

Error Types

With the earlier caveat on the degree of judgement involved here, it was possible to delineate four major types of error:

(i) Prior knowledge errors;
(ii) Linguistic errors;
(iii) Conceptual errors; and
(iv) Operational errors.

Each of these is discussed in detail below, but a brief overview may serve as an introduction. Functional reading ability is comprised of higher-order reading skills - searching, skimming, evaluating, rather than decoding print, etc. However, where the lower-order skills are missing or inadequately developed, it cannot be expected that the higher order skills can be used either consistently or accurately. This comprised one aspect of prior knowledge errors: lack of prior 'reading skill' knowledge. The other major aspect was the lack of what might be described as lack of prior 'real-world' knowledge. Here, the error involved the testees not knowing some common abbreviation or vocabulary item, etc.

A number of errors also occurred which may be broadly described as linguistic, in that they appeared to relate to some misapprehension of grammar, syntax or semantics. Also included here were literal and inferential comprehension, about which two specific points must be raised. The first of these is whether comprehension errors are linguistic or conceptual, and the second is whether comprehension can be considered in the two forms used above or whether it is a unitary skill. Lunzer, Waite & Dolan write that "Comprehension means understanding. But there are at least two levels at which understanding may operate. At the lower level, it is sufficient that the matter which he reads makes some sort of sense. To do this he must know the meaning of most of the words and he must see that they hang together grammatically and conceptually ... This level is important ... But it is clearly far from sufficient to enable the reader to learn from what he has read ... To learn by reading, a student needs to penetrate beyond the verbal forms of the text to the underlying ideas" (1979, pp. 37-38). So it can be seen that comprehension implies both linguistic and conceptual factors at work, the former to lift the ideas from the page, the latter to integrate those ideas into the testee's existing conceptual framework. The assignment of error, therefore, becomes somewhat of a 'chicken-and-egg' problem, in that one cannot have understanding without lifting the ideas from the page in the first place, but without being able to integrate the ideas, merely reading them does not mean understanding them. It may well be that these types of errors arise both from those who are unable to read the ideas and from those who, having read them, are unable to integrate them conceptually. The simplest of explanations is that testees are unable to deal with the linguistic side of comprehension as this is presumably the temporally prior step. Comprehension errors were therefore assigned to the linguistic error category.

Lunzer, Waite & Dolan (1979) have advanced a unitary model for comprehension as part of the Schools Council 'Effective Use of Reading' Project, and suggest that 'one cannot reliably measure different skills in comprehension' (p. 69). Whether or not this is the case, and their work is convincing, in this study we are concerned with reading not for learning but in relation to some task or role, and it is justifiable to distinguish tasks which involve different <u>applications</u> of comprehension. Errors of literal and inferential comprehension <u>tasks</u> were used as separate categories, the former referring to tasks where simple meanings were to be discovered, the latter to tasks requiring the testees to think beyond the text.

Conceptual errors were all related to pupils dealing with processes or hierarchically-organised materials and these errors are seen as particularly important due to the frequency of occurrence of such materials at work.

Operational errors were concerned with the failure of the testees to approach the reading passage or question with the appropriate strategy. Most test users will be familiar with these errors.

Each error was assigned a weight and this is discussed below, following which each error type is given in greater detail.

Error Weights

The importance of an error is a function of the frequency with which it occurs and of the number of testees who make the error. A simple weighting statistic was therefore devised by summing the percentages making each identified occurrence of each error and expressing the weight of each error as a percentage of the total identified error (there was a large proportion of responses which could not be assigned to any particular type of error). The error weights are given in Figure 19.1 overleaf.

Prior knowledge errors

It may seem facile to suggest that error arises primarily from a testee's inability to perform the reading task. It is, however, an important point to be underlined that functional reading tasks in this area are not simple and that pupils with low general reading ability will make errors, not by mistake or misapplication of known reading strategies, but by inability to approach the task at all. The first prior knowledge was an inability to read. Having said that, it must be pointed out that no pupil has ever scored less than 3 out of 31 on the FRT A (during the test-retest reliability study) and that the lowest score made by a pupil with known reading ability was 7 (for a pupil with ERT 4 quotient of 70).

Error Types	Error Weight	Category
Literal comprehension	13.9	Linguistic
Location-Guesswork	13.7	Operational
Inferential comprehension	11.4	Linguistic
Keyword misapplication	10.9	Operational
Conventions	9.8	Prior Knowledge
Reading the Question	8.3	Operational
Analysis of Process	7.0	Conceptual
Abbreviations	4.5	Prior Knowledge
Either/or criteria	4.2	Operational
Technical vocabulary	3.8	Prior Knowledge
Headings/subheadings	2.7	Conceptual
Contextual cues	2.2	Linguistic
Interference of Prior Knowledge	2.2	Operational
Common Vocabulary	1.7	Prior Knowledge
Colloquial/formal	1.1	Linguistic
Character recognition	1.0	Prior Knowledge
Active/passive	0.5	Linguistic
Similar/different	0.4	Linguistic

Figure 19.1: Error weights for identifiable occurrences of errors (other identified errors carried negligable weights)

An error linked to basic reading ability was that of a failure to recognise characters. There was a demonstrable failure of some pupils correctly to associate given chemical symbols with the answer options, although clearly stated both in text and question. For instance, the symbol for carbon (C) was often confused with the symbol for chlorine (Cl). This seemed to indicate a failure to recognise and discriminate between the characters as unique patterns.

Testees frequently failed to assign the correct meaning to every-day, commonplace, non-technical words or phrases, or failed to distinguish its commonplace meaning from a specific, less common usage. A typical example is the word 'bungalow': in one item it is given its commonplace meaning - a single storey dwelling place - and yet a number of testees assumed it was synonymous with any sort of house. Again, the word 'replace' caused problems: in the text its less common meaning - to substitute - was used, whilst a fair proportion of testees took its meaning to be the more common 'to put back'.

Specific shortfalls in the prior knowledge of testees arose for common abbreviations (e.g.; i.e.; etc.) and with alphabetic conventions for filing. In this latter item, some effort had been made in content validation to check the opinions of employers about firstly, the correct answer and, secondly, whether they would expect a school-leaver to know the convention. They confirmed that a firm's name not involving personal initials (e.g. Jupiter Rentals Ltd.) would be filed under 'J', and that they would expect a school-leaver to know this. The largest single error here was to file under 'R'.

Only rarely in the course of the whole research was any question related to technical vocabulary. When such was the case, the commonest error associated with such an item was, of course, the lack of knowledge of the meaning of the technical vocabulary.

Except for the basic reading ability errors, these prior knowledge errors were essentially trivial despite some high error weights, in that they could be corrected in a few minutes.

Linguistic Errors

In several cases, confusion arose where a formal form of a phrase has been interpreted colloquially, and vice versa. In one item, the phrase "Do not walk away and leave your machine running" is used, whilst the relevant answer-option was "If you are called away, you can leave your machine running". About 10% of testees chose this option as correct, presumably failing to relate the colloquial 'called away' to the formal 'walk away', as well as the slight inference involved.

The incorrect or inappropriate use of contextual cues tended to be a common error. The testees assigned an incorrect meaning to a word, despite strong contextual cues. This error was distinguished from lack of knowledge where the incorrect meaning used was an alternate or other use of the correct meaning.

A number of pupils failed to assign a stable meaning to a word or phrase, as if not distinguishing between similar but different terms. The term 'Production Manager' was often confused with 'Production Control Manager'. Whether or not the testees knew the conceptual difference, they should have been able to distinguish between them at the 'lifting from the page' stage discussed above.

The passive voice is generally considered to be more difficult for testees to understand than the active voice (see e.g. Slobin, 1971). In a set of instructions including "Have your work checked", testees often passed over this instruction, deeming it not to be an active, when asked what to do next.

The criteria to which a job must conform are specifications and the steps to perform a task are instructions. Frequently, they use similar terms ('must', 'after', 'do'). A small number of testees were unable to distinguish between the two types, however. It is to be noted that these are two specifically different categories of content as used in this work.

Errors in literal comprehension tasks arose frequently as many items were testing in this area. Error arose from the question itself (not understanding the question) and from the passage to read. This latter was seen as the more important and a good example was where testees were required to decide which of a set of statements was true or not, by reference to the passage to read. The majority of incorrect responses appeared to come from some specific distractions in the answer options, (rather than spread across all answers) indicating a failure to understand the passage sufficiently well to distinguish the correct from the incorrect.

Inferential comprehension task errors were less frequent (as the tasks are less common in job-related reading) but made by more testees when they did occur. For instance, a substantial proportion of testees failed to realise that 'twin and earth' cable contained three wires, as they did not make the inference that 'twin' implied two wires.

Linguistic errors are seen as rather more important and contain some of the highest error weights. Many of these errors seem to imply the lack of appropriate strategies for analysing the text, implying in turn that the testees approach the test with little conception of how reading materials can be structured.

Conceptual errors

The most major type of conceptual error was the failure by testees to split up a process description into its component steps, or to do only partially. This led to typical errors such as answering a question from the wrong area of the passage. In the items concerned with the Access guide, testees often attempted to answer the questions from the 'authorisation procedure' section, rather than from the 'voucher completion' section, having failed to realise or analyse the flow of the whole procedure.

A related type of error was the failure to distinguish between an overall heading and the content of its sub-categories. Here, the testee treats the heading and its subheading as equal data, or fails to realise that the subcategory was not included in the heading. For example, in the passage on Engineering Drawings, a specific description of the content of each category of drawing is given, with an accompanying diagram to indicate the hierarchy of detail between categories.

A number of testees, however, appeared to assign the content of Categories lower in the hierarchy to the higher categories as well as the content of the higher category. That is, 'general arrangement' drawings were assumed to contain not only 'general arrangement' but also 'sub-assembly' and 'detail' drawings. In other cases, the opposite error occurred, where equal data was treated as hierarchical.

The fieldwork and content validation procedures made it perfectly clear that the ability to follow written instructions and (descriptions of) processes were fundamental requirements in job-related reading. This means that the conceptual errors take on particular importance.

Operational errors

It must be said at once that every type of error previously outlined could be a product of a failure to locate the required information, combined with guesswork. This was only likely to be true in individual cases, however, rather than across the whole group tested, for as was noted above, guesswork would tend to produce results were each answer box gained an equal number of responses. A number of items did exhibit this spread of responses and the use of guessing was adopted as explanation for these incorrect responses. When otherwise not attributed to another type of error, the location error was considered likely. Such a failure to locate the required information might have been slow speed of response or lack of an effective strategy for searching.

In comprehension exercises at school, it is common practice to scan for a 'keyword' or phrase in the passage to read, having assigned that status to some aspect of the test question. This trategy was no doubt used a great deal in answering the items in the FRT A, particularly as the questions are posed before the passage and the testees instructed to "read the question first and then try to find the answers in the passage to read" (Manual, Appendix VI). Unfortunately, many pupils appeared not to have taken the next, necessary, step that of verification. On encountering the 'keyword' in the text, it was assumed to be the correct occurrence, rather than further checking being used to see if it is the correct one. A number of questions used a late occurrence of a keyword, with specific incorrect answer options related to early occurrences in the text. Responses show a substantial proportion of testees choosing the answer option related to the first occurrence of the keyword.

An error that will no doubt strike a chord for all teachers and testers alike was a failure to respond to the actual question, rather than what the testee assumed the question said. The testees seemed to be leaping at the text, not reading the question. This was probably facilitated by posing the question first.

One particular item caused a great deal of error: that involving a cheque verification procedure. In part, this was undoubtedly due to the novelty of the task, as it involved facsimile cheques. The various patterns of responses, however, lead one to suspect that a large proportion of the error was related to the 'either/or' criteria for accepting the cheque. A particular pattern seemed to be that several of the criteria had to be fulfilled for the testee to accept the cheque, rather than a single one which the 'either/or' specified.

Finally, despite the administration instruction to "use only the information given" in the passage to read, there were a number of errors where prior knowledge interfered with what the passage actually said. A number considered themselves 'undismissable' despite a disciplinary procedure passage to the contrary.

Conclusion

Four major categories of functional reading error have been outlined and their implications for the diagnosis of specific difficulties will be discussed in the next chapter.

Error weights, representing the product of frequency and magnitude of occurrence, have been listed, but it is still necessary to reiterate the qualifications surrounding this error analysis.

The last qualification given above is probably the most important. All errors could be explained by a failure to locate the required information combined with guesswork. This would not, however, explain why different incorrect answer options have been selected more often than others. Whereas the location-guesswork explanation is valid for each individual respondent, it fails to explain differential responses across the whole group. This tends to suggest other processes at work. The location problem is, in fact, that one cannot tell which individual errors are location errors or 'other process' errors. Nevertheless it does seem worthwhile to use a process of induction to speculate about 'other process' errors.

It is in splitting up 'other process' errors into separate types that the next qualification is enjoined. The categories used above are not exclusive or exhaustive, for their boundaries are not sufficiently sharply defined. A number of problems of overlap have already been the subject of some allusion in the discussion above; there are others. The sharpness of category boundaries can, of course, be assessed empirically by the use of judges, as a statistic allied to Cohen's κ (κ max) (Cohen, 1960) will yield a measure of this. Time and resources have not, unfortunately, allowed for this to be undertaken. There seem to be particular difficulties related to the linguistic-conceptual distinction, as one would expect from a test of higher-order reading skills in which linguistic and conceptual variables are inextricably intertwined. It may be that further work will demonstrate it to be a false distinction, to be discarded.

The correlation of functional reading test scores and general reading test quotients has already been demonstrated (Chapter 14, above) and the poor performance on the former test leads one to conclude that the FRT A should not be used with pupils of low general reading ability. Scores of 7 to 11 out of 31 are only just above chance in the circumstances of the test and very little information is thus gained. If one thus excludes the non-reader or extremely poor reader, the remaining errors designated 'prior knowledge' errors are negligable. If uncorrected, they may provide employers with grounds for complaint, but if such trivia are the only grounds, one can be well pleased.

The remaining categories of error are those about which one may be justifiably concerned. With few exceptions, there appears to be an underlying factor to all the errors and that is the use of inappropriate or mistaken strategies for approaching the functional reading task. The reading passages are, of course, unfamiliar to school pupils and the most complex (the cheque verification) completely floors 80% of the testees. The process of reading education is surely supposed to yield pupils who can cope effectively with unfamiliar material and yet the group mean score is only 68% correct. Whilst this is very reassuring in answering the critics of reading standards, it certainly falls far short of an ideal 100% correct. The following chapter will advance some proposals for the diagnosis of specific difficulties and indicate some areas for remediation.

CHAPTER 20

DIAGNOSIS OF FUNCTIONAL READING DIFFICULTIES

Introduction

The functional reading tests described in this work were developed to assess levels of occupational functional reading ability amongst school pupils and, as such, the items contained in the tests are a sample of tasks from a broad domain. The amount of specifically diagnostic information available from the test items is therefore quite limited, for the number of questions uniquely associated with any specific type of error is quite small. Had items been constructed for a mastery test, it would have been much easier to identify *a priori* the expected errors in relation to a narrow objective, and then to devise response options deliberately linked to each error. In the present case, the task of devising a test to encompass the domain and still be within the administrative restrictions imposed was sufficiently large in itself, without the added difficulties of the continual retesting required to include diagnostic response options. One might run the risk of making the test uninterpretable by including too much, and possibly conflicting, evidence.

The solution to such a problem in the present case seems to be the construction of a separate diagnostic test or tests. Several advantages may be associated with such a move. The number of items aimed at testing for the presence of a particular error may be increased and hence increase the likelihood of a correct identification of its presence. Also, diagnostic tests can be used to validate remedial procedures more accurately than the general functional reading test, in a pre- versus post- test study. Further, the construction of such tests allows one to investigate empirically the utility and validity of the error categories discussed in the previous Chapter.

A number of those errors, however, were quite rare and related to one or two particular answer boxes in the functional reading test. It would seem to be more efficient, therefore, to recommend procedures for dealing with them at that stage, rather than including them in any diagnostic instrument.

The purpose of this Chapter is to elucidate the necessary concepts and procedures that would be needed in the construction of diagnostic tests related to the FRT A. It is not proposed to describe an actual process of construction, as that has not yet been undertaken. In fact, such a task requires some further empirical study, the nature of which will also be discussed below.

<u>Selection of Errors</u>

In the preceding Chapter, four major categories of error were delineated:

(i) prior knowledge errors;
(ii) linguistic errors;
(iii) conceptual errors; and
(iv) operational errors.

As has been suggested, several of the errors in these categories were linked to specific items in the FRT A and it will be more valuable to concentrate on the commoner and more general errors. This is especially the case when using some 'cut-off' score or range of scores, for most low-scorers will tend to be making the commoner errors and, thus, the diagnostic test(s) will be concentrating on the specifics relevant to those put forward for remediation. To this end, the whole of 'Prior Knowledge' errors may be excluded from further consideration, as they are more obviously to be dealt with directly from the FRT A. In fact, a number of them are so simple that a few sentences from a teacher should serve to clear up the difficulty. On the other hand, a number are so severe that no functional remediation programme is likely to make any difference.

Five errors are particularly common:

(i) Literal comprehension;

(ii) Inferential comprehension;

(iii) Keyword strategy;

(iv) Location of information;

(v) Inability to follow a procedural description.

By concentrating on these five types, it may also be possible to minimise the number of questions and so relieve the testing burden on a group of already demonstrated poor readers.

Reading Materials and Item Types

All the reading passages included in the FRT A were derived from, or constructed to reflect, job-related reading tasks. As such, it is reasonable to suppose that the materials were of a kind unfamiliar to most school-pupils, at least insofar as their content and general format were probably beyond their experience. Indeed, a fairly high degree of interest was evinced by pupils taking the tests at all stages of development, often being expressed because the materials were identified as from the 'real-world', to do with the world of work, rather than school.

The question immediately arises as to whether it was the nature of the reading tasks with which the low-scoring pupils failed to cope, or whether it was the sheer novelty of the types of materials. In terms of encountering the materials at work, the question is irrelevant, of course, for the pupils are expected to deal with them regardless of the properties which produce error. In terms of diagnosis, however, the question is crucial. If it can be demonstrated that a pupil can cope quite adequately with a particular form of functional reading task when presented in a 'school-language' context, but not in an 'occupational-language' context, then the problem can presumably be solved by teachers introducing job-related reading materials into the classroom in order to familiarize their pupils with them. The continual improvement in test scores shown in the test, retest and retest study of practice effect (in Chapter 12) demonstrates that increased familiarity may well be a factor (time on test is likely to be the other major factor).

It would seem to be the case, therefore, that any diagnostic test should include reading passages from both contexts, so that a comparison may be made. This also implies that similar reading tasks be associated with similar passages from each context. A degree of matching between passages and tasks would be called for. In turn, this raises another point: it is conceivable that <u>all</u> functional reading errors arise from this lack of familiarity with job-related reading materials. If so, the use of such materials in the classroom as a part of everyday teaching might lead to the gradual decrease in <u>functional</u> reading errors, until a point is reached where functional can be perfectly correlated with general reading ability. That is, there comes a time when functional is no longer distinguishable from general. This implies some changes in how general reading ability is considered, of course, and a substantial modification of tests of general reading ability. In a society increasingly concerned with the purposes of reading, however, such a situation is at least within the boundaries of belief.

The item-types associated with such passages should differ little from those developed in the SOFRP, as these have been shown to be effective in reflecting job-related reading tasks. The practice effect study also showed there to be no significant burden associated with the novel item-types and other administration procedures. It is to be noted, however, that there <u>was</u> some difference between the testings, although not statistically significant. This difference m might well have arisen from the lower-scorers 'finding their feet', as it were. To cover this possibility and to ease administration, it might well prove advantageous to limit the item-types to only the

5-answer multiple-choice variety, unless this results in some distortion in reflecting a reading task.

Testing for Errors

As the errors for which one is testing have been identified, it would be counter-productive merely to use similar items to see if the patterns of response are repeated. That merely lends weight to the categorisation, not the diagnosis for individuals. Instead, one may devise items aimed at certain errors, in such a way that, if a testee does not answer correctly, he will tend to select certain responses according to his particular functional reading difficulty. Of course, one must ensure the testee has a fair chance of answering correctly. After all, the FRT A, like all tests, is not foolproof.

It was argued in the last Chapter that all functional reading errors might be a function of a single type: the failure to locate the relevant information. This turns out to be the most difficult error for which to test, for one cannot point a testee directly to an answer and then hope he will not get it right! One solution that has a number of advantages is, in fact, to do almost that. One takes a passage and produces another, identical in content, difficulty and format but not in phraseology and organisation and then numbers the sections and paragraphs of one. Several questions are then posed, not identical but matching in task and difficulty, about each passage. For the passage with numbered sections, the relevant number is given in the question. The hypothesis is this: if the testee has location difficulties with ordinary texts, he will tend to answer the questions on the unnumbered passage incorrectly, and correctly those on the numbered passage where the location is done for him. A considerable amount of effort is likely to devise and validate such passages and questions, however.

The keyword error arises from the misapplication of a useful search strategy. Similarly, the inability to follow a procedural direction is related to search strategies as well as the conceptual difficulties it suggests. It is with these errors that the judicious construction of response options can assist in diagnosis. The use of a reading passage with a number of occurrences, and a question posed concerning one of these which is not the first occurrence in the text. The thinking behind the keyword error suggests that a testee with difficulties in this area will select a response option related to at least any earlier occurrence of the keyword, and most usually, the first occurrence. The passage for use in the second error here will naturally be a set of procedural directions. Response options should be aimed at different sections of the text, and should include a 'Don't know' category. The different response options will indicate any confusion as to where in the text the testee is searching and the 'Don't know' category will indicate either a location error which should coincide with that section, or a general inability to deal with such tasks. Further, questions should be related to steps in the procedure both before and after the one required for the correct answer, to see if this elicits any unique response pattern.

In testing to diagnose difficulties of literal and inferential comprehension, one is concerned with testing understanding of what is in the passage and what can be said using the passage. It should be said at once that diagnostics from items associated with comprehension difficulties will only be useful if there are no 'location' difficulties, for it is clearly impossible to distinguish between them. If, however, no location difficulties can be found, some weight can be placed on these measures. There are a number of ways of assessing levels of comprehension and it may be that it is here that differences between school and occupational contexts cause the most problems. Straightforward multiple-choice questions about a passage, with no particular attention to other response options, might be as effective a method as any other, but it may be that other item-types are better suited, despite the comments on uniformity of presentation above.

Cloze procedures offer an option not open in the FRT A, to elaborate upon the specific understanding of the pupil, but there is little opportunity to compare school material with occupational material.

Inferential comprehension tasks are rather more difficult to deal with, however, and the multiple-choice does seem to be the best measure.

Administration

It is to be hoped that not many pupils reach the diagnostic stage, but it is perhaps better that they do, than enter the world of work less prepared than they might. This being the case, group rather than individual administration would be demanded. All administration should continue, as in FRT A, to be read out, and examples undertaken and explained. An added help might be to have the tester read a question, or part of a question, aloud to a testee on request.

If all items are to have the same format, only one section of the test is needed, and it is recommended that this be untimed, or given within a wide time scale. With more than one type, it is recommended that the shorter be given first, within a wide time scale.

Further Research

It has been indicated that the error analysis used only two of the three data sources which might have been available if diagnosis were the main investigation, rather than a successor to it. Only the response rates and the physical format and content of the items have been used, and discussions with testees have not. It would be unwise to proceed with the construction of any diagnostic tests until such discussions have taken place. Those taking the tests are valuable informants as to how they approached the text, what they thought the questions and passages meant and why they chose particular options. One piece of further research in this area, then, is to administer the FRT A to a sample of Fifth Form pupils of all abilities, and then to interview those making a substantial number of errors in terms of the information suggested above. Alternatively, one could take a small group of pupils and have them talk aloud as they answered each item about what they were doing. One has to balance the unreliability of recall in the first study with the change in motivation, interest and attention levels in the second.

Another study in this area, over a rather longer time, would be to undertake an evaluation of the effects, in functional reading terms, of introducing job-related reading materials as normal comprehension exercises (for instance) into the classrooms of Fourth and Fifth year pupils at secondary schools.

Conclusion

With the identification of functional reading errors, one arrives at a prescription for diagnosing reading difficulties. Putting that prescription into practice is not always as easy as that sounds, however. Except for Prior Knowledge errors, most other errors were related to the use of inappropriate, or misuse of appropriate strategies for dealing with text. The diagnostic techniques suggested above are essentially concerned with testing which, if any, strategy is being used and seeing whether it is the appropriate one, or one of them, for the circumstances.

Before making any final conclusions in this area, it is necessary to undertake a number of further studies, details of which have been outlined. Otherwise one ends up having constructed a model of reading difficulty to which the numbers fit nicely but with which the perceptions of testees jibe continuously.

Nonetheless, certain indications are available to suggest the appropriate format both for diagnosis and subsequent remediation. The former has been considered in this Chapter. A few words on the latter must suffice. It would appear that familiarity with the materials used in the FRT A played a part in the increases in scores in all test-retest situations. The interest in the test materials themselves is singular in comparison with the lack of interest shown by the same pupils to the ERT 4 (although this was, of course, administered second). This suggests firstly, that difficulties may be pre-empted by familiarisation (and research has been suggested for the study of that area) and secondly, that remediation might well take the form of helping pupils deal with job-related reading materials, as well as the specific skills and strategies with which they have difficulty.

CHAPTER 21

CONCLUSION

In 1980 it appears to have entered the common folk lore that school leavers are often less than able to cope with the reading requirements of employment. One notes,however, a decrease in the number of public comments to this effect, either in the Press or as political statements. Also, the sharp upward trend of unemployment amongst school-leavers has meant that employers now have many more applicants than they have posts to offer. As such employers are presumably able to select the most able or most suited for their posts, it is perhaps unsurprising that the complaints have been less often voiced. This is not to say, however, that matters have necessarily dramatically improved. It is more an acknowledgement that work is being seriously undertaken to meet the needs of employers, and that employers are scrutinising more closely what their requirements actually are.

One observes, also, changes of attitudes within the education system towards accountability and a greater emphasis on vocational preparation. These changes may again be linked to the prevailing economic conditions, particularly with pressure to curb public expenditure. These changes also reflect a genuine concern by members of the teaching profession to assist their pupils in an increasingly difficult employment market.

The present study has been concerned with the development of a criterion-referenced test of occupational functional reading ability for use with pupils coming up to leaving school. As such, a fairly detailed investigation was made of the reading requirements of a wide range of jobs in different industries. From this, it was possible to demonstrate qualitative differences between the type of task and the type of reading materials encountered at work and those encountered at school. This is not to say that there was no overlap, certainly there was, but both the circumstances and requirements of use were generally different between the two. Such a result suggests that both the concepts of functional literacy in general and occupational functional reading ability in this specific application are useful and valid.

There is much more to be discovered about the nature of job-related reading, however. Despite the extensive investigation carried out as part of the SOFRP, there is room for much further work. In particular, one might like to consider some more detailed, linguistically based investigations of the whole language milieu of employment for school-leavers, and the comparison of this with the school and the home. Other, interrelated functional skills would also be of value to investigate. Measures of readability of materials would be an especially welcome research tool and teaching aid.

In specific terms, a number of investigations were not carried to their full conclusion during the stage of the SOFRP reported here. The identification of reading tasks encountered in the Metropolitan District Council was not undertaken due to the workload associated with the sweep testing. The response from the trade unions and associations contacted was not complete, and no items were constructed for their materials. Moreover, several other specific studies might be undertaken to supplement the work already done. Some extra study courses in Colleges of Further Education would prove a welcome addition to the purely work-related orientation of the Project, thus making it relevant to all the non-school destinations of 16+ year old school-leavers. Also, whilst it was assumed that very small companies were unlikely to take on school-leavers in general, an investigation of some of these firms would help to substantiate this view. The number of leavers entering small firms in the distributive trades would be of special interest, where a greater degree of responsibility may be placed upon the individual school-leaver.

Criterion-referenced testing has progressed a long way since Glaser's first paper in 1963, and the related technology has grown accordingly. Perhaps one of the least attractive facets of this technology is its concentration on paper-and-pencil, multiple choice item-types, and it would be of value to see measures designed which did not conform to such a model. Of course, for the SOFRP and many other projects, the design has been restricted by the usability of the product for teachers. An investigation of how pupils deal with the new and unfamiliar tasks associated with employment, purely for research purposes, might do well to diversify its measurement approach. A concentration on 'hands-on' performance assessment might prove worthwhile.

Content validation is of extreme importance for criterion-referenced measures, and other workers may wish to consider other applications of the panel studies used here. It might perhaps be of value to have employer experts involved at the item construction phase, and let them write items. Another possibility might be to validate in two stages: firstly the choice of materials around which items were to be constructed; and then secondly the items themselves when constructed.

The problem of restricted variability and its effect on the measurement of correlation has been noted throughout this work, in item analysis, reliability estimation and validation. No one solution has been offered as the product-moment-type coefficient is useful if of a high value, but suspect of lower; agreement coefficients presuppose a cutting score and have a high rate of indeterminacy even when the strength of the relationship is obviously high; and absolute difference measures are useless when there is a practice effect. The use of a calculation of the proportion of response changes was advocated in the reliability study with particular emphasis on changes at retest from correct to incorrect. The proportion of pupils improving their score at retest is also of interest. Yet such measures are not applicable for other correlations such as the calculation of discrimination indices. There remains a great deal of statistical work to be done in this area, using the 'real-world', domain-referenced type of test developed here, rather than the artificially simple mastery tests.

A novel component of this work has been the investigation of the effect of the administrative instructions on response improvements at retest. Continual improvements in score at a retest are not seen as important to reliability, providing they are a reflection of the extra time spent with the materials - a learning effect - rather than a reflection of the fact that the lower scores on the first occasion were due to unfamiliarity with the item-types and administrative instructions. It was demonstrated that there were no significant differences between successive retests which could be laid at the door of administrative procedure or novel item-types.

The concurrent validity of the FRT A was established in relation to the ERT 4 quotient. The correlation obtained was quite a lot less than perfect, and this was attributed not only to the usual factors affecting such measures but also to the fact that, the two tests are related but different in application.

The predictive validation of the test, on the other hand, was much less successful. The results of the study have been reported in Chapter 15 and the likely factors involved in the failure to demonstrate any useful relationship between test score and job performance have been discussed at some length. It will be of more value to consider here possible improvements that might be made, in order that a future study should not suffer similar problems.

The largest single problem associated with the prediction study was the low response rate from employers. To improve this rate, a future study might consider a more aggressive 'marketing' approach to the follow-up, with continual calls and reminders to employers, but whilst this might have the desired short-term effect, long-term relationships might be irrevocably damaged. One might attempt to gain the backing and active co-operation of some official employers' organisation, but the extent to which this might actually improve responses is questionable, as such a body would presumably have only as much leverage as the interest of the employer in the subject of the study - no more than before. A future study might test a different type of sample or a larger one, with fewer higher ability pupils, now that the ranges and types of their performances have been demonstrated. This would mean that a greater proportion of the sample would be in the labour market, but not necessarily that the response rate would improve.

The major alternative to attempting to improve on the existing method would be to redesign the study completely. Bearing in mind the various restricting factors, it is very difficult to offer suggestions. One might arrange with a sample of employers to test their intake and be guaranteed performance scales for each, but this would hardly constitute a prediction from ability at school to performance at work. It would be a prediction, but not the appropriate one. Perhaps one could sample employers and arrange a definite commitment to the return of the rating scale should one of the sample tested be taken on, but there is no guarantee any leaver would join any of the firms involved.

It may be that the rating scales used were not the most appropriate type of instrument to be used. The reasons for their use were discussed in Chapter 13 and will not be repeated here. Suffice it to say that little could be achieved in terms of sensitivity by changing the response options from five to any other number, whilst changing the content endangers the already precarious response rate. A possibility to be considered, however, might be to substitute marks out of five on given traits, whilst removing the apparently afunctional non-reading performance scales.

In considering error analysis and diagnosis, it must be stressed again that much of the discussion in earlier chapters should be approached with the caution associated with judgemental processes. Much collaborative research of the type indicated will be needed to bring diagnostic tests to fruition. In fact, one way to evaluate the categories of error put forward, would be to use a pre-instruction post-instruction testing paradigm. In this way, any improvements would suggest the effectiveness of the diagnostic procedures in identifying the errors for which remediation was necessary, providing that remediation was of known quality, perhaps through use in other contexts. It has been suggested that many of the main functional reading errors might be remediated by the explication of the structural properties of text, with the use of job-related reading materials becoming an everyday part of school activity.

With the demonstrated shortfall in predictive validity, it proved impossible to arrive at any single score or range of scores appropriate for use as a cutting score. As such, the point at which a tester should 'look again' at the functional reading performance of his pupil remains indeterminate. However, the descriptive use of the FRT A has been suggested, rather than a predictive use, and this can be linked to error types and so to diagnosis.

Occupational functional reading is but one of the aspects of functional reading undertaken in everyday life by young people. There remains to study the young person as consumer and citizen, and the world of work has still many secrets to offer. In this study, a test has been developed based on the actual reading tasks encountered by school-leavers in the first six weeks at work. The identification, classification and analysis of such tasks and the construction of test items were aimed at making the test as close a reflection of the 'real world' situation as possible, and hence to make known the reading requirements of such jobs. In a few years time, technological advance may make most of the content no longer valid, but the present study has created a base-line and a basic set of methods for future investigation. Further, the work has served to bring functional reading to the attention of teachers, and in a form which is designed for their use. On the other hand, employers have been able to see that their voices have not been crying in the wilderness.

There is much more to be done in this field, but as this Project has developed a test, an evaluation of the test itself should form the appropriate conclusion to this work. The functional Reading Test, Form A has been demonstrated to be reliable in terms of reproducibility of scores and the relationship between subtest totals and total score. It has been shown to have content and concurrent validity. Unfortunately, no predictive validity could be shown to exist in terms of job performance, employment status and job-type. It is suggested that, while the test is still valid for use, its interpretation be restricted to descriptions of the specific tasks that pupils can and cannot perform.

Ahmann, J.S. (1975) An Exploration of Survival levels of achievement by means of Assessment Techniques.
in Nielsen, D.M. & Hjelm, H.F., Reading and Career Education
IRA Newark, Delaware pp 38 - 42

Algra, C. (1978) Meeting the Challenge of a Minimum Reading Graduation Requirement.
Journal of Reading, Feb.1978 pp 392 - 397

Anderton, W.G. (1979) An investigation into factors affecting the selection, training in part-time further education courses, and the employment of apprentices in Coventry, and their relationship with the labour market.
Unpublished Ph.D. thesis. University of Leicester·School of Education

Ansfield, P.J. (1973) A User-Oriented Computing Procedure for Compiling and Generating Examinations.
Educ.Tech. March 1973, 12 - 13

Ashton, D. (1979) The function of academic and non-academic criteria in employers' selection strategies
University of Leicester. Youth in the Labour Market Project.

Assessment of Performance Unit (1978) Language Performance
Dept. of Education and Science.

Ayrer, J.E. (1977) Problems in the Development of a Test of Functional Literacy.
Journal of Reading, 20(8), 697-705.

Baker, F.B. (1973) An Interactive Approach to Test Construction.
Educ.Tech., March 1973, 13-15.

Baker, F.B. (1977) Advances in Item Analysis.
Review of Educational Research 47(1), 151-178.

Beggs, D.L. & Lewis,E.L. (1975) Measurement and Evaluation in the Schools
Boston, Mass. Houghton Mifflin Co.

Berk, R.A. (1976) Determination of Optional (sic) Cutting Scores in Criterion-referenced Measurement.
Journal of Experimental Education 45,2, 4-9.

Berk,R.A. & DeGagni,G.A. (1979) Technical Consideration in the Evaluation of Pediatric Motor Scales.
American Journal of Occupational Therapy, 33,4, 240-244.

Blanton, W.E., Farr, R. & Tuinmann, J.J. (eds) Measuring Reading Performance
IRA. Newark, Delaware.

Bormuth, J (1966) Readability: a new approach.
Reading Research Quarterly, I, (3), pp 79-132.

Bormuth, J.R. (1973-4) Reading Literacy: its Definition and Assessment.
Reading Research Quarterly, 9, (1), pp 7-66.

Boulmetis, J (1973) Cognitive Evaluation in ABE
Adult Leadership, Sep. 1973, pp 108 - 9.

Brown, F.G. (1970) Principles of educational and psychological writing.
Hinsdale, Ill.; The Dryden Press Inc.

Brown, W.A. (1973) A Computer Examination Compositor for the IBM 360/40.
Educ. Tech., March 1973, 15-16.

Buckley-Sharp, M.D. (1973) A Multiple Choice Question Banking System.
Educ. Tech., March 1973, 16-18.

Buckley-Sharp, M.D. & Harris, F.T.C. (1970) A Computer Program for Banking Multiple Choice Questions.
The Computer Journal, 13, 3, pp 230-236.

Burnett, R.W. (1966) Technical Report No. 1 for the Basic Reading Inventory.
Scholastic Testing Service, Inc.
Bensenville, Ill.
ED 019 548.

Burstall, C. (1978) Assessment - the American Experience
Dept. of Education & Science, London.

Butcher, H.J. (1966) Sampling in Educational Research
Statistical Guides in Educational Research No. 3, Manchester Univ. Press.

Byrne, C.J. (1975) Computerized Question-Banking Systems
Institute of Educational Technology.
The Open University, Milton Keynes.

Cameron, J; Livson, N. & Bayley, N. (1967) Infant vocalizations and their relationship to mature intelligence.
Science, 157, 331-333.

Carver, R.P. (1972) Reading Tests in 1970 versus 1980: psychometric versus edumetric.
The Reading Teacher, 26, 3.

Cassidy, J. (1978) High School Graduation: Exit Competencies?
Journal of Reading: Feb. 1978 pp 398-402.

[illegible] (1972) Should the 'Survival Literacy Study' Survive?
Journal of Reading, 15(6), 429-435

Choppin, B. (1977) Item Banking and the Monitoring of Achievement: An Introductory Paper.
Unpublished document, NFER.

Close, S.J. (1978) Drumcondra Criterion Referenced Maths Test.
St.Patrick's College, Drumcondra, Dublin.

Cohen, J. (1960) A coefficient of agreement for nominal scales.
Educational & Psychological Measurement, xx,1, 37-46.

Cohen, J. (1968) Weighted Kappa: nominal scale agreement with provision for scaled disagreement or partial credit.
Psychological Bulletin, Vol.70 No.4 pp 213-220.

Cohen, P.S. & Cohen, L.R. (1973) Computer-Generated Tests for a Student-Paced Course.
Educ. Tech., March 1973, 18-19

Conrad, H.S. (1951) The Experimental Tryout of Test Materials.
in Linduist, E.F. (ed.) Educational Measurement
Washington. American Council of Education.

Cox, R. (1970) Evaluative aspects of Criterion-Referenced measures.
Paper presented to the American Educational Research Association, Minneapolis.

Cox, R.C. & Graham, R.T. (1966) The development of a sequentially scaled achievement test.
Journal of Educational Measurement, 3, 147-150.

Cronbach, L.J. & Gleser, G.C. (1965) Psychological Tests and Personnel Decisions.
University of Illinois Press/ Urbana.

Curtis, H.A. (ed) (1972) The Development and Management of Banks of Performance-Based Test Items
Harcourt Brace Jovanovich Inc., N.Y., April 1972
ED 072 099.

Dahl, T. (1971) Toward an Evaluative Methodology for Criterion-Referenced Measures: Objective-Item Congruence.
Paper presented at the annual meeting of the California Educational Research Association, San Diego.

in Gilliland, J. (ed)
Reading: Research and Classroom Practice.
Proceedings of the 13th U.K.R.A. Conference, Durham.

Davis, F.B. (1974) Criterion-Referenced Tests: A Critique. in Blanton, W.E.; Farr, R. & Tuinmann, J.J. (eds) Measuring Reading Performance IRA Newark, Delaware pp 44-50.

Department of Education & Science (1974) Educational Disadvantage and the Educational Needs of Immigrants H.M.S.O. Cmnd 5720.

Department of Education & Science (1975) A Language for Life (The Bullock Report) H.M.S.O. London

Department of Education & Science (1977) Education in Schools: A Consultative Document Green Paper, Cmnd 6869, H.M.S.O.

Department of Education & Science (1978) Assessing the Performance of Pupils D.E.S. Report on Education No.93.

Department of Education & Science (1979) Aspects of Secondary Education H.M.S.O.

Derby, T.L. (1975) Informal Testing in Vo-Ed Reading. Journal of Reading, 18, No.7.

Diack, H. (1974) A Way of Measuring Literacy. New Society Vol.27 No.590 pp 191-2

Divgi, D.R. (1978) A New Index for the Accuracy of a Criterion-Referenced Test Paper presented at the National Council on Educational Measurement. ED 160 636.

Dixon, J. (1979) Education 16-19: The role of English & Communication. MacMillan Education/Schools Council. London.

Ebel, R.L. (1972) Essentials of Educational Measurement Englewood Cliffs, N.J. Prentice-Hall.

Educational Testing Service (1973a) State Testing Programs: 1973 Revision ERIC Clearinghouse on tests, measurement and evaluation. Princeton, N.J. ED 087 789

Look.
ETS, Princeton, N.J.
ED 081 783

Elley, W.B. & Reid, N.A. (1969) Progressive Achievement Tests. Reading Comprehension and Reading Vocabulary
New Zealand Council for Educational Research.

Elsner, P.J.E. (1973) Study of Criterion-Referenced Assessment and its Classroom Uses as Viewed by Teachers.
Journal of Educational Research,66 p 220.

Emrick, J.A. (1971) An evaluation model for mastery testing
Journal of Educational Measurement 8, 321 - 326.

Epstein, M.G. (1973) Computer-Assisted Assembly of Tests at Educational Testing Service.
Educ. Tech.,March 1973, 23-24

Ewen, E.L.; Gipps,C.V. & Sumner, R. (1975) Language Proficiency in the Multi-Racial Junior School: A Comparative Study.
Slough: N.F.E.R.

Fillbrandt, J.R. & Merz, W.R. (1977) The Assessment of Competency in Reading and Mathematics using Community Based Standards.
Educational Research Quarterly,2(1) 3 - 11.

Fisher, A.H.Jnr. & Brown, G.H. (1971) Army 'New Standards' Personnel: Effect of Remedial Literacy Training on Performance in Military Service.
ED 056 272.

Fitzgerald, A (1979) Measuring the Arithmetical Competence of Pupils in Relation to Employment.
The Vocational Aspect of Education, Vol.xxxi,No.78,pp 9-21.

Fleming, D. & Lavercombe, S.(1979) Ready for the Real World?
Sheffield City Polytechnic.

Freedman, E.S. (1979) The Certificate of Secondary Education and the Employer.
The Vocational Aspect of Education, 31,80,73 - 81.

Fremer, J. (1973) Developing a Criterion-Referenced Assessment Program
Paper presented at National Council on Measurement in Education, New Orleans.

Fremer,J; Zieky,M; Marco,G; Valentine, S. (1974) Development of Handbook for Conducting Task Analyses and Developing Criterion-Referenced Tests of Language Skills.
Educational Testing Service, Princeton, N.J.

Frisby, C.B., Vincent, D.F. & Lancashire, R. (1959) *Tests for Engineering Apprentices.* National Institute of Industrial Psychology Report 14, London.

Gadway, C.J. & Wilson, H.A. (1976) *Functional Literacy: Basic Reading Performance.* N.A.E.P.

Ganopole, S.J. (1978) *The Construction and Validation of a Criterion-Referenced Test of Fundamental Reading Competencies.* Unpublished doctoral dissertation, University of California, Los Angeles.

Gatewood, R.D. & Schoenfeldt, L.F. (1977) Content Validity and EEOC: A Useful Alternative for Selection. *Personnel Journal*, August 1977

Ghiselli, E.E. (1966) *The Validity of Occupational Aptitude Tests.* John Wiley & Sons. New York.

Glaser, R. (1963) Instructional Technology and the Measurement of Learning Outcomes: some questions. *American Psychologist*, 18, 519-521.

Glaser, R. & Nitko, A.J. (1971) Measurement in Learning and Instruction *in* R.L. Thorndike (ed) *Educational Measurement* (2nd ed.) Washington, American Council on Education.

Glass, G.V. & Smith, M.L. (1978) The Technology and Politics of Standards. *Educational Technology*, 18, 5, 12 -18.

Goldstein, H. & Blinkhorn, S. (1977) Monitoring Educational Standards - An Inappropriate Model. *Bulletin of British Psychological Society*, 30, 309 -311.

Gordon, P. & White, J. (1979) *Philosophers as Educational Reformers* London; Routledge & Kegan Paul.

Gorman, T. (1977) *Adult Literacy Research Project: An Outline of Test Rationale and Objectives* N.F.E.R. Confidential communication.

Gorman, T. (1980) *Developments in Language Assessment: A Report on the National Monitoring Project.* Paper presented to the United Kingdom Reading Association Conference, Warwick University.

Gorth, W.P; Allen, D.W. & Grayson, A. (1971) Computer Programs for test objective and item banking. *Educational & Psychological Measurement*, 31, 245-50.

Greenbaum,W. (1977) *Measuring Educational Progress* McGraw-Hill. New York.

Gronlund,N.E. (1977) *Constructing Achievement Tests* 2nd ed. Englewood Cliffs,N.J. Prentice-Hall.

Grundin, H.U. (1976) Evaluating Competence in Functional Reading. *in* Merritt,J.E. *New Horizons in Reading*, pp 198-207. (Proceedings of the 5th IRA World Congress on Reading 1974). IRA. Newark, Delaware.

Hambleton,R.K. & Novick,M.R.(1978) Toward an integration of theory and method for criterion-referenced tests. *Journal of Educational Measurement*, 10, 159-170.

Hambleton,R.K.; Roberts,D.M. & Traub,R.E. (1970) A comparison of the reliability and validity of two methods of assessing partial knowledge on a multiple-choice test. *Journal of Educational Measurement*, 7, 2.

Hambleton,R.K.; Swaminathan,H. & Algina J. (1976) Some Contributions to the Theory and Practice of Criterion-Referenced Testing. *in* De Gruijter,D.N.M. & Van der Kamp,L.J.T. *Advances in Psychological and Educational Measurement*. London: John Wiley & Sons.

Hambleton, R.K.; Swaminathan,H; Algina,J. & Coulson,D. (1975) *Criterion-Referenced Testing and Measurement: Review of Technical Isses and Developments.* An invited symposium presented at the annual meeting of the American Educational Research Association, Washington,D.C.

Hammond, L.; Dean,C.G.& Morgan,G. (1976) *Final Report on the SHEAF Evaluation Project* Sheffield City Polytechnic.

Harman,D. (1970) Literacy: an Overview. *Harvard Educational Review*, 40 pp 226-43.

Harris, C.W. (1972) An interpretation of Livingston's reliability coefficient for criterion-referenced tests. *Journal of Educational* Measurement,9, 27-29.

Harris, C.W. et al (1974) Problems in Criterion-referenced Measurement.
CSE Monograph Series in Evaluation. California Univ:Los Angeles Center for the Study of Evaluation.
ED 171 732.

Harris,L. & Associates (1970) Survival Literacy Study.
Washington,D.C; National Reading Council.
ED 068 813.

Harrison,C. (1979) Assessing the Readability of School Texts.
in The Effective Use of Reading
ed. Lunzer,E. & Gardner,K.
London. Heinemann -Schools Council.

Hazlett,C.B. (1970) MEDSIRCH: A computerized system for the retrieval of multiple choice items.
Alberta University, Edmonton, Dept.of Educational Administration.
Royal College of Physicians and Surgeons, Ottawa, Ontario.
Report No. DERS-3-70.
ED 049 633.

Heding,H.W. (1967) Missouri Adult Vocational Literacy Materials Development Project, Final Report.
ED 014 650.

Hemphill,J. & Westie,C.M. (1950) The Measurement of Group Dimensions.
Journal of Psychology,29, 325-342.

Hillerich, R.L. (1976) Towards an assessable definition of literacy.
English Journal,65, pp 50-55.

Hirst,P.H. & Peters,R.S. (1970) The Logic of Education
London. Routledge & Kegan Paul.

House of Commons (1976-7) Tenth Report of the Expenditure Committee: The Attainments of the School Leaver.
Paper No.526 I,II & III. H.M.S.O.

Hoyles, M.(ed) (1977) The Politics of Literacy
Writers and Readers Publishing Co-operative. London.

Huynh, H. (1976) On the reliability of decisions in domain-referenced testing.
Journal of Educational Measurement, 13, 253-264.

Ivens, S.H. (1970) An interpretation of item analysis, reliability, and validity in relation to criterion-referenced tests.
Unpublished doctoral dissertation. Florida State University.

Princeton, N.J. Educational Testing Service: ERIC Clearinghouse on Tests, Measurement and Evaluation.

Jackson, R. (1972) The Development of a Collection of Adult Tasks for Assessment of Literacy
Paper presented to the American Educational Research Association, Chicago, Illinois.
ED 061 023.

Jeffries, Sir C. (1967) Illiteracy: A World Problem
London. Pall Mall Press.

Kagan, J. (1971) Change and Continuity in Infancy
New York: Wiley.

Kirsch, I. & Guthrie, J.T. (1978) The Concept and Measurement of Functional Literacy.
Reading Research Quarterly, 13(4), 485-508.

Klare, G.R. (1974/5) Assessing Readability
Reading Research Quarterly, 10(1), 62-102.

Klein, S.P. (1974) Procedures and Issues in the Development of Criterion-Referenced Tests.
in Cronbag, H.F. & DeGruijter, D.N. (eds) Contemporary Issues in Educational Testing.
Mouton. The Hague.

Knox, C. (1977) Numeracy and School-leavers: a survey of employers' needs.
Sheffield Regional Centre for Science and Technology, Paper No.12

Koos, E.M. & Chan, J.Y. (1972) Criterion-referenced tests in biology
Paper presented to the American Educational Research Association.

Kriewall, T.E. (1972) Aspects and applications of criterion-referenced tests.
Paper presented to American Educational Research Association.

Lichtman, M. (1974) The Development and Validation of R/EAL, an Instrument to assess functional literacy.
Journal of Reading Behavior, 6(2), 167-182.

Lippey, G; Toggenburger, F. & Brown, D.C. (1971) Computer based exercises and classroom teachers.
AEDS Journal pp 75-84.

Livingston, S.A. (1972a) Criterion-referenced applications of classical test theory.
Journal of Educational Measurement, 9, 13-26

Livingston, S.A. (1972b) A reply to 'An interpretation of Livingston's reliability coefficient for criterion-referenced tests'. Journal of Educational Measurement, 9, 31

Livingston, S.A. (1976) The criterion-referenced reliability of a single score: Report 76-01
E.T.S. Princeton, N.J.
ED 129 866.

Lockwood, Sir J. (Chmn)(1964) The Examining of English Language
Report of the Working Party on the Schools' Curriculum and Examinations.
D.E.S.

Lord,F.M. & Novick,M.R. (1968) Statistical Theories of Mental Test Scores.
Reading, Mass. Addison-Wesley.

Lunzer,E; Waite,M. & Dolan,T. (1980) Comprehension and Comprehension Tests
in Lunzer,E. & Gardner,K. (eds) The Effective Use of Reading
Heinemann - Schools Council, London
pp 37-71.

MacClure, S. (1979) Introduction - Background to the Accountability Debate Chapter I in Becher,A. & MacClure,S. (eds) Accountability in Education
SSRC/NFER. Slough.

McDonald,T.F. & Moorman,G.B. (1974) Criterion Referenced Testing for Functional Literacy.
Journal of Reading, 17(5), 363-366.

Manpower Services Commission (1979) Sheffield & Rotherham Labour Market Study.
MSC. Sheffield.

Marshall,J.L. & Haertel,E.H. (1976) The mean split-half coefficient of agreement: A single administration index of reliability for mastery tests.
Unpublished manuscript, University of Wisconsin.

Martuza,V.R. (1977) Applying Norm-Referenced and Criterion-Referenced Measurement in the Classroom
Allyn & Bacon. Boston, Mass.

Maryland State Dept. of Education (1974) Functional Reading for Maryland
M. S. Dept. of Education, Baltimore.
ED 106 825.

Massey, A.J. & Newbould, C.J. (1976) Computerised Item Banking System (CIBS) Handbook.
Occasional Publication No.3.
Test Development and Research Unit, Cambridge.

Melnik, A. & Merritt, J. (eds) (1972) *Reading: Today and Tomorrow*
London. Univ. of London Press.

Merritt, J.E. (ed) (1976) *New Horizons in Reading*
(Proceedings of the 5th IRA World Congress on Reading, Vienna, 1974)
IRA Newark, Delaware.

Meskauskas, J.A. (1976) Evaluation Models for Criterion-Referenced Testing: Views Regarding Mastery and Standard-Setting.
Review of Educational Research, 16,1, 133-158.

Mill, J.S. (1924) *Autobiography of John Stuart Mill.*
J.J.Cross(ed). New York.

Millman, J. (1972) Criterion Referenced Measurement: An Alternative.
The Reading Teacher, Vol.26, No.3.

Millman, J. (1974) Criterion-referenced Measurement.
in Popham, W.J. (ed)
Evaluation in Education, Current Applications
Berkeley, Calif: McCutchen.

Moore, T. (1967) Language and Intelligence. A longitudinal study of the first eight years.
Human Development, 10, 88-106

Moser, C.A. & Kalton, G. (1971) *Survey Methods in Social Investigation*
London. Heinemann.

Murphy, R.T. (1973) *Adult Functional Reading Study: Project I: Targeted Research and Development Reading Program Objective, Sub-parts 1,2 & 3. Final Report with Appendices A,B & C.*
Educational Testing Service.
Princeton, N.J.
ED 109 650.

Murphy, R.T. (1975) Assessment of Adult Reading Competence
in Nielsen, D.M. & Hjelm, H.F. (eds)
Reading and Career Education
IRA. Newark, Delaware pp 50-61

Murray, D. (1975) *Rasch Item Analysis & Scaling*
Paper presented to the Annual Conference of the British Psychological Society, Nottingham

Nafziger, D.H; Thompson, P.B; Hiscox, M.D; Owen, T.R. (1975) *Tests of Functional Adult Literacy: An Evaluation of Currently Available Instruments.*
North West Regional Educational Laboratory.
ED 109 265.

Nedelsky, L (1954) Absolute grading standards for objective tests.
Educational and Psychological Measurement, 14, 3-19.

Content Area Teacher.
Journal of Reading April 1978

Nitko, A.J. (1970) Criterion-referenced testing in the context of instruction.
in Testing in turmoil: a conference on problems and issues in educational measurement.
The 35th Annual Conference of Educational Records Bureau: Framingham, Mass.

Northcutt, N.W. (1975) Functional Literacy for Adults.
in Nielsen, D.M. & Hjelm, H.F. (eds) Reading and Career Education
IRA. Newark, Delaware. pp 43 - 9.

Norwood, Sir C. (Chmn) (1943) Curriculum and Examination in Secondary Schools.
Report of the Committee of the Secondary School Examination Council.
Board of Education.

Osburn, H.G. (1968) Item Sampling for Achievement Testing.
Education and Psychological Measurement, 28, 95 - 104.

Ozenne, D.G. (1971) Toward an evaluative methodology for criterion-referenced measures: test sensitivity.
Report No.72, Center for the Study of Evaluation, UCLA Graduate School of Education, Los Angeles, Calif.

Panell, R.C. & Laabs, G.J. (1979) Construction of a criterion-referenced, diagnostic test for an individualised instruction program.
Journal of Applied Psychology, 64,3, 255-261.

Peck, C van N & Kling, M. (1977) Adult Literacy in the Seventies: Its Definition and Measurement.
Journal of Reading, Vol.20, No.8.

Petre, R.M. & Majer, K. (1976) Reading Literacy: Functionally Defined and Practically Assessed.
Mimeographed. Maryland State Dept. of Education.

Popham, W.J. (ed) (1971) Criterion-Referenced Measurement- An Introduction.
Educational Technology Publications, Englewood Cliffs, N.J.

Popham, W.J. (1975) Educational Evaluation
Prentice Hall. Englewood Cliffs, N.J.

Popham, W.J. (1978) Practical Criterion-Referenced Measures for Intrastate Evaluation.
Educational Technology 18(5) pp 19 - 23.

Popham, W.J. & Husek, T.R. (1969) Implications of Criterion-Referenced Measurement.
Journal of Educational Measurement 6,1, 1 - 9.

Powell, W.R. (1972) Levels of Literacy.
Journal of Reading, Vol.20, No.6 March 1972.

Rasch, G. (1960) Probabilistic Models for Some Intelligence and Attainment Tests.
Copenhagen. Danmarks Paedagogiske Institut.

Rasch, G. (1961) On General Laws and the Meaning of Measurement in Psychology.
Proceedings of the 4th Berkeley Symposium on Mathematical Statistics.
Berkeley. Univ.of California Press.

Rasch, G. (1966) An Item Analysis which takes Individual Differences into Account.
British Journal of Mathematic and Statistical Psychology, 19(1) 49 - 57.

Reid, J.B. & Roberts, D.M. (1978) A Monte Carlo comparison of Phi and Kappa as measures of criterion-referenced reliability.
ED 159 226.

Rosen, P. (ed) (1973) Test Collection Bibliographies: Criterion-Referenced Measures.
ETS Princeton, N.J.
ED 104 910.

Rothwell, R. & Zegweld, W. (1979) Technical Change and Employment
Frances Pinter. London.

Roudabush, G.E. (1973) Item selection for criterion-referenced tests.
Paper presented to the American Educational Research Association, New Orleans.

Roudabush, G.E. (1974) Models for a beginning theory of criterion-referenced tests.
Paper presented to the National Council for Measurement in Education, Chicago.

Roudabush, G.E. & Green, D.R. (1971) Some Reliability Problems of a Criterion-Referenced Test.
Paper to A.E.R.A., N.Y.
ED 050 144.

Rovinelli, R. & Hambleton, R.K. (1973) Some Procedures for the Validation of Criterion-Referenced Test Items: Final Report.
Albany: Bureau of School and Cultural Research, New York State Education Department.

(1976) the Assessment of Criterion-Referenced Test Item Validity.
Paper presented to the American Educational Research Association, San Francisco.

Royal, M. (1974) Performance objectives and C-R Tests - we wrote our own!
The Reading Teacher, 27,7, 701-3.

Shavelson,R.J; Block,J.H; Ravitch,M.M. (1972) Criterion-referenced testing: comments on reliability.
Journal of Educational Measurement, 9, 133-136.

Shoemaker,D.M. (1973) Principles and Procedures of Multiple Matrix Sampling.
John Wiley, London.

Shoemaker,D.M. & Hornke,L.F. (1976) Multiple Matrix Sampling und seine Anwendungen in der Schulforschung.
Zeitschrift fur Entwicklungpsychologie und Pädagogische Psychologie, 8(1), 13-23.

Smart,N. (ed) (1968) Crisis in the Classroom
quoted by Barzun,J.
The Centrality of Reading, p 73.
Black Paper 3: Cox C.B.& Dyson,A.E.
London. Critical Quarterly.

Smith, C.W. (1974) Criterion-referenced assessment.
in Crombag,H.F.& De Gruijter,D.N.(eds) Contemporary Issues in Educational Testing.
Mouton. The Hague.

Sockett,H. (ed) (1980) Accountability in the English Educational System.
Unibooks (Hodder & Stoughton). London.

Spache,G.D. (1970) The Research and Development Program on Reading.
National Reading Conference.
ED 050 892.

Stanley,M. (1972) Literacy: The Crisis of a conventional wisdom.
School Review. May 1972, 373-408

Start,K.B. & Wells,B.K. (1972) The Trend of Reading Standards.
NFER. Slough, Windsor.

Sticht,T.G. (1973) Research toward the design, development and evaluation of a job-functional literacy training program for the United States Army.
Literacy Discussion, 4(3), 339-369.

Sticht,T.G. (ed) (1975) Reading for Working: A functional literacy anthology.
Human Resources Research Organisation.
Alexandria, Va.

Reading Task Tests.
HumRRO, Alexandria, Va.
ED 064 683.

Sticht, T.G; Caylor, J.S; Kern, R.P; Fox, L.C. (1972) Project REALISTIC; determination of functional literacy skill levels. Reading Research Quarterly, 7, pp 424 - 465.

Sticht, T.G. & McFann, H.H. (1975) Reading Requirements for Career Entry. in Nielsen, D.M. & Hjelm, H.F. (eds) Reading and Career Education. IRA, Newark, Delaware pp 62-76.

Subkoviak, M. (1976) Estimating reliability from a single administration of a criterion-referenced test. Journal of Educational Measurement, 13, 265-76.

Subkoviak, M. (1978) Empirical Investigation of Procedures for estimating reliability for mastery tests. Journal of Educational Measurement, 15,2, 111-116.

Sumner, R. & Robertson, T.S. (1977) Criterion-referenced measurement and criterion-referenced tests; some published work reviewed. NFER Publishing Co., Windsor, Berks.

Super, D.E. & Crites, J.O. (1962) Appraising Vocational Fitness. Harper & Row. New York, 2nd edition.

Swaminathan, H; Hambleton, R.K. & Algina, J. (1974) Reliability of Criterion-Referenced Tests: a Decision-Theoretic Formulation. Journal of Educational Measurement, 11,4, 263-7

Thomas, R. & Wetherall, D. (1974) Looking Forward to Work HMSO

Thompson, R.A. & Dziuban, C.D. (1973) Criterion referenced reading tests in perspective. The Reading Teacher, 27, No.3

Ward, J. (1970) On the concept of criterion-referenced measurement. Brit. Journal of Educ. Psych. 40(3) pp 314-323.

Wedman, I. (1974) On the evaluation of criterion-referenced tests. in Crombag, H.F. & De Gruijter, D.N. (eds) Contemporary Issues in Educational Testing. Mouton. The Hague.

for further education.
Schools Council Examinations Bulletin 24
Evans/Methuen Educational, London.

Willmott,A. (1978) Assessment & Performance.
Oxford Rev. Educ. 4(1), 51-64.

Willmott,A.S. & Fowles,D.F. (1974) The objective interpretation of test performance: the Rasch model applied.
NFER, Slough.

Wood,R. & Skurnik,L.S. (1969) Item Banking.
NFER. Windsor, Berks.

Wright,B.D. (1967) Sample-free Test Calibration and Person Measurement.
Proceedings of the 1967 Invitational Conference on Testing Problems, E.T.S., Princeton, N.J.

Wright,B.D. (1977) Solving Measurement Problems with the Rasch model.
Journal of Educational Measurement, 14(2), 97-116.

Young,J.C; Knapp,R.R. & Michael,W.B. (1970) The validity of the tests of achievement in basic skills for predicting achievement in general mathematics and algebra.
Educational and Psychological Measurement, 30, 951-954.

Sheffield City Polytechnic

anguage Development Centre

A

UNCTIONAL READING TESTS

AME ______________________________

CHOOL ______________________________

DO NOT TURN OVER UNTIL YOU ARE TOLD TO DO SO.

PLEASE READ THIS

On the following pages, there are different sorts of questions, and next to each one is something to read. Look at the question first and try to find the answers in the passage to read.

All the answers can be found in the passage and you should use only the information given.

For most questions, there is usually an example that shows you how to answer the question.

If you wish to change your mind, fill in the box completely, like this: and then tick the box you think is right.

TURN OVER AND LOOK AT THE EXAMPLES

Here is a list of duties: place a cross (X) in the boxes of those which are the same as the duties of a Warehouse and General Assistant, given on the opposite page.

- [ ] A. Is responsible for balancing the till at the end of the day
- [ ] B. Services the shop floor with the required merchandise as directed
- [ ] C. Reports immediately any problem or queries to the Manager
- [ ] D. Controls workflow to and from the warehouse
- [ ] E. Rotates stock in accordance with Company practice
- [ ] F. Carries out Trolley Duties
- [ ] G. Changes customer receipt roll as required
- [ ] H. Obeys all the Company rules and regulations

TITLE: WAREHOUSE AND GENERAL ASSISTANT

RELATIONSHIPS: Responsible to the Warehouse Manager/Foreman as applicable.
Works in co-operation with Goods Inward, Warehouse and other store staff as directed.

DUTIES/RESPONSIBILITIES:

Company:

1. Maintains the highest standards of customer relations.
2. Complies with all Company rules and regulations.

Warehouse:

3. Exercises great care in the handling of goods to avoid damage.
4. Manually unloads delivery vehicles in the loading bay area.
5. Removes goods from the loading bay area and safely stacks them manually as necessary.
6. Rotates stock in accordance with Company practice.
7. Selects merchandise for shop floor requirements as directed.
8. Re-prices warehouse stocks according to current 925 as required.
9. Sweeps and maintains the warehouse in a clean, safe and tidy manner and removes all waste/cardboard.
10. Uses warehouse equipment - pallet trucks, conveyors, etc. in a safe, efficient manner.
11. Services the shop floor with the required merchandise as directed.

General:

12. Carries out other duties as directed, including:

Car Park Duties	Baler/Compactor Duties
First Aisle Duties	Cardboard Collection Duties etc.
Trolley Duties	
Damages Duties	Van Driver's Mate

13. Assists in achieving the highest standards of teamwork and efficiency of the Goods Inwards and Warehouse operation.

JOB PURPOSE: To assist in the maintenance of a safe, clean, smooth running Foods Inward and Warehousing operation and to carry out other duties (as shown above) efficiently and correctly.

Here is a list of things you might expect to be told when you start work. Put a tick (✓) next to each one that Peter Jones would tell John Smith when he starts at <u>that</u> steel works. The first one is done for you.

- [x] Hot bars burn
- [ ] Location of canteen
- [ ] Location of toilets
- [ ] Need to use protective equipment
- [ ] Details of pension scheme
- [ ] Unsafe clothing
- [ ] Basic hygiene regulations
- [ ] Artificial respiration
- [ ] Location of Ambulance Room
- [ ] Driving a fork-lift truck

BASIC TRAINING PROGRAMME

Name: JOHN SMITH Clock No. 0371 Date 19.10.79

Training given by P. JONES

1. Supervision

a) During the initial part of your employment with this Company, you will be responsible to:-

- i Foremen P. JONES
- ii Assist. Manager N. BROWN
- iii Manager F. BLACK

b) In matters of discipline, the Foreman has the Company's authority to act.

c) In the case of employee complaints on all matters, these should be directed through the Foreman, to the Manager, finally to the Works Manager, when necessary. (Shop Steward to be advised if necessary).

2. Safety

The Company and all employees are subject to the Health & Safety at Work Act, 1975, and any codes of practises issues by the authorities.

a) Mill Safety

- i Hot bars burn
- ii Black hot bars burn
- iii Never turn your back on mill
- iv Never bring unauthorised personnel into mill
- v Unsafe clothing in mill
- vi Hazards peculiar to 9" mill Straightening bed

b) General

- i Need to use protective equipment
  - a) boots
  - b) glasses (protection of eyes regulations)
  - c) gloves
  - d) gaiters
- ii Never pass under loaded cranes
- iii Horseplay, or fooling of any kind, must not take place in these premises
- iv Location of Ambulance Room
- v Location of Toilets

3. Shift Pattern

a) Details of your shift pattern are:
6.00 a.m. - 2.00 p.m., 2.00 p.m. - 9.30 p.m., 10.00 p.m. - 6.00 a.m., 8.00 a.m. - 4.30 p.m. or as directed

b) Details of early day system - in mill only

Signed Peter Jones
Basic Instructor

Signed John Smith
Employee

Put a tick (√) in the box next to each element in the following list which is contained in Acid or Basic Iron. The first one is done for you

- [x] Silicon (Si)
- [ ] Magnesium (Mg)
- [ ] Manganese (Mn)
- [ ] Phosphorus (P)
- [ ] Platinum (Pt)
- [ ] Carbon (C)
- [ ] Chlorine (Cl)
- [ ] Tin (Sn)
- [ ] Sulphur (S)
- [ ] Oxygen (O)

STEEL

Steel is an alloy of iron and iron and carbon plus trace elements, manganese, silicon, sulphur and phosphorus, not exceeding one per cent. This is steel in the most simple form.

Carbon - Carbon dust, graphite, coal, coke, soot, pencils, diamonds (most pure form).

Iron - Derived from iron ore. Elemented iron 5% Fe is the fourth most abundant element in the earth's crust. The other three are (a) Al (clay), (2) Oxygen, (3) Si (sand). Rocks with over 20% Fe are termed ores. Iron is extracted from iron ore by a reduction process in a blast furnace. After processing ore by a reduction process in a blast furnace. After processing the iron is tipped into pig casting machine or a sand bed.

Types of iron produced

	Si	P	S	C	Mn
ACID IRON	2.3%	.04%x	.04%x	3.4%	1.2%

This is a high silicon iron with low sulphur and phosphorus.

	Si	P	S	C	Mn
BASIC IRON	1.2%	2.5%x	.08%x	2.5% / 3.5%	.6 - 1.5%

This is a medium silicon high phosphorus and sulphur iron.

Steel is a Crystaline Structure

All metals are made up of grains. Their grains are made up of crystals. Their crystals are very small and there are millions of crystals to the grain. There are thousands of grains to one cubic inch. If the metal is worked the grains distort and slip along the slip planes.

CHEQUES

A cheque can be accepted from a customer where:

(a) It is in payment of a Credit Account.

or

(b) The customer is known to the Manager.

or

(c) The customer presents a Bank Cheque card or a Barclaycard as evidence of his identity, provided certain points are checked first.

or

(d) The amount of the cheque does not exceed £5.00.

or

(e) The customer wishes or is willing to have the goods delivered to his home. Seven days should be allowed for clearance of the cheque before the goods are despatched.

Read the details given above and then look at each of the cheques opposite.

If you think you can accept the cheque, put a tick (✓) in the box next to it.

1st July 1979

S
B C
Sheffield Banking Company Ltd

Pay Your Shop & Co. Ltd or order

Twenty-seven pounds and 99 pence only £27-99

T. Smith

T. Smith

00000 11 4444 88900000

'Bankers Cheque Card checked and found correct'

1st July 1979

S
B C
Sheffield Banking Company Ltd

Pay Your Shop & Co Ltd or order

Seventeen pounds only £17

T. Smith

T. Smith

00000 11 4444 88900000

1st July 1979

S
B C
Sheffield Banking Company Ltd

Pay Your Shop & Co. Ltd or order

Three pounds and 20 pence only £3-20

T. Smith

T. Smith

00000 11 4444 8890000

'No Cheque Card'

Here is a list of statements.

Read each one and compare it to the list of safety rules given opposite.

If the statement is <u>true</u>, put a tick (✓) in the box next to it.

The first one is done for you.

- [x] Compressed air can kill
- [ ] You must not tamper with a machine
- [ ] Tools whould be put down near where you are working
- [ ] If you are called away, you can leave your machine running
- [ ] You must not throw things
- [ ] You must not replace damaged tools
- [ ] You must switch off faulty machines
- [ ] Anyone can operate a machine
- [ ] You should check oil levels before starting

DO

1. Keep machines and all equipment clean, and in good condition.

2. Before starting a machine ensure that you know how to stop it.

3. Switch off the machine immediately anything goes wrong.

4. Keep the machine and surrounding area tidy.

5. Check oil levels before first starting machines.

6. Switch off machine at mains at end of each day.

7. Check that chucks or cutters rotate in the correct direction before commencing cutting operations.

8. Use the correct tool or cutter for the job.

9. Replace tools that are worn or damaged.

10. Keep tools and cutters in boxes or racks when not in use.

11. Report immediately to your instructor any mechanical or electrical fault.

12. Ensure that all machine guards are in position before starting the machine.

13. Check that the work area is clear before starting the machine.

14. Ensure that everything is properly secured before starting the machine.

15. Ensure that feed mechanisms are not engaged before starting the machine.

DO NOT

1. **Do not** attempt to operate a machine until you know how to use it correctly.

2. **Do not** tamper with a machine.

3. **Do not** remove any stops in an effort to obtain a greater cutting range, or the machine may be seriously damaged.

4. **Do not** try and reverse the direction of a spindle while it is in motion.

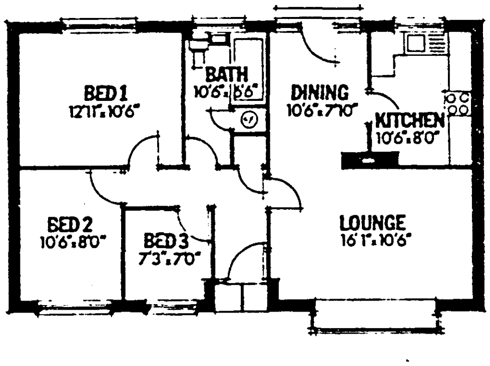
5. **Do not** try to change a spindle speed while it is in motion.

6. **Do not** throw things.

7. **Do not** walk away and leave your machine running.

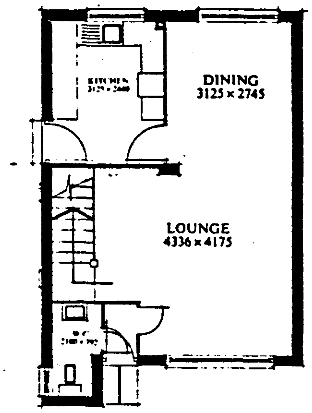
8. **Do not** direct compressed air at yourself or workmates. It can kill.

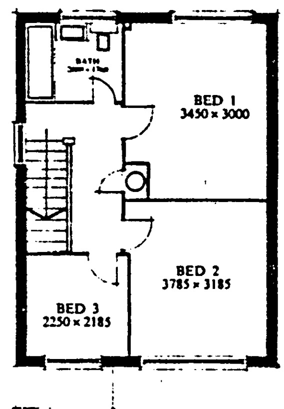
9. **Do not** leave crane hook over machine, or the surrounding area after use.



A spacious three bedroom bungalow with attractive broad frontage comprising of Hall, generous sized fitted Kitchen, large Lounge with separate Dining Room. Two double Bedrooms and one single Bedroom. Bathroom with W.C.

BRENTFORD





This attractive spacious 3 bedroom detached house comprises of an entrance lobby with cloakroom and W.C. off. The large lounge interestingly designed, features an open plan staircase and leading from the lounge is the separate dining area. There is a generous sized fitted kitchen. On the first floor there are 2 large and a smaller third bedroom. Bathroom with W.C.

Imperial Equivalents

Kitchen	10' 3" x 8' 0"
Dining	10' 3" x 9' 0"
Lounge	14' 3" x 13' 8"
W.C.	6' 10" x 2' 7"
Bathroom	6' 6" x 5' 10"
Bed 1	11' 4" x 9' 10"
Bed 2	12' 5" x 10' 6"
Bed 3	7' 4" x 7' 2"

A young couple are seeking a house. Here is a description of the sort of house they want.

2 <u>or</u> 3-bedroomed semi-detached house

Large kitchen with Dining Room

A separate Lounge

One bedroom must be at least 12'0 x 10'0 in size.

From the details of the five types of houses given on these pages, find the house which fits their description.

Put a cross (X) in the box next to its name in the list,

☐ Fife

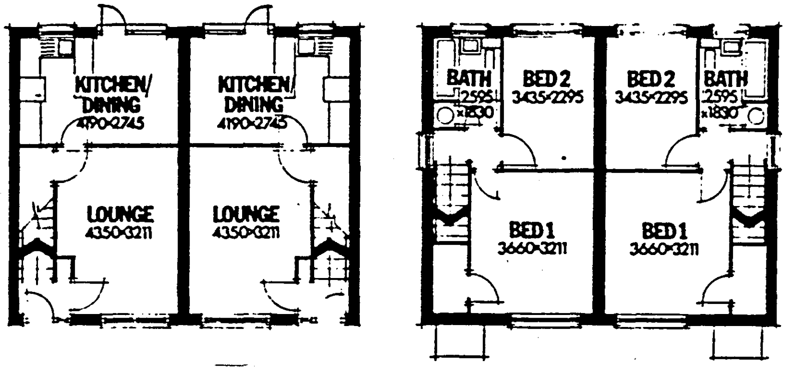
☐ Bren

☐ Mont

☐ Cumb

☐ Gram

Ground Floor **First Floor**

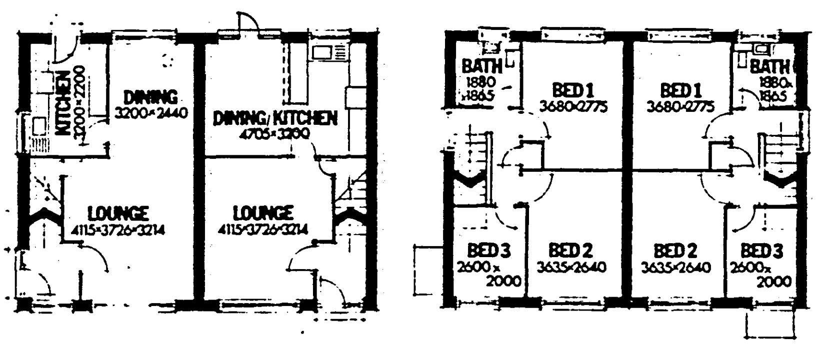


Imperial equivalents:
Kitchen/
Dining Room 9′ 0″ × 13′ 9″
Lounge 10′ 6″ × 14′ 3″

Bed 1 10′ 6″ × 12′ 0″
Bed 2 7′ 6″ × 11′ 3″
Bathroom 6′ 0″ × 8′ 6″

An ideal home for first time buyers, the Cumbria is a spacious two bedroom semi-detached house. A large Lounge with a modern open plan Dining Room and Kitchen makes this an exceptional house design. Upstairs there is a large main bedroom and one medium room. together with bathroom and W.C.

GRAMPIAN MONTGOMERY



Imperial equivalents:
Lounge 13′ 6″ × 12′ 3″ × 10′ 6″
Dining Room 10′ 6″ × 8′ 0″
Kitchen 10′ 6″ × 7′ 3″
Bed 1 12′ 0″ × 9′ 2″

Bed 2 11′ 11″ × 8′ 8″
Bed 3 8′ 6″ × 6′ 6″
Bathroom 6′ 2″ × 6′ 1″

Two modern styles of three bedroom semi-detached houses. The Grampian comprises of a spacious open plan Lounge and Dining Room with a generous sized kitchen. The Montgomery has an alternative plan with a separate Lounge and a large open plan Dining Room and Kitchen. Upstairs, they both have two spacious bedrooms and one smaller, together with bathroom and W.C.

Place a tick (✓) in the box next to each subject you should normally learn in your first year as an apprentice:

The first one is done for you.

- [x] Sterilisation of tools and equipment
- [ ] Permanent waving
- [ ] Combing and Brushing
- [ ] Structure of hair and skin
- [ ] Reception of Clients
- [ ] Reversed Curling
- [ ] Finger Waving
- [ ] Measuring for a Wig
- [ ] Composition and nature of hairdressing materials

INSTRUCTION IN LADIES' HAIRDRESSING

1st YEAR

*1. Deportment, cleanliness, dress, care of hands, and nails.
2. Care and handling of tools and equipment.
3. Sterilization of tools and equipment.
4. Care and handling of electrical equipment.
5. Maintenance and care of the salon and equipment.
*6. Reception of clients: in person and by telephone including booking appointments.
7. The structure of the hair and skin.
8. Instruction in the composition and nature of materials used in hairdressing, Shampoos, Permanent Waving materials, Tints, Dyes, Setting Lotions, Lacquers, etc. Also manicure preparations.
9. The elementary principles of and instruction in Marcel waving.
10. Practical preparation of materials for use.
11. Shampooing. Methods, objects and effect of different materials.
    Note: Elementary massage movements can be taught during shampoo instruction.
12. Combing and Brushing.

2nd YEAR

13. Finger Waving natural movements.
14. Application of temporary colours.
15. Cutting with both scissors and razor.
16. Elementary practice in winding on rollers and curlers.
17. Reversed Curling.
18. Setting. A full pli including pin curls and rollers.
19. Planning a pli for final results and dressing out the style.

3rd YEAR

20. Permanent Waving.
21. Bleaching.
22. Tinting—including semi permanent colours.
23. Continued practice in cutting and setting.
24. Decolourising and advanced tinting.
25. Treatments, Massage and high frequency.
26. Cleaning and dressing wigs and transformations.
27. Measuring for a wig, and placing workroom order.
    Minor repairs to foundation and knowledge of knotting, sufficient to repair a wig, should be included. An appreciation of hair preparation and weaving should be taught.

*Parents and/or Guardians should be jointly responsible for the dress, appearance and speech of the apprentice, with particular attention to the hair style and clothes worn.

Note:
1. It is recommended that the apprentice be encouraged to take the Guild General Certificate of Hairdressing Qualification before the end of his/her apprenticeship.
2. This suggested course for the apprentice is laid out for the 1st, 2nd and 3rd years but as some apprentices learn much faster than others, the order of the subjects is suggested as a reasonable sequence of instruction.
3. If at the end of the Apprenticeship the Apprentice continues in employment as Assistant to the Master, the parties shall execute an Assistants Agreement. Time served as Apprentice will then count for calculating the period of notice subject to the provisions of the Contracts of Employment Act 1972.

These names have to be filed alphabetically.

Place them in alphabetical order by writing a figure 1 in the box next to the one that comes earliest in the alphabet, a figure 2 in the next earliest, a figure 3 in the next, and so on until all eight have been ordered.

The first one is done for you.

N Harris Esq
917 Colebrook Road
Ashton-sur-Vale
Cheshire
CH6 9AB

Jupiter Rentals Ltd
9 Fawcett Road
Welwyn Garden City
Herts

1

J P Brass & Co Ltd
Rutland Avenue
Merseytown
Lancs

Mr J Reckitt
Middle Chambers
Law Row
OXFORD
Oxon

N P Tolman Esq
57 Englebrook Street
Sheffield
S1 3PZ

Anne Charpal
Avenue des Champs-Elysees
PARIS - 84103
France

Mrs J Morgan
High Tower
Chesterman
Essex
CL15 9PT

Mr F Price
3 Providence Mews
LONDON
N6 3YP

PLEASE READ THIS

On each of the following pages, there are three (3) questions to answer and something to read.

The questions can <u>all</u> be answered using the information given in the passage to read and you should use <u>only</u> the information given.

Put a tick (✓) in the box next to the answer that you think is the correct one. If you wish to change your mind, fill in the box completely, like this: ■ and then tick the box you think is right.

TURN OVER AND LOOK AT THE EXAMPLE

1. When should a file-holder be used?

- [ ] A. When the file is sharp at the edges
- [ ] B. When a lot of filing is to be done
- [ ] C. When the surface is larger than the file
- [ ] D. To avoid blunting teeth
- [ ] E. To ensure a smooth finish

2. In this passage, 'deliberately'means

- [ ] A. On purpose
- [ ] B. Accidentally
- [ ] C. With thought
- [ ] D. Freely
- [ ] E. None of these

3. Which of these statements is <u>untrue</u>?

- [ ] A. A file can be cleaned with a strip of soft metal
- [ ] B. Cross filing is done at 90° to the original strokes
- [ ] C. Files should be stored in a box or rack
- [ ] D. Chalk causes clogging of files
- [ ] E. Care must be taken with finished surfaces

Filing

r filing in one direction, the work is filed n with the strokes at 90° to the original ones.

Filing

e the surface has been filed down almost to draw filing can give a fairly accurate and oth finish. Sometimes particles of material me embedded in the teeth of the file, enting a smooth finish being achieved. k may be rubbed into the teeth to avoid ging and so ensure as smooth a finish as ible.

g Large Surfaces

re the surface is larger than the file, a file er should be used.

of Files

void blunting teeth and breaking files they ld be stored carefully in a tool box or rack.

Cleaning Files

A file should be cleaned by brushing with a file card in the direction of the cut, or by means of a strip of thin, soft metal.

Protecting Finished Surfaces

Once the first surface of the workpiece has been finished, care must be taken to maintain that finish when subsequent faces are being worked on. The workpiece must be carefully handled. It should be positioned deliberately and gently. If the component has to be held in a vice or clamps, soft jaws or soft metal protecting pieces should be used. If a hammer has to be used on the workpiece, it must be rawhide or soft faced. Finished faces should be protected by some sort of shield if there is the slightest danger of tool slip.

1. If the card has expired, what would you do?

- [ ] A Wait until told what to do next
- [ ] B Telephone for the police 037
- [ ] C Refuse to serve the customer
- [ ] D Call the Manager
- [ ] E Telephone for an authorisation

2. With what should the customer sign the sales voucher?

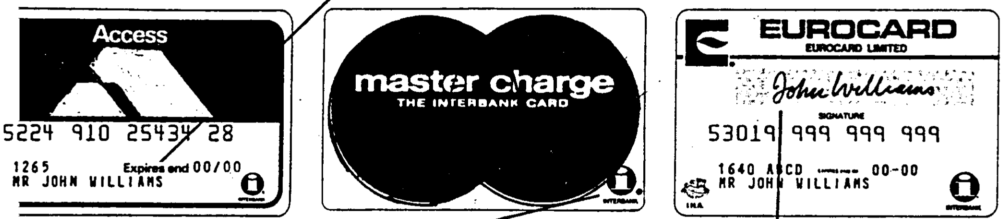
- [ ] A With a ball point pen
- [ ] B With any writing instrument 038
- [ ] C With <u>your</u> pen
- [ ] D With <u>his</u> pen
- [ ] E With a fountain pen

3. If a sale is authorised what do you then do?

- [ ] A Place the card face upwards in the imprinter
- [ ] B Use Access stationary only 039
- [ ] C Write the authorisation on the sales voucher
- [ ] D Carry on as if nothing had happened
- [ ] E Hand the customer his receipt

Access Guide for
Retail Sales

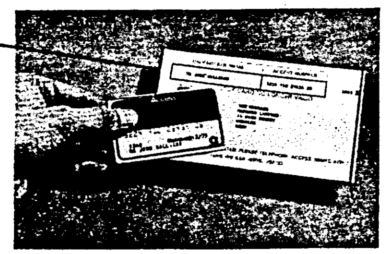
ur Important Points for your Protection



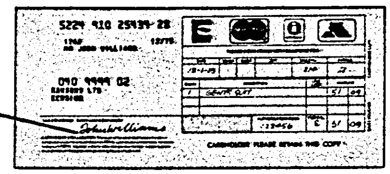
Do you recognise the d? It will carry the cess symbol and/or the erbank symbol.

2 Has the card expired?

3 Is it on your void card list?

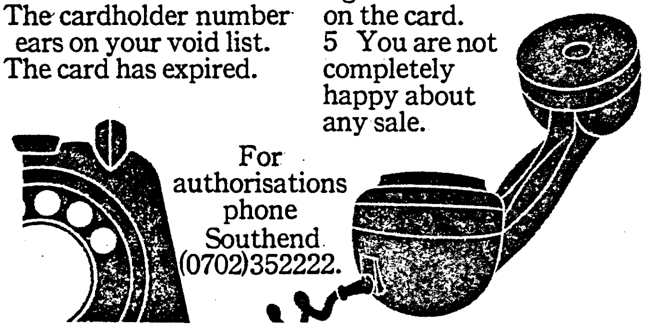


4 Is the customer's signature the same as the signature on the card?



uthorisation is needed if

The sale exceeds your r limit.
The cardholder number ears on your void list.
The card has expired.
4 The customer's signature differs from that on the card.
5 You are not completely happy about any sale.

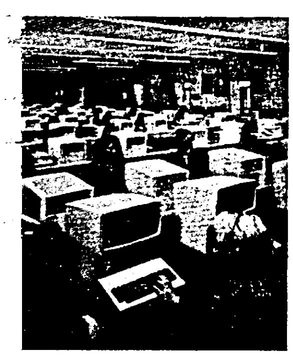


For authorisations phone Southend (0702)352222.

ays take the customer's card with you when you make an authorisation call.

ocedure

all Authorisation tre 8 a.m.–10 p.m. nday to Saturday usive.



2 The operator will ask you:
What is the cardholder number?
It's on the card. Please quote the FULL number.
What is your retailer number?
It's in the front of your Retailers Guide.
What is the amount ?
Quote to the nearest £ above.
3 If authorised, write Authorisation number on sales voucher.
4 If declined, you will be advised what to do next.

Completing the Sales Voucher

1 Use only Access Stationery.

2 Place card face upwards in the imprinter, place sales voucher over the card also face upwards, with the bottom left corner beneath the triangular voucher guide.

3 Pull the imprinter handle from left to right and return.

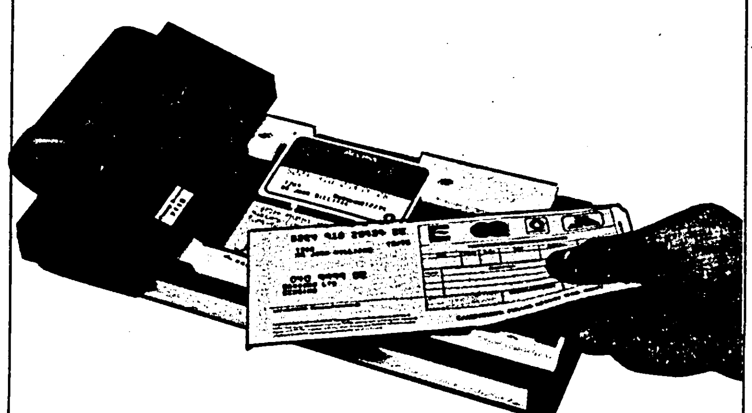
4 Remove sales voucher and, writing firmly with a ball point pen, enter the date, sales description and total of sale.

5 Customer signs the voucher with a ball point pen – check signature with the card.

6 Enter authorisation number if needed.

7 Check that details are legible on all four copies of the voucher.

8 Hand top copy to the customer with his card.



We will pay a reward of £25 where a void or stolen card is recovered.

1. Why do the processes need rigidly controlled conditions? **108**

- [ ] A. To melt the steel quicker
- [ ] B. To replace the Bessemer processes
- [ ] C. To make different sorts of alloys
- [ ] D. To reduce impurities to a minimum
- [ ] E. To treat pig iron

2. What is added to the furnace to remove carbon? **109**

- [ ] A. Three carbon electrodes
- [ ] B. Silicon chips
- [ ] C. Phosphorus
- [ ] D. Non-metallic inclusions
- [ ] E. Iron ore

3. What happens <u>after</u> the slag composed of lime, fluospar and carbon is added?

- [ ] A. The temperature is checked
- [ ] B. A sample of steel is analysed
- [ ] C. The slag is raked off
- [ ] D. The furnace is tilted
- [ ] E. The metal is tapped into a ladle

ELECTRIC-FURNACE STEEL

For high grade alloy steels such as are used for many cutting tools, die steels, and stainless steel, it is necessary that the metal shall be refined and melted under rigidly controlled conditions and in such a way that impurities are reduced to a minimum. Where fuel is burnt in the furnace some contamination is unavoidable; and this led steel-makers to realize that electric melting was likely to be technically more desirable, even though more costly, than the methods of the open-hearth and Bessemer processes. The electric furnace is intended chiefly to refine and produce alloy steels of good quality. Pig iron is not directly treated in the electric furnaces, though sometimes it is partly refined in an open-hearth furnace and then transferred to electric furnaces for final treatment and alloying.

The hearth can be either acid or basic lined, according to needs. Acid furnaces are mainly used in steel foundries and are rarely of more than 10 tons' capacity. Basic furnaces of up to 80 tons' capacity are now used for making alloy and special steels; they are becoming increasingly used even for common steels, when they may be as big as 150 tons.

The bottom of the furnace is covered with lime; scrap steel of known quality is then put inside. Next the three carbon electrodes are automatically lowered to the surface of the metal and melting begins. When melting is complete the slag will have already removed much of the silicon, manganese, and phosphorus from the molten scrap. Iron ore is added for the removal of carbon, the remainder of the phosphorus, and non-metallic inclusions.

Next the furnace is tilted and the slag raked off into the slag ladle It is replaced by a slag composed of lime, fluospar, and carbon which removes sulphur from steel. A sample of steel is analysed and adjustments to the composition are made by adding ferro-alloys. Finally the temperature is checked, the furnace tilted, and the metal tapped into the ladle

1. Who gives the first warning?

- [ ] A Departmental Manager
- [ ] B Personnel Manager 094
- [ ] C Training Officer
- [ ] D Foreman
- [ ] E Shop Steward

2. When will a written warning be sent?

- [ ] A At the end of the probationary period
- [ ] B After 3 weeks 095
- [ ] C When no other way is possible
- [ ] D After 8 weeks
- [ ] E If the offence continues

3. At which stage is the employee dismissed?

- [ ] A Stage 1
- [ ] B Stage 3 096
- [ ] C Stage 4
- [ ] D Stage 2
- [ ] E The employee cannot be dismissed

DISCLIPLINARY PROCEDURE FOR LATENESS

Nothing in this procedure shall apply to employees with less than eight weeks continuous service. Such employees shall be considered to be serving a probationary period.

When an employee persistently comes late the following procedure will be adopted:-

Stage 1

A verbal warning will be administered by the Departmental Manager in the presence of the employee's shop steward. This will be recorded and dated in writing, a copy being sent to the Personnel Department.

Stage 2

If the offence continues, the employee will receive a written warning indicating the rule being breached and asking for improvement in timekeeping. Copies of the written warning will be sent to the Personnel Department, the employee's shop steward and the chairman of the shop steward's committee.

Stage 3

If bad timekeeping continues, the employee will receive a final written warning stating that breaking the rule again will result in dismissal. Copies of this final warning will be sent to the Personnel Department, the employee's shop steward and the chairman of the shop steward's committee.

Stage 4

Dismissal. Notice will be given in accordance with the terms of the Contract of Employment Act 1972 as amended by the Employment Protection Act 1975.

1. What will happen if drawings are left in strong sunlight?

- [ ] A They will crinkle at the edges
- [ ] B Coffee may be spilt on them
- [ ] C They will be torn
- [ ] D They will fade
- [ ] E The ink will run

047

2. What sort of drawings show the details of individual manufacture?

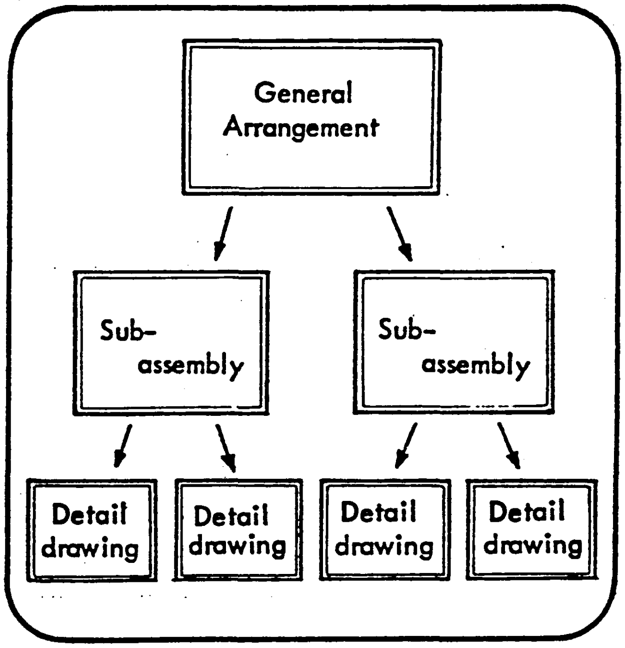
- [ ] A Sub-Assembly drawings
- [ ] B General arrangement drawings
- [ ] C Detail drawings
- [ ] D All of these
- [ ] E None of these

048

3. What are the most common methods of reproducing drawings?

- [ ] A Photocopying and Offset Litho
- [ ] B Dye line and Photocopying
- [ ] C Dye line and Photogravure
- [ ] D Lithography and Dye line
- [ ] E Photocopying and Photogravure

049



Engineering drawings give information about shapes, dimensions, surface finishes, materials, assemblies and connections of components to enable the reader to understand what has to be made.

Drawings
Main forms of drawing are:

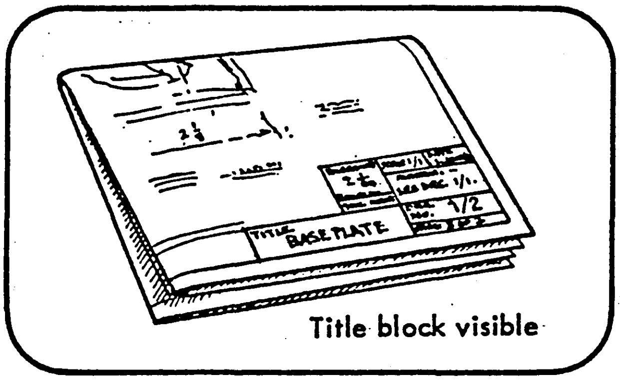
1. General Arrangement drawings which show the complete arrangement of all the components.

2. Sub-Assembly drawings which show in greater detail the way parts are assembled.

3. Detail drawings which show the details of individual manufacture.

4. Circuit Diagrams.

5. Wiring Diagrams.



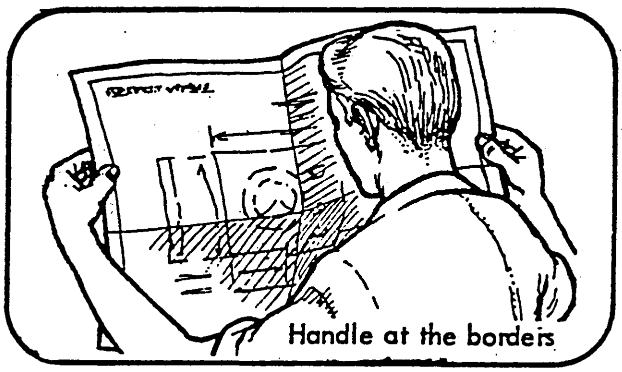
Title block visible

Drawings and diagrams are usually numbered by starting with the number or code of the General Arrangement, and progressively adding other numbers to it.

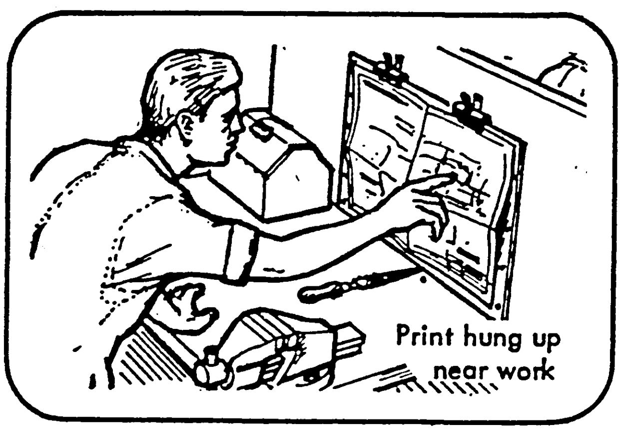
The most common methods of reproducing drawings are:

Dye Line
Photocopying

When made, duplicate prints should be folded to leave the title block and drawing number visible.


Handle at the borders

Before using a drawing, make sure that it is the correct one. Sheets should be handled by the borders and should not be left in strong sunlight or they will fade.


Print hung up near work

It is good practice to hang the print near the work for easy reference.

To read engineering drawings correctly, a knowledge of drawing conventions is essential. These are taught in the Technical College course. Careful attention must be paid to B.S. 308 which is the British Standard for engineering drawings.

1. Where would you file this letter? 002

- [ ] A. Under 'Moulding & Turners Ltd'
- [ ] B. Under 'I. S. Ingrams'
- [ ] C. Under 'Steel & Co Ltd'
- [ ] D. Under 'Ironco Ltd'
- [ ] E. Under 'London'

2. Who would you pass this letter on to?

- [ ] A. Mr Sykes, Company Secretary
- [ ] B. Mr Pierce, Production Manager 003
- [ ] C. Mr Jones, Dispatch Manager
- [ ] D. Mrs Pascalle, Chief Clerical Officer
- [ ] E. Mr Peters, Research Officer

3. Who sent the letter?

- [ ] A. Production Manager
- [ ] B. I S Ingrams
- [ ] C. S J Marshall
- [ ] D. Mr T Jones 004
- [ ] E. None of these

STEEL & CO LTD

247 Newcliffe Road,
Sheffield,
S9 2ZL

Telephone: Sheffield (STD Code 0742)
440010

5 February 1979

Moulding and Turners Ltd
97 Stoke Lane
Clapham
London
SW17 9QT

for the attention of the Production Manager

Dear Sir,

Thank you for your recent letter concerning the quality of billets in our last consignment to you.

Regrettably, we shall be unable to supply replacements for at least three weeks, owing to forging problems, which were the cause of the original faults in the billets.

I have therefore arranged for our parent company, Ironco Ltd, to supply the replacements. Mr Ingrams of that firm will contact you this week.

I hope these arrangements are satisfactory.

Yours faithfully,

S J Marshall

S J Marshall
Production Control Manager

c.c. I S Ingrams, Ironco Ltd

1. What must you do if you have to leave your selling position?

- [ ] A. Remove all cash from the till
- [ ] B. Tell your supervisor
- [ ] C. Lock your cash register
- [ ] D. Turn all customers away
- [ ] E. None of these

2. "Respect the cash drawers of your colleagues" means

- [ ] A. Cash registers are very expensive
- [ ] B. Don't interrupt them whilst they are selling
- [ ] C. Be polite to your fellow workers
- [ ] D. Don't ask them to neglect their tills
- [ ] E. None of these

3. Which of these statements is <u>untrue</u>?

- [ ] A. A gift voucher is equivalent to cash
- [ ] B. A Refund slip is a Cash Register document
- [ ] C. You should keep your Locking Key chained to your overall
- [ ] D. You should never leave money unattended
- [ ] E. You should wrap the goods first of all

CARE WITH CASH

Follow these sensible rules for looking after cash:

Register the sale, accept customers money, give any change required and close the cash drawer before wrapping the goods. CASH BEFORE WRAP!

Present the cash register ticket to the customer, if it is not parcelled with the goods.

Always close the cash register drawer on completion of sale.

Never leave money lying about unattended. It can vanish!

Unattended cash registers must be kept locked. If you leave your selling position for any reason lock the register first, using the Read or Locking Key.

Always keep this key attached to your overall with the chain provided.

Respect the cash drawers of your colleagues.

Ensure that they respect yours. You are responsible for the cash and cash equivalent documents in your care, until they are handed over, checked and accepted by your Manager or the official Cashier. Keep nothing in your cash drawer but cash, equivalent documents and Cash Register Documents. Documents equivalent to cash are Cheques, Gift Vouchers, Manufacturer's Redeemed Coupons, Access and Barclaycard Sales Vouchers and I.O.U.'s issued by the Cash Office. C ash Register documents are Refund Slips, Cancelled and No Salestickets.

At cashing up time the cash and all these documents are removed and the total value of each is entered in the appropriate space on the Assistants Cash Slip.

1. What should you do before setting the time clock to zero?

- [ ] A. Depress RH lever to set to zero
- [ ] B. Do not force the winder 1
- [ ] C. See that the clock is wound up
- [ ] D. Refit the boiler lid
- [ ] E. None of these

2. How do you start the clock? 13

- [ ] A. Switch on the heater
- [ ] B. Press the right-hand lever
- [ ] C. Lift the LH lever
- [ ] D. Lift the right-hand lever
- [ ] E. Wind up the clock

3. How many thermometer readings must you take each time?

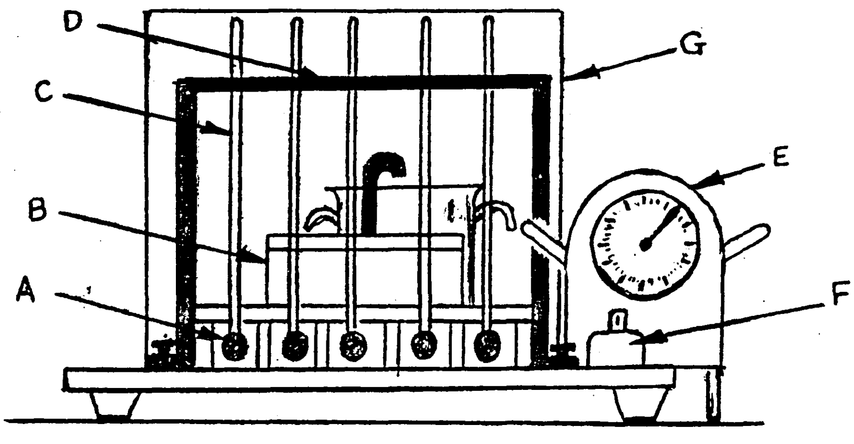
- [ ] A. Readings at 6,8, 10, 12, 14, 16, 18, 20 minutes
- [ ] B. 5 readings
- [ ] C. 45 readings in all 132
- [ ] D. After 4 minutes
- [ ] E. None of these

TOPIC: Thermal Conductivity (Heat Conducted)

AIM: To compare the thermal conductivities of 5 metal rods, i.e. Aluminium, Brass, Cast Iron, Copper and Mild Steel.

Equipment Required on Shelf No. 27

A copper boiler with the 5 metal rods attached and fitted to a baseboard. The boiler is heated by an electric element clamped underneath. The bulbs of five thermometers are placed in sockets drilled bear the outer end of each rod and a bridge supports the top ends of the thermometers. A time clock and switch are also attached to the baseboard.



The Apparatus Set Up for Use

A. 5 metal rods
B. Boiler
C. Thermometers
D. Bridge
E. Time Clock
F. Switch
G. Perspex guard

PROCEDURE

Complete the questions or tasks on your report sheet by following the procedure set out below: PLEASE READ CAREFULLY BEFORE STARTING WORK.

1. Set up the apparatus facing you as shown in the sketch, taking great care not to damage the thermometers when moving the apparatus.
2. Remove the boiler top carefully and fill the boiler up to the rivet head (inside the curved end) with COLD water, using the copper can with the handles. Refit the boiler lid and place the copper can under the steam outlet to catch the water droplets.
3. See that the time clock is wound up (do not force the winder) and check that it is set to zero. (Depress RH lever to set zero).
4. See that the switch is OFF and plug the lead into a mains socket.
5. After a few minutes read the 5 thermometers to the NEAREST DEGREE from LH to RH and enter the 5 readings on the 0 line in the table on your report and under the name of the metal concerned.
6. Start the clock (lift RH lever) and SWITCH ON the heater.
7. Wait 4 minutes, take the 5 thermometer readings as before (from LH to RH) and enter them correctly on the 4 minute line in the table.
8. Repeat the readings at 6, 8, 10, 12, 14, 16, 18 and 20 minutes.
9. Now switch off, pull out the mains plug, stop the clock (depress L/hand lever) and leave the apparatus to cool while you complete your lab. sheet.

1. What would you do first of all?

- [ ] A. Fix 4 plastic surface boxes to a workboard 127
- [ ] B. Wire 4 -13A sockets on a ring circuit
- [ ] C. Join all the boxes together
- [ ] D. Draw a circuit diagram
- [ ] E. List the readings obtained

2. What would you do after connecting the red, black and green/yellow wires? 128

- [ ] A. Carry out insulation tests
- [ ] B. Have your work checked
- [ ] C. List the readings obtained
- [ ] D. Strip the cable
- [ ] E. Run back to the fuse board

3. 'Twin and earth cable' has: 129

- [ ] A. Two wires : red and earth
- [ ] B. Two wires : red and neutral
- [ ] C. Three wires : red, neutral and fuse
- [ ] D. Three wires : red, neutral and earth
- [ ] E. None of these

Object: To wire, in Twin and earth cable, 4-13A sockets on a ring circuit.

Note: Draw a circuit diagram before starting any work.

1. Fix 4 plastic surface boxes to a workboard.

2. Fix a 2-way fuseboard to the centre of the workboard.

3. Run $2.5mm^2$ Twin and earth cable from the fuse board to the nearest box.

4. Join all the boxes together in this way.

5. From the last socket, run back to the fuse board.

6. Strip the cable and fix the sockets.

7. Connect the cable into a) the same fuse (red core)
   b) the neutral block (black core)
   c) the earth block (green/yellow sleeved earth)

   in the fuse board.

8. Connect 1 metre of $6mm^2$ double sheathed Red, 1 metre $6mm^2$ double sheathed Black and 1 metre of $6mm^2$ Green/Yellow to the incoming side of the fuse board.

9. Have your work checked.

10. Use an insulation and resistance tester and carry out the necessary tests as prescribed in the I.E.E. Regs.

11. List the readings obtained.

APPENDIX VIII

EDINBURGH READING TEST, FORM 4

EXTRAPOLATED QUOTIENTS

EXTRAPOLATED QUOTIENTS : EPT4

SCORE	16. 1	16. 2	16. 3	16. 4	16. 5	16. 6	16. 7	16. 8	16. 9	16.10	16.11	17. 0
1	70	70	70	70	70	70	70	70	70	70	70	70
2	70	70	70	70	70	70	70	70	70	70	70	70
3	70	70	70	70	70	70	70	70	70	70	70	70
4	70	70	70	70	70	70	70	70	70	70	70	70
5	70	70	70	70	70	70	70	70	70	70	70	70
6	70	70	70	70	70	70	70	70	70	70	70	70
7	70	70	70	70	70	70	70	70	70	70	70	70
8	70	70	70	70	70	70	70	70	70	70	70	70
9	70	70	70	70	70	70	70	70	70	70	70	70
10	70	70	70	70	70	70	70	70	70	70	70	70
11	70	70	70	70	70	70	70	70	70	70	70	70
12	70	70	70	70	70	70	70	70	70	70	70	70
13	70	70	70	70	70	70	70	70	70	70	70	70
14	70	70	70	70	70	70	70	70	70	70	70	70
15	70	70	70	70	70	70	70	70	70	70	70	70
16	70	70	70	70	70	70	70	70	70	70	70	70
17	70	70	70	70	70	70	70	70	70	70	70	70
18	70	70	70	70	70	70	70	70	70	70	70	70
19	70	70	70	70	70	70	70	70	70	70	70	70
20	70	70	70	70	70	70	70	70	70	70	70	70
21	70	70	70	70	70	70	70	70	70	70	70	70
22	70	70	70	70	70	70	70	70	70	70	70	70
23	70	70	70	70	70	70	70	70	70	70	70	70
24	70	70	70	70	70	70	70	70	70	70	70	70
25	70	70	70	70	70	70	70	70	70	70	70	70
26	70	70	70	70	70	70	70	70	70	70	70	70
27	70	70	70	70	70	70	70	70	70	70	70	70
28	70	70	70	70	70	70	70	70	70	70	70	70
29	70	70	70	70	70	70	70	70	70	70	70	70
30	70	70	70	70	70	70	70	70	70	70	70	70
31	70	70	70	70	70	70	70	70	70	70	70	70
32	70	70	70	70	70	70	70	70	70	70	70	70
33	70	70	70	70	70	70	70	70	70	70	70	70
34	70	70	70	70	70	70	70	70	70	70	70	70
35	70	70	70	70	70	70	70	70	70	70	70	70
36	70	70	70	70	70	70	70	70	70	70	70	70
37	70	70	70	70	70	70	70	70	70	70	70	70
38	70	70	70	70	70	70	70	70	70	70	70	70
39	70	70	70	70	70	70	70	70	70	70	70	70
40	71	70	70	70	70	70	70	70	70	70	70	70
41	71	71	71	70	70	70	70	70	70	70	70	70
42	72	72	71	71	71	71	71	70	70	70	70	70
43	72	72	72	72	72	72	71	71	71	71	71	71
44	73	73	73	72	72	72	72	72	71	71	71	71
45	74	74	73	73	73	73	73	73	72	72	72	72
46	74	74	74	74	74	73	73	73	73	73	73	72
47	75	75	74	74	74	74	74	74	73	73	73	73
48	75	75	75	75	75	74	74	74	74	74	74	73
49	76	75	75	75	75	75	75	74	74	74	74	74
50	76	76	76	75	75	75	75	75	75	74	74	74
51	77	76	76	76	76	76	76	75	75	75	75	75
52	77	77	77	76	76	76	76	76	75	75	75	75
53	77	77	77	77	77	77	76	76	76	76	76	76
54	78	78	77	77	77	77	77	77	76	76	76	76
55	78	78	78	78	78	78	77	77	77	77	77	77
56	79	79	79	79	79	78	78	78	78	78	78	77
57	79	79	79	79	78	78	78	78	78	78	77	77
58	80	79	79	79	79	79	79	78	78	78	78	78
59	80	80	80	79	79	79	79	79	79	78	78	78
60	80	80	80	80	80	79	79	79	79	79	79	78
61	81	81	80	80	80	80	80	80	79	79	79	79
62	81	81	81	81	80	80	80	80	80	79	79	79
63	82	81	81	81	81	81	80	80	80	80	80	79
64	82	82	81	81	81	81	81	80	80	80	80	80
65	82	82	82	82	81	81	81	81	81	80	80	80

SCORE	16. 1	16. 2	16. 3	16. 4	16. 5	16. 6	16. 7	16. 8	16. 9	16.10	16.11	17. 0
66	83	82	82	82	82	82	81	81	81	81	81	81
67	83	83	83	82	82	82	82	82	81	81	81	81
68	83	83	83	83	83	82	82	82	82	82	81	81
69	84	83	83	83	83	83	82	82	82	82	82	81
70	84	84	84	83	83	83	83	83	82	82	82	82
71	84	84	84	84	84	83	83	83	83	82	82	82
72	85	84	84	84	84	84	83	83	83	83	82	82
73	85	85	85	84	84	84	84	84	83	83	83	83
74	86	85	85	85	85	84	84	84	84	84	83	83
75	86	86	85	85	85	85	84	84	84	84	84	83
76	86	86	86	85	85	85	85	84	84	84	84	83
77	86	86	86	86	86	85	85	85	85	84	84	84
78	87	87	86	86	86	86	85	85	85	85	84	84
79	87	87	87	86	86	86	86	86	85	85	85	85
80	88	87	87	87	87	86	86	86	86	85	85	85
81	88	87	87	87	87	86	86	86	86	85	85	85
82	88	88	88	87	87	87	87	86	86	86	86	85
83	89	88	88	88	87	87	87	87	86	86	86	86
84	89	89	88	88	88	88	87	87	87	87	86	86
85	89	89	89	88	88	88	88	87	87	87	87	86
86	90	89	89	89	88	88	88	88	87	87	87	87
87	90	90	90	89	89	89	88	88	88	88	87	87
88	90	90	90	89	89	89	89	88	88	88	88	87
89	91	90	90	90	90	89	89	89	89	88	88	88
90	91	91	90	90	90	89	89	89	89	88	88	88
91	92	91	91	91	90	90	90	90	89	89	89	89
92	92	91	91	91	91	90	90	90	89	89	89	89
93	92	92	92	91	91	91	90	90	90	90	89	89
94	93	92	92	92	91	91	91	91	90	90	90	89
95	93	93	92	92	92	91	91	91	91	90	90	90
96	93	93	93	92	92	92	91	91	91	90	90	90
97	94	93	93	93	92	92	92	92	91	91	91	90
98	94	94	93	93	93	92	92	92	91	91	91	91
99	94	94	94	93	93	93	92	92	92	92	91	91
100	95	94	94	94	93	93	93	92	92	92	91	91
101	95	95	94	94	94	93	93	93	92	92	92	91
102	95	95	95	94	94	94	93	93	93	92	92	92
103	96	95	95	95	94	94	94	93	93	93	92	92
104	96	96	95	95	95	94	94	94	93	93	93	92
105	96	96	96	95	95	95	94	94	94	93	93	93
106	97	96	96	96	95	95	95	94	94	94	93	93
107	97	97	96	96	96	95	95	95	94	94	94	93
108	98	97	97	96	96	96	95	95	95	94	94	94
109	98	98	97	97	96	96	96	95	95	95	94	94
110	98	98	98	97	97	96	96	96	95	95	94	94
111	99	98	98	98	97	97	96	96	96	95	95	95
112	99	99	98	98	97	97	97	96	96	95	95	95
113	100	99	99	98	98	98	97	97	96	96	96	95
114	100	100	99	99	98	98	98	97	97	96	96	95
115	101	100	100	99	99	98	98	98	97	97	96	96
116	101	101	100	100	99	99	98	98	98	97	97	96
117	101	101	101	100	100	99	99	98	98	98	97	97
118	102	102	101	101	100	100	99	99	99	98	98	97
119	102	102	102	101	101	100	100	99	99	98	98	98
120	103	103	102	102	101	101	100	100	99	99	99	98
121	103	103	102	102	102	101	101	100	100	99	99	98
122	104	103	103	103	102	102	101	101	100	100	99	99
123	105	104	104	103	103	102	102	101	101	100	100	99
124	105	105	104	104	103	103	102	102	101	101	100	100
125	106	105	105	104	104	103	103	102	102	101	101	100
126	106	106	105	105	104	104	103	103	102	102	101	101
127	107	106	106	105	105	104	104	103	103	102	101	101
128	107	107	106	106	105	105	104	103	103	102	102	101
129	108	107	107	106	106	105	105	104	104	103	102	102
130	108	108	107	107	106	106	105	104	104	103	103	102

SCORE	16. 1	16. 2	16. 3	16. 4	16. 5	16. 6	16. 7	16. 8	16. 9	16.10	16.11	17. 0
131	109	108	108	107	107	106	106	105	104	104	103	103
132	110	109	109	108	107	107	106	106	105	104	104	103
133	111	110	109	109	108	107	107	106	106	105	104	104
134	111	111	110	109	109	108	107	107	106	106	105	104
135	112	111	111	110	109	109	108	107	107	106	105	105
136	113	112	111	111	110	109	109	108	107	107	106	105
137	114	113	112	112	111	110	110	109	108	107	107	106
138	115	114	114	113	112	112	111	110	109	109	108	107
139	117	116	116	115	114	114	113	112	112	111	110	110
140	119	118	118	117	116	116	115	115	114	113	113	112
141	121	120	120	119	119	118	118	117	116	116	115	115
142	123	122	122	122	121	121	120	120	120	119	119	118
143	125	125	124	124	124	124	123	123	123	122	122	122
144	127	127	126	126	126	126	126	126	125	125	125	125
145	128	128	128	128	128	128	128	128	128	128	127	127
146	130	130	129	129	129	129	129	129	129	129	129	129
147	130	130	130	130	130	130	130	130	130	130	130	130
148	130	130	130	130	130	130	130	130	130	130	130	130
149	130	130	130	130	130	130	130	130	130	130	130	130
150	130	130	130	130	130	130	130	130	130	130	130	130

APPENDIX I

TABLES OF DATA

Tables of data are numberes in sequential order within Chapters. Hence, Table 5.3 is the third Table within the set of tables for Chapter 5. Each set of Tables is separated by a sheet specifying the relevant chapter.

INDUSTRY	Apprentice	Profess'l	Clerical	Operative	Others	Totals
Agriculture	17	-	-	1	22	40
Mining	35		2	1		38
Food, Drink & Tobacco	8	1	1	3	36	49
Chemical & Allied	4		1		2	7
Metal Manufacture	245	8	34	11	35	333
Engineering	115	2	2	9	13	141
Electrical Goods	36	1	2	8	2	49
Engineers Machine Tools	147	2	13	9	28	199
Vehicle Manufacture	4			2	1	7
Hand Tools	96	1	7	10	18	132
Cutlery	69	2	12	11	93	187
Other Metal Goods	87		4	4	12	107
Textiles	10			9	3	22
Clothing	4			2		6
Bricks, Pottery & Glass	6	3	4	2	14	29
Timber & Furniture	39		3	11	18	71
Paper, Printing & Publishing	30	1	3	3	9	46
Other Manufacture	1			1	1	3
Construction	504	15	8	2	31	560
Gas, Electricity & Water	36		3	2		41
Transport & Communications	32		7	4	5	48
Distributive	55	2	28	74	189	348
Insurance, Banking & Finance	5	3	44	2		54
Professional & Scientific Services	8	39	8	4	4	63
Hairdressing	10			2	1	13
Motor Repairs	166	1	6	6	17	196
Miscellaneous Services	42			11	37	90
Public Administration	85	16	22	62	6	191
Totals	1,896	97	214	266	597	3,070

Table 5.1: Analysis of Employment (Boys) 1.10.73 to 30.9.74

INDUSTRY	Apprentice	Profess'l	Clerical	Operative	Others	Totals
Agriculture	1				3	4
Mining			2			2
Food, Drink & Tobacco	3		41	3	38	85
Chemical & Allied			7		6	13
Metal Manufacture	7		124	1	9	141
Engineering	3		53		7	63
Electrical Goods	1		13	3	8	25
Engineers Machine Tools	3		78	2	13	96
Vehicle Manufacture						
Hand Tools	1		54		14	69
Cutlery	1		36		81	118
Other Metal Goods	1		30	1	10	42
Textiles			1	1		2
Clothing	1		1	84	11	97
Bricks, Pottery & Glass	1	1	6		1	9
Timber & Furniture			7	3	4	14
Paper, Printing & Publishing	1		18	6	31	56
Other Manufacture			4		4	8
Construction	4	1	39		1	45
Gas, Electricity & Water	2		9			11
Transport & Communications			25	1	1	27
Distributive	18		157	221	354	750
Insurance, Banking & Finance		8	296	1		305
Professional & Scientific Services	6	105	145	52	5	313
Hairdressing	113		2	3	2	120
Motor Repairs						
Miscellaneous Services	7		84	8	68	167
Public Administration	4	1	90	21	1	117
Totals	178	116	1,322	411	672	2,699

Table 5.2: Analysis of Employment (Girls) 1.10.73 to 30.9.74

INDUSTRY	Apprentice	Profess'l	Clerical	Operative	Others	Totals
Agriculture	1	1			10	12
Mining	28		1	1	3	33
Food, Drink & Tobacco	7		1	5	26	39
Chemical & Allied	1		1	1		3
Metal Manufacture	224	6	23	25	16	294
Engineering	95		1	21	8	125
Electrical Goods	8			10		18
Engineers Machine Tools	121	6	10	18	10	165
Vehicle Manufacture	5			3		8
Hand Tools	112	1	11	31	24	179
Cutlery	57		4	40	74	175
Other Metal Goods	81		4	24	11	120
Textiles	7			4	3	14
Clothing	3		1		1	5
Bricks, Pottery & Glass	6		1	3	5	15
Timber & Furniture	29	1	2	8	13	53
Paper, Printing & Publishing	12		1	7	4	24
Other Manufacture	9		1	3	2	15
Construction	299	12	11	3	34	359
Gas, Electricity & Water	20	2	4	3		29
Transport & Communications	25	2	4	11	10	52
Distributive	101	7	24	158	200	490
Insurance, Banking & Finance	1	16	12	2	1	32
Professional & Scientific Services	14	20	12	7	1	54
Hairdressing	11					11
Motor Repairs	111		2	6	20	139
Miscellaneous Services	23	1	1	11	22	58
Public Administration	61	40	34	44	7	186
Totals	1,472	115	166	449	505	2,707

Table 5.3: Analysis of Employment 'Boys' 1.10.74 to 30.9.75

INDUSTRY	Apprentice	Profess'l	Clerical	Operative	Others	Totals
Agriculture		1	1			2
Mining			5			5
Food, Drink & Tobacco	2		12	10	67	91
Chemical & Allied			5	1	10	16
Metal Manufacture		1	87	2	6	96
Engineering	1		38	14	3	56
Electrical Goods			3	2	3	8
Engineers Machine Tools	1		41		6	48
Vehicle Manufacture				1		1
Hand Tools	1		56	2	24	83
Cutlery			40	7	80	127
Other Metal Goods			17	2	9	28
Textiles	1		6	10		17
Clothing	1		5	78	4	88
Bricks, Pottery & Glass			2	1		3
Timber & Furniture			4	8	5	17
Paper, Printing & Publishing	2		11	3	18	34
Other Manufacture			2		5	7
Construction			63		1	64
Gas, Electricity & Water			6			6
Transport & Communications			21	3	2	26
Distributive	16	1	116	269	264	666
Insurance, Banking & Finance		10	111	2		123
Professional & Scientific Services	7	81	128	49	9	274
Hairdressing	73	1	5	3	1	83
Motor Repairs			30	1	5	36
Miscellaneous Services	7	1	23	13	53	97
Public Administration	5	6	62	7	2	82
Totals	117	102	900	488	577	2.184

Table 5.4: Analysis of Employment (Girls) 1.10.74 to 30.9.75

INDUSTRY	Apprentice	Profess'l	Clerical	Operative	Others	Totals
Agriculture	4	2		2	12	20
Mining	20		2	3	3	28
Food, Drink & Tobacco	6	1	3	11	34	55
Chemical & Allied	1	2	1	2	1	7
Metal Manufacture	246	11	21	76	28	382
Engineering	66	1	2	31	10	110
Electrical Goods	11			12	1	24
Engineers Machine Tools	125	6	8	52	17	208
Vehicle Manufacture	3			6	1	10
Hand Tools	55	3	6	56	24	144
Cutlery	29		4	74	85	192
Other Metal Goods	66	1	2	52	24	145
Textiles	1		1	4	4	10
Clothing				5		5
Bricks, Pottery & Glass	3		3	8	5	19
Timber & Furniture	14	1		22	15	52
Paper, Printing & Publishing	8	2		6	4	20
Other Manufacture	5		2	8	4	19
Construction	268	9	6	21	73	377
Gas, Electricity & Water	25	3	2	3		33
Transport & Communications	23		5	8	6	42
Distributive	65	14	19	141	215	454
Insurance, Banking & Finance	3	11	10	4	3	31
Professional & Scientific Services	10	12	13	4	4	43
Hairdressing	5			1		6
Motor Repairs	75	1	1	8	30	115
Miscellaneous Services	32		2	12	25	71
Public Administration	65	27	29	45	9	175
Totals	1,234	107	142	677	637	2,797

Table 5.5: Analysis of Employment (Boys) 1.10.75 to 30.9.76

INDUSTRY	Apprentice	Profess'l	Clerical	Operative	Others	Totals
Agriculture			1		1	2
Mining			3			3
Food, Drink & Tobacco			30	27	45	102
Chemical & Allied			4	2	5	11
Metal Manufacture	1	3	85	10	4	103
Engineering		1	42	6	7	56
Electrical Goods			5	2	2	9
Engineers Machine Tools	2	1	48	2	12	65
Vehicle Manufacture			1			1
Hand Tools	1		32	4	17	54
Cutlery			36	12	121	169
Other Metal Goods	2		14	9	12	37
Textiles			3	13	6	22
Clothing	1	1	3	74	12	91
Bricks, Pottery & Glass			11	1	1	13
Timber & Furniture	4		6	3	5	18
Paper, Printing & Publishing	1		14	9	6	30
Other Manufacture			1	3	4	8
Construction			34		1	35
Gas, Electricity & Water			8			8
Transport & Communications		1	27	1	1	30
Distributive	11	3	132	227	364	737
Insurance, Banking & Finance	1	8	103	3	3	118
Professional & Scientific Services	19	58	114	63	23	277
Hairdressing	91	2	1	6	1	101
Motor Repairs		1	29	2	3	35
Miscellaneous Services	4	2	10	19	81	116
Public Administration	6	10	53	5	3	77
Totals	144	91	850	503	740	2,328

Table 5.6: Anal sis of Em lo

INDUSTRY	Apprentice	Profess'l	Clerical	Operative	Others	Totals
Agriculture	3			6	20	29
Mining	27	1	1	5	5	39
Food, Drink & Tobacco	9		3	16	36	64
Chemical & Allied	2		2	1		5
Metal Manufacture	214	11	27	95	27	374
Engineering	57	2	7	31	4	101
Electrical Goods	11	1	1	9	4	26
Engineers Machine Tools	79		6	64	10	159
Vehicle Manufacture	1		1	1	1	4
Hand Tools	48	2	4	66	28	148
Cutlery	19	1	7	77	125	229
Other Metal Goods	135	2	12	85	53	287
Textiles	1			2		3
Clothing	2		1	3	6	12
Bricks, Pottery & Glass	4	1	2	4	7	18
Timber & Furniture	9		2	24	8	43
Paper, Printing & Publishing	4		4	14	4	26
Other Manufacture	3			5	1	9
Construction	248	10	8	53	50	369
Gas, Electricity & Water	20	1	6	7		34
Transport & Communications	33		3	7	5	48
Distributive	21	3	11	155	232	422
Insurance, Banking & Finance	10	10	15	5		40
Professional & Scientific Services	5	7	8	11	7	38
Hairdressing	4					4
Motor Repairs	124		4	39	19	186
Miscellaneous Services	13		10	20	28	71
Public Administration	64	43	13	45	20	185
Totals	1,170	95	158	850	700	2,973

Table 5.7: Analysis of Employment (Boys) 1.10.76 to 30.9.77

INDUSTRY	Apprentice	Profess'l	Clerical	Operative	Others	Totals
Agriculture				2	1	3
Mining			2			2
Food, Drink & Tobacco		1	21	15	92	129
Chemical & Allied			4			4
Metal Manufacture	3	2	86	9	13	113
Engineering			31	3	4	38
Electrical Goods		1	11	5	11	28
Engineers Machine Tools	1		30	2	14	47
Vehicle Manufacture		1	3			4
Hand Tools	1		29	10	42	82
Cutlery	1		24	25	155	205
Other Metal Goods	2		41	17	38	98
Textiles		1	1	4	5	11
Clothing			4	56	37	97
Bricks, Pottery & Glass			8		1	9
Timber & Furniture			3	9	18	30
Paper, Printing & Publishing			9	3	18	30
Other Manufacture			1	3	15	19
Construction	2		47	1		50
Gas, Electricity & Water			13			13
Transport & Communications		1	25	4		30
Distributive	9	5	144	470	354	787
Insurance, Banking & Finance		8	138	6	2	154
Professional & Scientific Services	15	69	110	30	11	235
Hairdressing	80		4	8		92
Motor Repairs	2		33	1	7	43
Miscellaneous Services	8	2	25	33	54	122
Public Administration	4	6	35	10	7	62
Totals	128	97	882	526	904	2,537

Table 5.8: Analysis of Employment (Girls) 1.10.76 to 30.9.77

Industry	Apprentice-ship	Profes-sional	Clerical	Operative Training	Others	Total	%	Index 73/74 = 100	Index 74/75 = 100	Index 75/76 = 100
Agriculture	3	-	-	8	21	32	0.58	76	200	134
Mining	27	1	3	5	5	41	0.74	107	95	123
Food, Drink and Tobacco	9	1	24	31	128	193	3.52	152	132	115
Chemical & Allied Industry	2	-	6	1		9	0.16	46	41	46
Metal Manufacture	217	13	113	104	40	487	8.84	108	111	93
Engineering	57	2	38	34	8	139	2.52	71	68	77
Electrical Goods	11	2	12	14	15	54	0.98	77	185	153
Vehicles	1	1	4	1	1	8	0.15	125	83	71
Engineering Small Tools	80	-	36	66	24	206	3.74	72	86	70
Hand Tools	49	2	33	76	70	230	4.17	120	78	108
Cutlery	20	1	31	102	280	434	7.88	147	128	112
Other Metal Goods	137	2	53	103	91	385	6.99	271	303	197
Textiles	1	1	1	6	5	14	0.25	60	40	40
Clothing	2	-	5	59	43	109	1.98	111	104	106
Bricks, Pottery and Glass	4	1	10	4	8	27	0.49	74	132	79
Timber and Furniture	9	-	5	33	26	73	1.32	90	92	96
Paper, Printing & Publishing	4	-	13	17	22	56	1.01	57	85	103
Other Manufacturing Industries	3	-	1	8	16	28	0.51	268	113	96
Construction	250	10	55	54	50	419	7.60	72	88	95
Gas, Electricity and Water	20	1	19	7	-	47	0.85	94	118	106
Transport & Communications	33	1	28	11	5	78	1.42	109	89	101
Distribution	20	8	155	425	591	1209	21.94	115	93	94
Insurance, Banking & Finance	10	18	153	11	2	194	3.52	57	110	121
Professional & Scientific Services	20	76	118	41	18	273	4.95	76	74	79
Miscellaneous Services	21	2	35	53	82	193	3.52	79	111	96
Hairdressing	84	-	4	8	-	96	1.74	75	91	83
Motor Repairs	126	-	37	40	26	229	4.16	122	116	142
Public Administration	68	49	23	55	27	247	4.48	84	82	91
Total	1298	192	1040	1376	1604	5510	100.01			

Table 5.9: Analysis of Employment 1976/9 showing trends

	Numbers			Percentages		
	Boys	Girls	Total	Boys	Girls	Total
Apprenticeships	1896	178	2074	61.8	6.6	36.0
Professional	97	116	213	3.1	4.3	3.7
Clerical	214	1322	1536	7.0	49.0	26.6
Operative Training	266	411	677	8.7	15.2	11.7
Others	597	672	1269	19.4	24.9	22.0
Totals	3070	2699	5769	100	100	100

Table 5.10 : Analysis of job-types

1/10/73 to 30/9/74

	Numbers			Percentages		
	Boys	Girls	Total	Boys	Girls	Total
Apprenticeships	1472	117	1589	54.4	5.4	32.5
Professional	115	102	217	4.2	4.7	4.4
Clerical	166	900	1066	6.1	41.2	21.8
Operative Training	449	488	937	16.6	22.3	19.2
Others	505	577	1082	18.7	26.4	22.1
Totals	2707	2184	4891	100	100	100.2

Table 5.11 : Analysis of job-types
1/10/74 to 30/9/75

	Numbers			Percentages		
	Boys	Girls	Total	Boys	Girls	Total
Apprenticeships	1234	144	1378	44.1	6.2	26.9
Professional	107	91	198	3.8	3.9	3.9
Clerical	142	850	992	5.1	36.5	19.4
Operative Training	677	503	1180	24.2	21.6	23.0
Others	637	740	1377	22.8	31.8	26.9
Totals	2797	2328	5125	100	100	100.1

Table 5.12: Analysis of job-types
1/10/75 to 30/9/76

	Numbers			Percentages		
	Boys	Girls	Total	Boys	Girls	Total
Apprenticeships	1170	128	1298	39.4	5.0	23.6
Professional	95	97	192	3.2	3.8	3.5
Clerical	158	882	1040	5.3	34.8	18.9
Operative Training	850	526	1376	28.6	20.7	25.0
Others	700	904	1604	23.5	35.6	29.0
Totals	2973	2537	5510	100	99.9	100.0

Table 5.13 : Analysis of job-types
1/10/76 to 30/9/77

	1974-1975	1975-1976		1976-1977		
	100 = 73/4	100 = 73/4	100 = 74/5	100 = 73/4	100 = 74/5	100 = 7
Apprenticeships	90	75	83	66	73	88
Professionals	119	105	82	95	80	90
Clerical	82	73	89	71	87	97
Operative Training	164	197	120	214	130	109
Others	100	122	122	132	131	108

Table 5.14: Indices of trends in job-types: 1973 to 1977

INDUSTRY	Apprentice	Profess'l	Clerical	Operative	Others	Totals
Agriculture	9.4	-	-	25.0	65.6	100
Mining	65.9	2.4	7.3	12.2	12.2	100
Food, Drink & Tobacco	4.7	0.5	12.4	16.1	66.3	100
Chemical & Allied	22.2	-	66.7	11.1	-	100
Metal Manufacture	44.6	2.7	23.2	21.4	8.2	100.1
Engineering	41.0	1.4	27.3	24.5	5.8	100
Electrical Goods	20.4	3.7	22.2	25.9	27.8	100
Engineers Machine Tools	38.8	-	17.5	32.0	11.7	100
Vehicle Manufacture	12.5	12.5	50.0	12.5	12.5	100
Hand Tools	21.3	0.9	14.3	33.0	30.4	99.9
Cutlery	4.6	0.2	7.1	23.5	64.5	99.9
Other Metal Goods	35.6	0.5	13.8	26.5	23.6	100
Textiles	7.1	7.1	7.1	42.8	35.7	99.8
Clothing	1.8	-	4.6	54.1	39.4	99.9
Bricks, Pottery & Glass	14.8	3.7	37.0	14.8	29.6	99.9
Timber & Furniture	12.3	-	6.8	45.2	35.6	99.9
Paper, Printing & Publishing	7.1	-	23.2	30.4	39.3	100
Other Manufacture	10.7	-	3.6	28.6	57.1	100
Construction	59.7	2.4	13.1	12.9	11.9	100
Gas, Electricity & Water	42.6	2.1	40.4	14.9	-	100
Transport & Communications	42.3	1.3	35.9	14.1	6.4	100
Distributive	2.5	0.7	12.8	35.2	48.9	100.1
Insurance, Banking & Finance	5.2	9.3	78.9	5.7	1.0	100.1
Professional & Scientific Services	7.3	27.8	43.2	15.0	6.6	99.9
Hairdressing	87.5	-	4.2	8.3	-	100
Motor Repairs	55.0	-	16.2	17.5	11.4	100.1
Miscellaneous Services	10.9	1.0	18.1	27.5	42.5	100
Public Administration	27.5	19.8	19.4	22.3	10.9	99.9
Totals	23.6	3.5	18.9	25.0	29.0	

Table 5.15: Percentage of school-leavers in each job-type by industry, 1976/7

INDUSTRY	Apprentice	Profess'l	Clerical	Operative	Others	Totals
Agriculture	0.2	-	-	0.6	1.3	0.58
Mining	2.1	0.5	0.3	0.4	0.3	0.74
Food, Drink & Tobacco	0.7	0.5	2.3	2.3	8.0	3.52
Chemical & Allied	0.2	-	0.6	0.1	-	0.16
Metal Manufacture	16.7	6.8	10.9	7.6	2.5	8.84
Engineering	4.4	1.0	3.7	2.5	0.5	2.52
Electrical Goods	0.8	1.0	1.2	1.0	0.9	0.98
Engineers Machine Tools	6.2	-	3.5	4.8	1.5	3.74
Vehicle Manufacture	0.1	0.5	0.4	0.1	0.1	0.15
Hand Tools	3.8	1.0	3.2	5.5	4.4	4.17
Cutlery	1.5	0.5	3.0	7.4	17.5	7.88
Other Metal Goods	10.6	1.0	5.1	7.4	5.7	6.99
Textiles	0.1	0.5	0.1	0.4	0.3	0.25
Clothing	0.2	-	0.5	4.3	2.7	1.98
Bricks, Pottery & Glass	0.3	0.5	1.0	0.3	0.5	0.49
Timber & Furniture	0.7	-	0.5	2.4	1.6	1.32
Paper, Printing & Publishing	0.3	-	1.3	1.2	1.4	1.01
Other Manufacture	0.2	-	0.1	0.6	1.0	0.51
Construction	19.3	5.2	5.3	3.9	3.1	7.60
Gas, Electricity & Water	1.5	0.5	1.8	0.5	-	0.85
Transport & Communications	2.5	0.5	2.7	0.8	0.3	1.42
Distributive	2.3	4.2	14.9	30.9	36.8	21.94
Insurance, Banking & Finance	0.8	9.4	14.7	0.8	0.1	3.52
Professional & Scientific Services	1.5	39.6	11.3	3.0	1.1	4.95
Hairdressing	6.5	-	0.4	0.6	-	1.74
Motor Repairs	9.7	-	3.6	2.9	1.6	4.16
Miscellaneous Services	1.6	1.0	3.4	3.9	5.1	3.52
Public Administration	5.2	25.5	4.6	4.0	1.7	4.48
Totals	100.0	99.7	100.4	100.2	100.0	100.01

Table 5.16: Percentage of leavers per category for each job type

Name of Category	B.S.I.C. Numbers	Number of firms	Size L	M	S
Food, Drink and Tobacco	211 to 240 inc	5	2	2	1
Metal Manufacture	311 to 323 inc	11	4	4	3
Engineering	331 to 370 inc	4	2	1	1
Engineering Small Tools	390	5	2	2	1
Hand Tools	391	5	2	2	1
Cutlery	392	10	4	3	3
Other Metal Manufacture	393 to 399 inc	9	3	3	3
Clothing	441 to 450 inc	3	1	1	1
Timber and Furniture	471 to 479 inc	3	1	1	1
Paper, Printing and Publishing	481 to 489 inc	3	1	1	1
Construction	500	10	4	3	3
Transport and Communication	701 to 709 inc	3	1	1	1
Distribution					
Wholesale Food and Drink	810	2)	1	1	0
Other Wholesale	(811) (812)	2)	1	1	0
Retail Food and Drink	820	7) 28	3	2	2
Other Retail	821	14)	5	5	4
Dealing in Supplies	(831) (832)	3)	1	1	1
Insurance, Banking and Finance	860 to 866 inc	5	2	1	1
Hotel, Restaurants and Catering	884,885,888 only	5	2	2	1
Hairdressing	889	3	1	1	1
Motor Repairs	894	5	2	2	1
Sheffield Metropolitan District Council	872 & 906, 3 only	1			
		118			

Table 5.17: Final Composition of the Sample of Firms

Total of firms sampled:		117
less:	wholly-owned subsidiary	4
	bankruptcy	2
	change of business	2
		109

Total of firms possible:	109
No. contacted	109
Appointments made	77
Appointments conducted	77
% responding:	70.6%

Non-responders:	
No school-leavers	13
Refusal	12
Appointment not arranged	7
	32

Table 6.1 : Analysis of Interview Response Rates for firms

Industry	Small		Medium (& Medium/Large)		Large	
	Inter-viewed	Total possible	Inter-viewed	Total possible	Inter-viewed	Total possible
Food, Drink and Tobacco	0	1	1	2	2	2
Metal Manufacture	1	3	3	4	3	4
Engineering	0	1	0	1	2	2
Engineering Small Tools	1	1	2	2	2	2
Hand Tools	1	1	2	2	2	2
Cutlery	1	4	4	5	2	2
Other Metal Goods	2	3	1	3	3	3
Clothing	1	1	1	1	-	-
Timber and Furniture	0	1	1	2	-	-
Paper, Printing and Publishing	0	1	1	1	1	1
Construction	1	3	2	3	3	4
Transport and Communications	0	1	0	1	1	1
Distribution	7	11	6	11	4	6
Insurance, Banking and Finance	1	2	2	2	1	1
Hotels, Restaurants & Caterers	1	2	3	3	-	-
Hairdressers	1	2	-	-	-	-
Motor Repairs	1	1	3	3	1	1
Total	19	39	32	46	27	31
% of Total	48.7%		69.5%		87%	

Table 6.2 Analysis of Response to Interview by category and size

Total Employees	Appren- tices	Profess- ional	Clerical	Operatives/ others	Total leavers
53,729	327	12	215	470	1147

(N = 77 firms : Individual job categories are low approximations.as firms were often unable to specify individual jobs)

Table 6.3 Number of Employees and School-leavers in Sample Firms

Job-Type	Industry	Number of Subjects
Apprentice	Motor Vehicle	10
Apprentice	Metal Manufacture	10
Apprentice	Engineering	10
Clerical	Engineering	4
Clerical	Distribution	10
Distributive Operative	Other Retail	4
Operative/ others	Metal Manufacture	19
Operative/ others	Food, Drink, Tobacco	8

Table 6.4 Analysis of School-leaver interviews

Clerical Courses

Basic Clerical Course	(Block Release)
" " "	(Day Release)
B E C General Course	

Apprentice Courses:

Basic Hairdressing
Meat Trades Certificate
Food Service Certificate
Basic Cookery
Basic Bakery and Flour Confectionery

Basic Engineering Craft Studies Pt 1	(Electrical Bias)
" " " "	(Fabrication & Welding Bias)
" " " "	(Mechanical Bias)

T E C in Fabrication and Welding Studies
Vehicle Trades Apprentice Stage V
Electrical Installation Pt 1
T E C in Mechanical & Production Engineering

Distribution Operative Courses

National Distribution Certificate

Operative/Others Courses

Silversmithing & Allied Crafts
Industrial Operations, Hollow-ware
Printing
Industrial Operatives Certificate (Cutlery & Engineering)
Mechanical Trades Principles

Table 6.5: Courses investigated at Colleges of Further Education

Department	Number of school-leavers 73	74	75	76	77
Housing (staff)	-	-	30	30	35
(works)	-	-	-	38	45
Cleansing	1	3	3	2	5
Treasury	18	30	13	13	27
Family & Community Services	-	2	4	3	0
Works	-	109	113	54	114
Recreation	16	16	16	14	19
Libraries	23	36	19	21	11
Estates	2	6	8	2	8
Environmental Health	1	6	5	2	4
Admin. & Legal (est)	4	3	3	4	4
Education	-	16	11	8	10
Totals	65	227	225	191	282

(Planning and Design not replying)

Table 6.6 : Employment of school-leavers by City of Sheffield Metropolitan District Council 1973-1977

Test	Coefficient	N Items	Mean (Test)	S.Dev.	Mean (Retest)	S.Dev.
Whole Test	0.86	31	21.85	5.67	23.13	5.34
CONT - 2	0.53	5	3.32	1.27	3.60	1.23
CONT - 3	0.75	10	6.94	1.90	7.43	1.90
CONT - 4	0.67	6	4.26	1.50	4.45	1.32
CONT - 5	0.63	4	2.11	1.29	2.30	1.08
CONT - 6	0.80	6	5.23	1.25	5.36	1.13
JOB - 1	0.80	14	10.06	2.98	10.47	2.95
JOB - 3	0.62	5	2.91	1.19	3.26	1.17
JOB - 4	0.63	4	2.68	1.02	2.85	1.04
JOB - 5	0.59	2	1.43	0.68	1.47	0.69
JOB - 6	0.58	6	4.77	1.25	5.09	0.93
LING - 1	0.85	17	12.91	3.11	13.49	3.01
LING - 2	0.72	13	8.47	2.89	9.19	2.69
LING - 4	0.36	1	0.47	0.50	0.45	0.50

(N = 47)

Table 12.1: Test-Retest analysis for FRT A: Product moment correlation

	Content Subtests					Job Subtests					Linguistic Task		
	2	3	4	5	6	1	3	4	5	6	1	2	4
Test	0.77	0.81	0.74	0.75	0.85	0.95	0.61	0.67	0.66	0.77	0.92	0.90	0.41
Retest	0.74	0.85	0.80	0.74	0.85	0.94	0.74	0.62	0.62	0.66	0.93	0.90	0.25

Table 12.2 Correlations of total test score with total subtest scores, FRT A

(N=47)

Cutting Score	κ	Standard Error	φ
1	-	-	-
2	-	-	-
3	0.0	0.99	-
4	0.0	0.99	-
5	0.0	0.99	-
6	0.0	0.99	-
7	0.0	0.69	-
8	0.66	0.34	0.70
9	0.66	0.34	0.70
10	1.00	0.00	1.00
11	1.00	0.00	1.00
12	0.64	0.25	0.64
13	0.64	0.25	0.64
14	0.64	0.25	0.64
15	0.63	0.21	0.64
16	0.81	0.13	0.81
17	0.81	0.13	0.81
18	0.76	0.13	0.76
19	0.85	0.10	0.86
20	0.63	0.14	0.68
21	0.65	0.13	0.70
22	0.56	0.13	0.57
23	0.51	0.13	0.52
24	0.66	0.11	0.66
25	0.61	0.12	0.61
26	0.47	0.15	0.48
27	0.46	0.17	0.51
28	0.61	0.16	0.66
29	0.41	0.25	0.50
30	-0.02	0.71	-0.02
31	0.0	0.99	-

Table 12.3: Agreement coefficients: FRT A (N=47)

Test	Retest	Number	Percentage
Incorrect	Correct	161	11.05
Correct	Correct	926	63.55
Incorrect	Incorrect	269	18.46
Correct	Incorrect	101	6.93
		1457	99.99

Table 12.4 : Analysis of Response Consistency at Test and Retest (N=47), FRT A

Score	Number	Percentage
Increase	27	57.4
Same	7	14.9
Decrease	13	27.7
	47	99.9

Table 12.5 Analysis of Score Consistency at Test and Retest, FRT A

Test	Correlation Coefficient	N Items	Mean (Test)	St.Dev.	Mean (Retest)	St.Dev.
Whole Test	0.56	30	24.73	2.22	25.98	2.22
CONT-2	0.39	12	10.41	1.00	11.11	0.89
CONT-3	0.48	14	11.23	1.63	11.57	1.53
CONT-4	-0.05	3	2.89	0.39	2.98	0.15
CONT-6	0.14	1	0.20	0.41	0.32	0.47
JOB-1	0.41	3	1.86	0.82	1.93	0.73
JOB-2	0.55	4	2.50	0.73	2.80	0.90
JOB-4	0.44	10	9.07	0.87	9.32	0.67
JOB-5	0.44	9	7.80	1.13	8.30	1.00
JOB-6	0.33	4	3.50	0.66	3.64	0.53
LING-1	0.51	14	11.27	1.35	11.73	1.47
LING-2	0.33	9	8.11	0.89	8.64	0.53
LING-4	0.38	7	5.34	1.12	5.61	0.97

Table 12.6 Pearson product-moment correlations for Test and Retest, FRT B (N = 44)

	Content Subtests					Job Subtests				Linguistic Tasks		
	2	3	4	6	1	3	4	5	6	1	2	4
Test	0.47	0.88	0.48	0.34	0.64	0.39	0.39	0.53	0.71	0.77	0.42	0.71
Retest	0.63	0.87	0.14	0.63	0.62	0.60	0.47	0.71	0.37	0.88	0.45	0.70

Table 12.7 Correlations of total test score with total subtest scores, FRT B

(N = 44)

Cutting Score	κ	S.Error.	ϕ
Scores 1 to 19 - no scores in this range	-	-	-
20	-	-	-
21	0.0	0.99	-
22	0.38	0.35	0.48
23	0.28	0.25	0.29
24	0.18	0.19	0.22
25	0.21	0.15	0.24
26	0.38	0.14	0.41
27	0.24	0.16	0.25
28	0.51	0.19	0.58
29	0.19	0.28	0.32
30	0.0	0.69	-

Table 12.8 Agreement coefficients, FRT B, whole test
(N = 44)

Test	Retest	Number	Percentage
Incorrect	Correct	112	8.48
Correct	Correct	1031	78.11
Incorrect	Incorrect	120	9.09
Correct	Incorrect	57	4.32
		1320	100.00

Table 12.9 Analysis of Response Consistency at test and retest, FRT B (N = 44)

Score	Number	Percentage
Increase	32	72.72
Same	8	18.18
Decrease	4	9.09
	44	99.99

Table 12.10 Analysis of Score consistency at test and retest, FRT B

	Test		Retest 1		Retest 2
Sum of scores	652		682		710
Difference		30	58	28	

Table 12.11 : Sum of scores for each testing occasion and differences between the occasions

(N = 31)

	Test & Retest 1		Test & Retest 2		Retest 1 & Retest 2	
	Number	Percentage	Number	Percentage	Number	Percentage
Incorrect to Correct	120	12.49	131	13.63	91	9.47
Correct each test	562	58.48	578	60.25	619	64.41
Incorrect each test	189	19.67	178	18.52	188	19.56
Correct to Incorrect	90	9.37	73	7.60	63	6.56
	961	100.0	961	100.0	961	100.0

Table 12.12: Response changes over three testing occasions (N= 31)

	School 1	School 2	School 3	School 4	School 5(L)	School 5(U)	School 6	School 7(L)	School 7(U)	Total
Male	21	26	11	20	8	10	37	57	62	252
Female	15	31	7	32	8	14	33	30	49	219
Total	36	57	18	52	16	24	70	87	111	471

(L = lower ability)
(U = upper ability)

Table 14.1 : Sex of testees by school group

	SCHOOL								
	I	II	III	IV	V(L)	V(U)	VI	VII(L)	VII(U)
High Ability				33					103
Medium Ability	36	57				24	62	74	
Low Ability			17	20	16				
Totals	36	57	17	53	16	24	62	74	103

Table 14.2 Ability ranges of testees by school group

	No.of items	Mean Score	Standard Deviation	Standard Error of Mean	Skew	Kurtosis	Minimum	Maximum	Ra
Whole Test	31	21.07	4.79	0.23	-0.58	-0.24	7.0	30.0	2
JOB-1	14	9.30	2.78	0.13	-0.54	-0.51	1.0	14.0	1
JOB-3	5	3.01	1.15	0.06	-0.17	-0.43	0.0	5.0	
JOB-4	4	2.66	0.88	0.04	-0.25	-0.23	0.0	4.0	
JOB-5	2	1.35	0.72	0.04	-0.64	-0.86	0.0	2.0	
JOB-6	6	4.76	0.93	0.05	-1.09	+1.81	1.0	6.0	
CONT-2	5	3.17	1.12	0.05	-0.50	-0.06	0.0	5.0	
CONT-3	10	6.91	1.82	0.09	-0.46	-0.13	1.0	10.0	
CONT-4	6	4.0	1.37	0.07	-0.40	-0.64	1.0	6.0	
CONT-5	4	1.73	1.19	0.06	+0.1	-0.88	0.0	4.0	
CONT-6	6	5.27	1.03	0.05	-1.53	+2.03	1.0	6.0	
LING-1	17	12.8	2.69	0.13	-0.82	+0.49	2.0	17.0	1
LING-2	13	7.94	2.53	0.12	-0.42	-0.42	0.0	13.0	1
LING-4	1	0.31	0.46	0.02	+0.81	-1.35	0.0	1.0	1

Table 14.3 : Summary statistics for FRT A (N=428)

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE (PCT)
7	1	0.2	0.2
8	2	0.5	0.7
9	5	1.2	1.9
10	4	0.9	2.8
11	4	0.9	3.7
12	9	2.1	5.8
13	8	1.9	7.7
14	15	3.5	11.2
15	17	4.0	15.2
16	13	3.0	18.2
17	18	4.2	22.4
18	23	5.4	27.8
19	24	5.6	33.4
20	25	5.8	39.3
21	36	8.4	47.7
22	36	8.4	56.1
23	38	8.9	65.0
24	37	8.6	73.6
25	26	6.1	79.7
26	39	9.1	88.8
27	26	6.1	94.9
28	13	3.0	97.9
29	5	1.2	99.1
30	4	0.9	100.0
TOTAL	428	100.0	

MEAN	21.070	STD ERR	0.231	MEDIAN	21.778
MODE	26.000	STD DEV	4.789	VARIANCE	22.936
KURTOSIS	-0.239	SKEWNESS	-0.575	RANGE	23.000
MINIMUM	7.000	MAXIMUM	30.000		

Table 14.4 : Frequency Distribution of Scores: FRT A

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
1	1	0.2	0.2
3	10	2.3	2.6
4	18	4.2	6.8
5	28	6.5	13.3
6	22	5.1	18.5
7	28	6.5	25.0
8	39	9.1	34.1
9	47	11.0	45.1
10	69	16.1	61.2
11	62	14.5	75.7
12	58	13.6	89.3
13	36	8.4	97.7
14	10	2.3	100.0
TOTAL	428	100.0	

MEAN	9.304	STD ERR	0.134	MEDIAN	9.804
MODE	10.000	STD DEV	2.779	VARIANCE	7.720
KURTOSIS	-0.507	SKEWNESS	-0.538	RANGE	13.000
MINIMUM	1.000	MAXIMUM	14.000		

Table 14.5: Frequency Distribution of Scores: subtest JOB-1

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
0	6	1.4	1.4
1	34	7.9	9.3
2	100	23.4	32.7
3	139	32.5	65.2
4	107	25.0	90.2
5	42	9.8	100.0
TOTAL	428	100.0	

MEAN	3.012	STD ERR	0.056	MEDIAN	3.032
MODE	3.000	STD DEV	1.150	VARIANCE	1.323
KURTOSIS	-0.425	SKEWNESS	-0.171	RANGE	5.000
MINIMUM	0.0	MAXIMUM	5.000		

Table 14.6 : Frequency Distribution of Scores: subtest JOB-3

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
0	4	0.9	0.9
1	31	7.2	8.2
2	147	34.3	42.5
3	172	40.2	82.7
4	74	17.3	100.0
TOTAL	428	100.0	

MEAN	2.657	STD ERR	0.043	MEDIAN	2.686
MODE	3.000	STD DEV	0.880	VARIANCE	0.774
KURTOSIS	-0.233	SKEWNESS	-0.246	RANGE	4.000
MINIMUM	0.0	MAXIMUM	4.000		

Table 14.7: Frequency Distribution of Scores: subtest JOB-4

SCORES	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
0	63	14.7	14.7
1	152	35.5	50.2
2	213	49.8	100.0
TOTAL	428	100.0	

MEAN	1.350	STD ERR	0.035	MEDIAN	1.493
MODE	2.000	STD DEV	0.723	VARIANCE	0.523
KURTOSIS	-0.855	SKEWNESS	-0.642	RANGE	2.000
MINIMUM	0.0	MAXIMUM	2.000		

Table 14.8 : Frequency Distribution of Scores: subtest JOB-5

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
1	2	0.5	0.5
2	12	2.8	3.3
3	22	5.1	8.4
4	91	21.3	29.7
5	227	53.0	82.7
6	74	17.3	100.0
TOTAL	428	100.0	

MEAN	4.755	STD ERR	0.045	MEDIAN	4.883
MODE	5.000	STD DEV	0.927	VARIANCE	0.860
KURTOSIS	1.812	SKEWNESS	-1.087	RANGE	5.000
MINIMUM	1.000	MAXIMUM	6.000		

Table 14.9 : Frequency Distribution of Scores: subtest JOB-6

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
0	6	1.4	1.4
1	32	7.5	8.9
2	65	15.2	24.1
3	146	34.1	58.2
4	139	32.5	90.7
5	40	9.3	100.0
TOTAL	428	100.0	

MEAN	3.168	STD ERR	0.054	MEDIAN	3.260
MODE	3.000	STD DEV	1.118	VARIANCE	1.250
KURTOSIS	-0.059	SKEWNESS	-0.497	RANGE	5.000
MINIMUM	0.0	MAXIMUM	5.000		

Table 14.10 : Frequency Distribution of Scores: subtest CONT-2

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
1	1	0.2	0.2
2	5	1.2	1.4
3	14	3.3	4.7
4	24	5.6	10.3
5	46	10.7	21.0
6	75	17.5	38.6
7	84	19.6	58.2
8	96	22.4	80.6
9	56	13.1	93.7
10	27	6.3	100.0
TOTAL	428	100.0	

MEAN	6.914	STD ERR	0.088	MEDIAN	7.083
MODE	8.000	STD DEV	1.823	VARIANCE	3.325
KURTOSIS	-0.129	SKEWNESS	-0.460	RANGE	9.000
MINIMUM	1.000	MAXIMUM	10.000		

Table 14.11: Frequency Distribution of Scores: subtest CONT-3

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
1	21	4.9	4.9
2	45	10.5	15.4
3	84	19.6	35.0
4	97	22.7	57.7
5	126	29.4	87.1
6	55	12.9	100.0
TOTAL	428	100.0	

MEAN	3.998	STD ERR	0.066	MEDIAN	4.160
MODE	5.000	STD DEV	1.368	VARIANCE	1.871
KURTOSIS	-0.642	SKEWNESS	-0.404	RANGE	5.000
MINIMUM	1.000	MAXIMUM	6.000		

Table 14.12 : Frequency Distribution of Scores: subtest CONT-4

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
0	81	18.9	18.9
1	100	23.4	42.3
2	130	30.4	72.7
3	87	20.3	93.0
4	30	7.0	100.0
TOTAL	428	100.0	

MEAN	1.731	STD ERR	0.057	MEDIAN	1.754
MODE	2.000	STD DEV	1.185	VARIANCE	1.405
KURTOSIS	-0.884	SKEWNESS	0.100	RANGE	4.000
MINIMUM	0.0	MAXIMUM	4.000		

Table 14.13 : Frequency Distribution of Scores: subtest CONT-5

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
1	2	0.5	0.5
2	9	2.1	2.6
3	20	4.7	7.2
4	52	12.1	19.4
5	103	24.1	43.5
6	242	56.5	100.0
TOTAL	428	100.0	

MEAN	5.269	STD ERR	0.050	MEDIAN	5.616
MODE	6.000	STD DEV	1.034	VARIANCE	1.068
KURTOSIS	2.028	SKEWNESS	-1.528	RANGE	5.000
MINIMUM	1.000	MAXIMUM	6.000		

Table 14.14: Frequency Distribution of Scores: subtest CONT-6

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
2	1	0.2	0.2
4	1	0.2	0.5
5	2	0.5	0.9
6	6	1.4	2.3
7	11	2.6	4.9
8	14	3.3	8.2
9	18	4.2	12.4
10	26	6.1	18.5
11	43	10.0	28.5
12	38	8.9	37.4
13	70	16.4	53.7
14	72	16.8	70.6
15	66	15.4	86.0
16	42	9.8	95.8
17	18	4.2	100.0
TOTAL	428	100.0	

MEAN	12.799	STD ERR	0.130	MEDIAN	13.271
MODE	14.000	STD DEV	2.689	VARIANCE	7.229
KURTOSIS	0.488	SKEWNESS	-0.818	RANGE	15.000
MINIMUM	2.000	MAXIMUM	17.000		

Table 14.15 : Frequency Distribution of Scores: subtest LING-1

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
0	1	0.2	0.2
1	1	0.2	0.5
2	8	1.9	2.3
3	12	2.8	5.1
4	24	5.6	10.7
5	32	7.5	18.2
6	41	9.6	27.8
7	54	12.6	40.4
8	61	14.3	54.7
9	62	14.5	69.2
10	58	13.6	82.7
11	51	11.9	94.6
12	21	4.9	99.5
13	2	0.5	100.0
TOTAL	428	100.0	

MEAN	7.939	STD ERR	0.122	MEDIAN	8.172
MOSE	9.000	STD DEV	2.528	VARIANCE	6.390
KURTOSIS	-0.415	SKEWNESS	-0.418	RANGE	13.000
MINIMUM	0.0	MAXIMUM	13.000		

Table 14.16 : Frequency Distribution of Scores: subtest LING-2

SCORE	ABSOLUTE FREQUENCY	FREQUENCY (PCT)	CUMULATIVE FREQUENCY (PCT)
0	294	68.7	68.7
1	134	31.3	100.0
TOTAL	428	100.0	

MEAN	0.313	STD ERR	0.022	MEDIAN	0.228
MODE	0.0	STD DEV	0.464	VARIANCE	0.216
KURTOSIS	-1.352	SKEWNESS	0.809	RANGE	1.000
MINIMUM	0.0	MAXIMUM	1.000		

Table 14.17 : Frequency Distribution of Scores: subtest LING-4

	TOTAL	JOB-1	JOB-3	JOB-4	JOB-5	JOB-6	CONT-2	CONT-3	CONT-4	CONT-5	CONT-6	LING-1	LING-2	LING-
TOTAL	1.00													
JOB-1	0.92	1.00												
JOB-3	0.64	0.44	1.00											
JOB-4	0.60	0.47	0.25	1.00										
JOB-5	0.51	0.36	0.26	0.21	1.00									
JOB-6	0.64	0.48	0.28	0.30	0.26	1.00								
CONT-2	0.68	0.56	0.30	0.28	0.51	0.75	1.00							
CONT-3	0.82	0.68	0.86	0.36	0.49	0.42	0.45	1.00						
CONT-4	0.72	0.73	0.30	0.70	0.25	0.32	0.31	0.42	1.00					
CONT-5	0.70	0.73	0.34	0.48	0.24	0.37	0.39	0.45	0.45	1.00				
CONT-6	0.66	0.63	0.31	0.37	0.33	0.52	0.38	0.46	0.40	0.28	1.00			
LING-1	0.89	0.76	0.74	0.41	0.60	0.53	0.55	0.92	0.48	0.45	0.74	1.00		
LING-2	0.88	0.86	0.39	0.67	0.32	0.59	0.65	0.54	0.82	0.75	0.45	0.60	1.00	
LING-4	0.37	0.41	0.17	0.20	0.06*	0.23	0.17	0.23	0.22	0.61	0.13**	0.21	0.30	1.00

(N = 428; $p < 0.001$ in all cases except those marked: * : not significant
** : $p < 0.004$

Table 14.18: Pearson product-moment correlation coefficients, total score and subtests, FRT A

	No. of items	Mean	Standard Deviation	Standard Error of Mean	Skew	Kurtosis	Minimum	Maximum	R
Raw Score	155	109.8	25.96	1.27	-0.82	+0.13	23	152	
Quotient	-	100.79	12.45	0.61	+0.02	+0.06	70	130	
Skimming	30	25.32	4.96	0.24	-1.51	+1.74	7	30	
Vocabulary	35	26.19	8.03	0.39	-0.90	-0.10	2	35	
Reading for Facts	30	22.43	4.83	0.24	-0.94	0.59	5	30	
Points of View	35	20.0	7.8	0.38	-0.33	-0.66	0	35	
Comprehension	25	15.8	4.46	0.22	-0.51	-0.14	0	25	

Table 14.19: Summary Statistics for ERT 4 (N = 418)

	ERT 4 Raw Score	ERT 4 Quotient
FRT A Total Score	0.79	0.74

($N = 375$, $p < 0.001$)

Table 14.20: Pearson correlation coefficients of total score measures on the two tests

	Raw Score	Quotient	Skimming	Vocabulary	Reading for Facts	Points of View	Comprehensio
Quotient	0.96	1.00					
Skimming	0.79	0.73	1.00				
Vocabulary	0.91	0.86	0.64	1.00			
Reading for Facts	0.87	0.84	0.69	0.75	1.00		
Points of View	0.88	0.87	0.56	0.75	0.68	1.00	
Comprehension	0.84	0.82	0.60	0.73	0.70	0.69	1 .00

($N = 412$, $p < 0.001$)

Table 14.21: Intercorrelation matrix of ERT 4 total score measures and subtests

	Functional Reading Test Content Subtests				
	CONT-2	CONT-3	CONT-4	CONT-5	CONT-6
ERT 4 Comprehension subtest	0.50	0.48	0.56	0.44	0.47

(N = 374, $p < 0.001$ in all cases)

<u>Table 14.22: Pearson correlation coefficients for FRT A content subtests and ERT 4 Comprehension subtest</u>

	ERT 4 subtests:	
	Reading for Facts	Skimming
FRT A : LING-1 (Referential Linguistic Task)	0.67	0.63

($N = 375$, $p < 0.001$)

Table 14.23: Pearson correlation coefficients for subtests requiring location of information, FRT A and ERT 4

Group	Mean	Standard Deviation	N
Entire population tested	21.07	4.79	428
Female	21.71	4.26	200
Male	20.43	5.01	228
School 1	20.44	3.92	34
Female	21.67	3.90	15
Male	19.47	3.75	19
School 2	19.19	6.22	57
Female	21.32	5.29	31
Male	16.65	6.39	26
School 3	20.67	3.85	18
Female	18.86	4.78	7
Male	21.82	2.79	11
School 4	23.74	4.06	39
Female	24.00	3.06	22
Male	23.41	5.16	17
School 5 (L)	14.69	3.09	16
Female	15.13	3.40	8
Male	14.25	2.91	8
School 5 (U)	23.04	3.30	24
Female	21.86	3.35	14
Male	24.70	2.54	10
School 6	19.70	4.79	63
Female	20.81	4.66	31
Male	18.63	4.74	32
School 7 (L)	20.41	4.19	73
Female	21.07	4.45	28
Male	20.00	4.02	45
School 7 (U)	23.19	3.55	104
Female	23.87	2.83	44
Male	22.68	3.95	60

Table 14.24: Mean quotients for groups shown: FRT A

Group	Mean	Standard Deviation	N
Entire population tested	100.79	12.45	418
Female	101.55	11.63	197
Male	100.70	12.91	221
School 1	95.0	8.74	24
Female	97.22	9.72	9
Male	93.67	8.15	15
School 2	93.72	13.81	57
Female	98.19	12.37	31
Male	88.38	13.76	26
School 3	91.0	8.57	14
Female	85.67	8.78	6
Male	95.00	6.21	8
School 4	101.0	11.33	48
Female	101.03	10.06	30
Male	100.94	13.50	18
School 5 (L)	86.87	9.45	15
Female	85.0	7.65	8
Male	89.0	11.40	7
School 5 (U)	100.36	6.11	22
Female	98.42	6.05	12
Male	102.7	5.58	10
School 6	99.42	10.91	59
Female	98.97	8.83	29
Male	99.87	12.74	30
School 7 (L)	98.53	9.40	75
Female	96.42	8.65	26
Male	99.65	9.67	49
School 7 (U)	111.73	9.27	104
Female	112.41	9.23	46
Male	111.19	9.34	58

Table14.25: Mean quotients of groups shown: ERT 4

	ERT 4 quotients			
	70 - 84	85 - 100	101 - 115	116+
Lowest FRT A	7	10	12	21
Highest FRT A	18	28	30	29
Mean FRT A	12.3	19.6	23.5	25.2

Table 14.26: FRT A Score ranges by normal standard deviation groups

ERT 4 QUOTIENTS

	73 -	74 - 86	87 - 100	101 - 114	115 - 121	129 +
Lowest FRT A	8	7	10	12	21	22
Highest FRT A	12	21	28	30	29	29
Mean FRT A	9.6	13.2	19.6	23.4	25.2	25.7
N	10	35	136	146	33	13

Table 14.27: FRT A score ranges by actual standard deviation groups

Test	Correlation	Significance
FRT A Total	0.09	0.04
JOB -1	0.10	0.02
JOB -3	0.05	0.19
JOB -4	0.09	0.05
JOB -5	0.02	0.38
JOB -6	0.01	0.41
CONT -2	0.03	0.31
CONT -3	0.04	0.22
CONT -4	0.11	0.02
CONT -5	0.13	0.01
CONT -6	0.04	0.20
LING -1	0.05	0.17
LING -2	0.11	0.01
LING -4	0.08	0.05

Table 14.28: Correlation of Age at test with FRT A Scores

(N = 374)

Entire population tested	16 : 0	4	417
Female	16 : 01	4	197
Male	15 : 11	4	221
School 1	16 : 1	3	24
Female	16 : 2	4	9
Male	16 : 0	3	15
School 2	16 : 1	3	57
Female	16 : 1	3	31
Male	16 : 1	3	26
School 3	16 : 2	4	14
Female	16 : 4	4	6
Male	16 : 1	3	8
School 4	16 : 3	4	48
Female	16 : 4	4	30
Male	16 : 2	3	18
School 5 (L)	15 : 9	4	14
Female	15 : 9	5	8
Male	15 : 9	5	6
School 5 (U)	15 : 10	3	22
Female	15 : 10	3	12
Male	15 : 10	3	10
School 6	15 : 10	3	59
Female	15 : 10	3	29
Male	15 : 10	3	30
School 7 (L)	15 : 11	4	75
Female	15 : 11	3	26
Male	15 : 11	4	49
School 7 (U)	16 : 0	4	104
Female	16 : 0	4	46
Male	15 : 11	4	58

Table 14.29: Mean ages (rounded to nearest month) of groups shown

Test	Significance	Direction
Total Score	0.003	Females > Males
JOB-1	0.01	Females > Males
JOB-3	N.S.	-
JOB-4	N.S.	-
JOB-5	N.S.	-
JOB-6	0.012	Females > Males
CONT-2	0.002	Females > Males
CONT-3	0.01	Females > Males
CONT-4	N.S.	-
CONT-5	0.02	Females > Males
CONT-6	N.S.	-
LING-1	0.015	Females > Males
LING-2	0.002	Females > Males
LING-4	N.S.	

(N = 428; 2-tailed probability) N.S.= non-significant probability

Table 14.30: Mann-Whitney U-tests for sex differences FRT A Totals

Test	Significance	Direction
Raw Score	N.S.	-
Quotient	N.S.	-
Skimming	0.0001	Females > Males
Vocabulary	N.S.	-
Reading for Facts	N.S.	-
Points of View	N.S.	-
Comprehension	N.S.	-

(N = 417; 2-tailed probability) N.S.= non-significant probability

Table14.31: Mann-Whitney U-tests for sex differences ERT 4 Quotients

	Males		Females		Total	
Destination	Number	%	Number	%	Number	%
Employment	106	42.1	67	30.7	173	36.8
Unemployment	40	15.9	35	16.1	75	16.0
Return to school	58	23.0	55	25.2	113	24.0
Further Education	10	4.0	31	14.2	41	8.7
Untraced	38	15.1	30	13.8	68	14.5
Total	252	100.1	218	100.0	470	100.0

Table 15.1: Destination of School-leavers by sex
(N=470)

	Males		Females		Total	
Job - type	Number	%	Number	%	Number	%
Apprentice	46	45.1	5	8.3	51	31.5
Clerical	4	3.9	34	56.7	38	23.4
Distribution Op/Other	14	13.7	10	16.7	24	14.8
Op/Others	38	37.3	11	18.3	49	30.2
Totals	102	100.0	60	100.0	162	100.0

Table 15.2: Occupation of Employed School-leavers by sex

(N = 162)

	Number	%
Number of school-leavers obtaining employment:	173	36.8 [a]
Number taking up work within follow-up period:	101	58.4 [b]
Number for whom employer was contracted:	101	100.0 [c]
Number of completed forms:	35	35.0 [d]
Number of usable forms *:	28	80.0 [e]

a: percentage of sample

b: percentage of those obtaining employment

c: percentage of those taking up work

d: percentage of those contracted

e: percentage of those completed

*: 1 form returned without name; 1 form returned completed for wrong person;
5 forms only partially completed.

Table 15.3: Analysis of follow-up response rates

	Reading performance	Non-reading performance	Oral fluency
Reading performance	1.0		
Non-reading performance	0.85	1.0	
Oral fluency	0.68	0.54	1.0

(N = 28. : $p < 0.001$ in all cases)

Table 15.4: Nonparametric correlations of performance totals

	FRT A	ERT 4	Induction subtest	Age
Performance total	-0.01	0.01	0.08	-0.28
Reading scales total	0.05	0.08	0.18	-0.22
Non-reading scales total	0.23	0.12	-0.002	-0.18
Oral fluency	-0.08	0.07	0.12	-0.25

(N = 28 : no values significant)

Table 15.5: Nonparametric correlations of performance with test scores and age at test (All job types)

	FRT A	ERT 4	Apprentice subtest	Induction	Age
Performance total	-0.5 (3)	-0.2 (5)	-0.5(3)	0.5 (3)	-0.4 (5)
Reading scales total	-0.5 (3)	-0.2 (5)	-0.5 (3)	0.5 (3)	-0.4 (5)
Non-reading scales total	-0.5 (3)	-0.5 (5)	-0.5 (3)	0.5 (3)	-0.3 (5)
Oral fluency	-0.87(3)	0.35(5)	-0.87	-0.87(3)	-0.35(5)

(No values significant)
(Numbers in parentheses = number of cases)

Table 15.6: Nonparametric correlations of performance with test scores and age at test for Apprentice job-type

	FRT A	ERT 4	Clerical subtest	Induction	Age
Performance total	0.34	0.31	-0.08	0.41	0.36
Reading scales total	0.33	0.28	-0.11	0.38	0.43
Non-reading scales total	0.51	0.35	-0.11	0.06	0.44
Oral fluency	-0.15	0.19	0.08	0.54	0.09

(N = 9)

(No values significant)

Table 15.7: Nonparametric correlations of performance with test scores and age at test for Clerical job-type

	FRT A	ERT 4	Distribution subtest	Induction	Age
Performance total	-0.94**(6)	-0.76*(6)	-0.42(6)	-0.69(6)	-0.81*(6)
Reading scales total	-0.94**(6)	-0.76*(6)	-0.42(6)	-0.69(6)	-0.81*(6)
Non-reading scales total	-0.24 (7)	-0.51 (7)	0.08 (7)	-0.78*(7)	-0.47(7)
Oral fluency	-0.39 (7)	-0.43 (7)	0.03 (7)	-0.36(7)	-0.52(7)

* : $p < 0.05$

** : $p < 0.01$

(N = number in parentheses)

Table 15.8: Nonparametric correlations of performance with test scores and age at test for Distribution job-type

	FRT A	ERT 4	Op/Others	Induction	Age
Performance total	0.44 (7)	0.47 (7)	-0.18(7)	0.14(7)	-0.83*(7)
Reading scales total	0.91**(7)	0.90**(7)	0.24(7)	0.7*(7)	-0.41 (7)
Non-reading scales total	0.49 (9)	0.50 (9)	0.07(9)	0.27(9)	-0.24 (9)
Oral fluency	0.41 (9)	0.41 (9)	0.37(9)	0.24(9)	-0.55 (9)

* : p 0.05

**: p 0.01

N = number in parentheses

Table 15.9: Nonparametric correlations of performance with test scores and age at test for Op/Others job-type

	Employment Status				
	Employed	Unemployed	School	F E	Untraced
Lowest FRT A	7	8	12	10	10
Highest FRT A	30	26	30	30	29
Mean FRT A	20.3	18.3	24.00	21.6	20.8
N	153	70	101	39	64

Table 15.10: Ranges and means of FRT A scores by Employment Status

	Job-type			
	Apprentice	Clerical	Distribution	Operative/Other
Lowest FRT A	12	15	11	7
Highest FRT A	30	29	28	26
Mean FRT A	21.2	23.0	20.3	17.4
N	46	32	23	45

Table 15.11: Ranges and means of FRT A scores by Job-type

Cutting Score	λ	u	χ^2	Significance
8	0	0	1.7	NS
9	0	0	6.3	NS
10	0.01	0.01	21.9	0.0002
11	0.01	0.01	12.4	0.02
12	0.01	0.01	14.01	0.007
13	0	0.01	10.87	0.03
14	0	0.01	14.78	0.005
15	0	0.03	27.3	0.0002
16	0	0.02	26.6	0.00001
17	0	0.03	32.1	0.00001
18	0	0.04	32.3	0.00001
19	0	0.03	35.1	0.00001
20	0	0.04	36.6	0.00001
21	0	0.03	42.8	0.00001
22	0.03	0.04	41.3	0.00001
23	0.04	0.03	46.2	0.00001
24	0.06	0.04	54.4	0.00001
25	0.04	0.03	45.8	0.00001
26	0.04	0.03	41.0	0.00001
27	0.03	0.02	27.1	0.00001
28	0	0.01	7.5	NS
29	0	0	4.9	NS
30	0	0	4.1	NS

Table 15.12: Uncertainty coefficients for each cutting score: Prediction of Employment Status

(NS = not significant; N = 470)

Cutting Score	λ	u	χ^2	Significance
8	0.009	0.005	2.3	NS
9	0.009	0.005	2.3	NS
10	0.018	0.011	4.7	NS
11	0.20	0.16	7.05	NS
12	0.03	0.02	5.26	NS
13	0.05	0.02	8.54	NS
14	0.05	0.03	10.7	0.03
15	0.07	0.04	15.5	0.004
16	0.11	0.04	18.9	0.0008
17	0.11	0.05	18.8	0.0009
18	0.12	0.05	20.11	0.0005
19	0.14	0.06	24.7	0.0001
20	0.14	0.05	22.0	0.0002
21	0.14	0.06	24.5	0.0001
22	0.14	0.05	22.2	0.0002
23	0.12	0.05	21.5	0.0002
24	0.01	0.04	15.8	0.003
25	0.07	0.03	13.3	0.01
26	0.06	0.03	12.6	0.02
27	0.02	0.02	8.6	NS
28	0.02	0.02	9.2	NS
29	0.01	0.01	4.0	NS
30	0	0.01	2.8	NS

Table 15.13: Uncertainty coefficients for each cutting score: prediction of Job-type

(NS = not significant; N = 173)

Section 1

				Answer Box										
Question Number	Key	Item No	Omit	1	2	3	4	5	6	7	8	9	Number Correct	Number taking Item
1	2-3-5-8	72	0	27	479	491	12	467	31	9	457	11	346	529
2	2-3-5-8	165	3	23	437	475	15	448	16	3	467	39	357	509
3	1-3	141	46	311	75	243	-	-	-	-	-	-	134	509
4	1-4-6-8	74	3	483	14	42	477	57	415	9	437	-	312	508
5	4	7	21	67	68	81	298	37	-	-	-	-	272	506
6	2-3-4-8	73	12	172	451	456	431	67	20	10	320	-	233	506
7	Complete	1	31	-	-	-	-	-	-	-	-	-	171	507

Table 19.1: Analysis of Item responses: Part 1: Action-type

Question	Key	Item	Omit	Answer Box 1	2	3	4	5	Correct	Number
8	5	37	6	5	4	21	22	451	451	509
9	1	38	2	483	2	13	6	6	483	509
10	3	39	10	102	60	283	17	43	283	509
11	4	108	5	13	9	13	446	10	446	506
12	5	109	6	55	7	14	8	415	415	506
13	2	110	6	20	382	56	60	26	382	506
14	1	94	1	403	3	0	1	0	402	506
15	5	95	2	7	4	9	6	478	478	506
16	3	96	2	1	38	443	5	18	443	506
17	4	47	2	10	2	1	491	2	491	508
18	3	48	3	22	10	424	22	19	424	508
19	2	49	1	4	478	8	6	10	478	508
20	3	2	5	186	40	212	55	10	213	508
21	2	3	11	25	425	31	14	6	428	508
22	3	4	3	21	15	453	3	19	450	508
23	3	104	6	8	9	472	2	12	472	507
24	4	105	22	28	77	51	125	204	125	507
25	5	106	16	29	50	9	14	492	489	507
26	3	130	24	185	24	314	61	14	311	509
27	4	131	29	63	55	69	264	37	261	509
28	2	132	38	130	284	11	21	127	284	509
29	4	127	41	175	14	9	267	6	266	509
30	2	128	62	19	352	9	24	43	352	509
31	4	129	71	40	26	26	178	167	178	509

Table 19.1(continued): Analysis of Item responses : Part 2: Multiple Choice Item

APPENDIX II

BRITISH STANDARD INDUSTRIAL CLASSIFICATION

MINIMUM LIST HEADINGS

BRITISH STANDARD INDUSTRIAL CLASSIFICATION

Summary of Orders and Minimum List Headings

Minimum List Heading	
	ORDER I - AGRICULTURE, FORESTRY, FISHING
001	Agriculture and horticulture
002	Forestry
003	Fishing
	ORDER II - MINING AND QUARRYING
101	Coal mining
102	Stone and slate quarrying and mining
103	Chalk, clay, sand and gravel extraction
104	Petroleum and natural gas
109	Other mining and quarrying
	ORDER III - FOOD, DRINK AND TOBACCO
211	Grain milling
212	Bread and flour confectionery
213	Biscuits
214	Bacon curing, meat and fish products
215	Milk and milk products
216	Sugar
217	Cocoa, chocolate and sugar confectionery
218	Fruit and vegetable products
219	Animal and poultry foods
221	Vegetable and animal oils and fats
229	Food industries not elsewhere specified
231	Brewing and malting
232	Soft drinks
239	Other drink industries
240	Tobacco

Heading

ORDER IV - COAL AND PETROLEUM PRODUCTS

261	Coke ovens and manufactured fuel
262	Mineral oil refining
263	Lubricating oils and greases

ORDER V - CHEMICALS AND ALLIED INDUSTRIES

271	General chemicals
272	Pharmaceutical chemicals and preparations
273	Toilet preparations
274	Paint
275	Soap and detergents
276	Synthetic resins and plastics materials and synthetic rubber
277	Dyestuffs and pigments
278	Fertilizers
279	Other chemical industries

ORDER VI - METAL MANUFACTURE

311	Iron and steel (general)
312	Steel tubes
313	Iron castings, etc
321	Aluminium and aluminium alloys
322	Copper, brass and other copper alloys
323	Other base metals

ORDER VII - MECHANICAL ENGINEERING

331	Agricultural machinery (except tractors)
332	Metal-working machine tools
333	Pumps, valves and compressors
334	Industrial engines
335	Textile machinery and accessories
336	Construction and earth-moving equipment

List
Heading

ORDER VII - MECHANICAL ENGINEERING continued

337	Mechanical handling equipment
338	Office machinery
339	Other machinery
341	Industrial (including process) plant and steelwork
342	Ordnance and small arms
349	Other mechanical engineering not elsewhere specified

ORDER VIII - INSTRUMENT ENGINEERING

351	Photographic and document copying equipment
352	Watches and clocks
353	Surgical instruments and appliances
354	Scientific and industrial instruments and systems

ORDER IX - ELECTRICAL ENGINEERING

361	Electrical machinery
362	Insulated wires and cables
363	Telegraph and telephone apparatus and equipment
364	Radio and electronic components
365	Broadcast receiving and sound reproducing equipment
366	Electronic computers
367	Radio, radar and electronic capital goods
368	Electric appliances primarily for domestic use
369	Other electrical goods

ORDER X - SHIPBUILDING AND MARINE ENGINEERING

370	Shipbuilding and marine engineering

Minimum
List
Heading

ORDER XI - VEHICLES

380	Wheeled tractor manufacturing
381	Motor vehicle manufacturing
382	Motor cycle, tricycle and pedal cycle manufacturing
383	Aerospace equipment manufacturing and repairing
384	Locomotives and railway track equipment
385	Railway carriages and wagons and trams

ORDER XII - METAL GOODS NOT ELSEWHERE SPECIFIED

390	Engineers' small tools and gauges
391	Hand tools and implements
392	Cutlery, spoons, forks and plated tableware, etc
393	Bolts, nuts, screws, rivets, etc
394	Wire and wire manufactures
395	Cans and metal boxes
396	Jewellery and precious metals
399	Metal industries not elsewhere specified

ORDER XIII - TEXTILES

411	Production of man-made fibres
412	Spinning and doubling on the cotton and flax systems
413	Weaving of cotton, linen and man-made fibres
414	Woollen and worsted
415	Jute
416	Rope, twine and net
417	Hosiery and other knitted goods
418	Lace
419	Carpets
421	Narrow fabrics (not more than 30 cm. wide)
422	Made-up textiles
423	Textile finishing
429	Other textile industries

Minimum List Heading	
	ORDER XIV - LEATHER, LEATHER GOODS AND FUR
431	Leather (tanning and dressing) and fellmongery
432	Leather goods
433	Fur
	ORDER XV - CLOTHING AND FOOTWEAR
441	Weatherproof outerwear
442	Men's and boys' tailored outerwear
443	Women's and girls' tailored outerwear
444	Overalls and men's shirts, underwear, etc
445	Dresses, lingerie, infants' wear, etc
446	Hats, caps and millinery
449	Dress industries not elsewhere specified
450	Footwear
	ORDER XVI - BRICKS, POTTERY, GLASS, CEMENT, ETC
461	Bricks, fireclay and refractory goods
462	Pottery
463	Glass
464	Cement
469	Abrasivesand building materials, etc. not elsewhere specified
	ORDER XVII - TIMBER, FURNITURE, ETC
471	Timber
472	Furniture and upholstery
473	Bedding, etc
474	Shop and office fitting
475	Wooden containers and baskets
479	Miscellaneous wood and cork manufactures

Minimum List Heading	
	ORDER XVIII - PAPER, PRINTING AND PUBLISHING
481	Paper and board
482	Packaging products of paper, board and associated materials
483	Manufactured stationery
484	Manufactures of paper and board not elsewhere specified
485	Printing, publishing of newspapers
486	Printing, publishing of periodicals
489	Other printing, publishing, bookbinding, engraving, etc
	ORDER XIX - OTHER MANUFACTURING INDUSTRIES
491	Rubber
492	Linoleum, plastics floor-covering, leathercloth, etc
493	Brushes and brooms
494	Toys, games, children's carriages, and sports equipment
495	Miscellaneous stationers' goods
496	Plastics products not elsewhere specified
499	Miscellaneous manufacturing industries
	ORDER XX - CONSTRUCTION
500	Construction
	ORDER XXI - GAS, ELECTRICITY AND WATER
601	Gas
602	Electricity
603	Water supply

Minimum List Heading	
	ORDER XXII - TRANSPORT AND COMMUNICATION
701	Railways
702	Road passenger transport
703	Road haulage contracting for general hire or reward
704	Other road haulage
705	Sea transport
706	Port and inland water transport
707	Air transport
708	Postal services and telecommunications
709	Miscellaneous transport services and storage
	ORDER XXIII - DISTRIBUTIVE TRADES
810	Wholesale distribution of food and drink
811	Wholesale distribution of petroleum products
812	Other wholesale distribution
820	Retail distribution of food and drink
821	Other retail distribution
831	Dealing in coal, oil, builders' materials, grain and agricultural supplies
832	Dealing in other industrial materials and machinery
	ORDER XXIV - INSURANCE, BANKING, FINANCE AND BUSINESS SERVICES
860	Insurance
861	Banking and bill discounting
862	Other financial institutions
863	Property owning and managing, etc
864	Advertising and market research
865	Other business services
866	Central offices not allocable elsewhere

Minimum List Heading	
	ORDER XXV - PROFESSIONAL AND SCIENTIFIC SERVICES
871	Accountancy services
872	Educational services
873	Legal services
874	Medical and dental services
875	Religious organisations
876	Research and development services
879	Other professional and scientific services
	ORDER XXVI - MISCELLANEOUS SERVICES
881	Cinemas, theatres, radio, etc
882	Sport and other recreations
883	Betting and gambling
884	Hotels and other residential establishments
885	Restaurants, cafes, snack bars
886	Public houses
887	Clubs
888	Catering contractors
889	Hairdressing and manicure
891	Private domestic service
892	Laundries
893	Dry cleaning, job dyeing, carpet beating, etc
894	Motor repairers, distributors, garages and filling stations
895	Repair of boots and shoes
899	Other services
	ORDER XXVII - PUBLIC ADMINISTRATION AND DEFENCE
901	National government service
906	Local government service

36 Collegiate Crescent
Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742)

Department of Professional Studies
Head of Department G. Q. Craig, MA.

Please reply t(Letter 1
Language Devel(
37, Clarkehouse Road,
Sheffield,
S10 2LD

WL/JHB

Dear,

FUNCTIONAL READING PROJECT

As you will see from the attached document, we are, in collaboration with the Education Department of Sheffield Metropolitan District Council, trying to develop a test which can be used to assess how ready are pupils, leaving at 16+, for the reading requirements of the beginning of their employment.

Our sampling procedure has resulted in several firms from your industry being selected and we would be very grateful for the co-operation of the Association in the project. The nature of such co-operation would depend upon the role the Association plays in relation to firms in the Sheffield area - whether informative, supportive or actually running training schemes itself.

The Research Assistant to the project, Mr Owen Parry, will contact you in the next few days to arrange an appointment if you are willing to help. Please do not hesitate to question him concerning the project, or to contact me direct if you wish.

We hope this project will be one step in decreasing the gap between education and the needs of industry and that you will be able to help us.

Yours sincerely,

W. Latham

W. Latham,
Co-ordinator,
Language Development Centre.

36 Collegiate Crescent
Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742)

Department of Professional Studies
Head of Department G. Q. Craig, MA.

Letter 2

Please reply to:
Language Development Centre,
37, Clarkehouse Road,
Sheffield,
S10 2LD

WL/JHB

Dear

FUNCTIONAL READING PROJECT

As you will see from the attached document, we are, in collaboration with the Education Department of Sheffield Metropolitan District Council, trying to develop a test which can be used to assess how ready are pupils, leaving at 16+, for the reading requirements of the beginning of their employment.

Our sampling procedure has resulted in several firms from your industry being selected and we would be very grateful for the co-operation of the Training Board in the project. The nature of such co-operation would depend upon the role the Training Board plays in relation to firms in the Sheffield area - whether informative, supportive or actually running training schemes itself.

The Research Assistant to the project, Mr Owen Parry, will contact you in the next few days to arrange an appointment if you are willing to help. Please do not hesitate to question him concerning the project, or to contact me direct if you wish.

We hope this project will be one step in decreasing the gap between education and the needs of industry and that you will be able to help us.

Yours sincerely,

W. Latham,
Co-ordinator,
Language Development Centre.

36 Collegiate Crescent
Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742)
ext 201

Department of Professional Studies
Head of Department G. Q. Craig, MA.

Please reply to:
Language Development Centre,
37 Clarkehouse Road,
Sheffield,
S10 2LD

WL/JHB

Dear

FUNCTIONAL READING PROJECT

As you will see from the attached document, we are, in collaboration with the Education Department of Sheffield Metropolitan District Council, trying to develop a test which can be used to assess how ready are pupils, leaving at 16+, for the reading requirements of the beginning of their employment.

As this test is being designed for use in Sheffield schools, it is to be constructed from items taken from reading used by Sheffield school-leavers when they commence employment.

Our sampling procedure has resulted in your firm being selected as one of those which will represent your industry and, thus, we would be very grateful for your co-operation in the project.

Co-operation would involve the following:-

(i) an initial discussion between your representative and the Research Assistant to the project, Mr Owen Parry;

(ii) the inspection of reading materials actually used on the job by school-leavers;

(iii) discussions with some of your recent school-leaver employees about the situations on the job in which reading is required, and

(iv) observation by Mr Parry of some of the leavers at work in situations which require reading.

We are interested in reading materials met by any category of employee who came to you at 16+

Mr Parry will contact you in the next few days to arrange an appointment if you are willing to help. Please do not hesitate to question him concerning the project, or to contact me direct if you wish.

We hope this project will be one step in decreasing the gap between education and the needs of industry and that you will be able to help us.

Yours sincerely,

W. Latham

W. Latham,
Co-ordinator,
Language Development Centre.

Name of firm ______________________________ Schedule 1

Contact ______________________ Telephone ______________

S.I.C. No. ______________________ Size Group ______________

Nature of Industry ______________________________

Total Employees [ ] Total leavers 1977 [ ]

Apprentices [ ] Profs [ ] Clerical [ ] Op T [ ] Others [ ]

Training: Own Scheme [ ] ITB [ ] GTA [ ] Other [ ]

For whom: App [ ] Prof [ ] Cler [ ] Op T [ ] Others [ ]

Personnel Selection: Tests [ ] Interview [ ] Other [ ]

Initial Training Period:

App

Prof

Cler

Op T

Others

Easter Leavers: Total ________

App ______ Prof ______ Cler ______ Op T ______ Others ______

Acad. Qualifs.

App______ Prof ______ Cler ______ Op T ______ Others ______

Reading Problems

Comments

36 Collegiate Crescent
Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742) Ext 201

Please reply to:-
Language Development Centre
37 Clarkehouse Road
Sheffield S10 2LD

Department of Professional Studies
Head of Department G. Q. Craig, MA.

OP/VDA

September 1978

Dear

re: Functional Reading Project

I am writing both to thank you for your help in the Project so far and to ask for your assistance in the next stage of our work.

You may remember in the letter that you originally received two other areas of co-operation were mentioned beyond the discussion with myself. These involved talking with your recent school-leavers about any difficulties that they may have encountered with your reading materials.

As a result of the information received in the initial interviews, we are asking several of the larger firms visited if they would be prepared to help us for a few hours more. I should like to visit you again, this time to spend about ten minutes talking to ten of the school-leavers you have taken on this year as

I shall take the opportunity of telephoning you in a few days time and will be glad to provide more details.

I do hope you will be able to help us for just two hours more.

Yours sincerely,

Owen Parry
Research Assistant.

Schedule II

Schedule 2

No. of test materials:

Encountered materials

Any difficulties: Yes........ No.........

Related task presented problem

Did it need reading help?

Parts were unclear?

still unclear

Whether sought help?

from whom?

often?

Do a lot of reading on job?

or ask?

Any reading problems in general at work?

36 Collegiate Crescent
Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742) ext 201

Department of Professional Studies
Head of Department G. Q. Craig, MA.

Please reply to:
Language Development Centre,
37 Clarkehouse Road,
Sheffield,
S10 2LD

WL/JHB

Dear .

Functional Reading Project

The research project, described in the attached document, is being undertaken at this Centre, in collaboration with Sheffield MDC. We would like to include departments of the Council, taking a minimum number of 16+ year old school-leavers, in our sample of employers, from which we hope to obtain information concerning reading requirements. In order that departments meeting our recruitment figure can be identified, I would be grateful if you would complete and return the attached slip.

Further details concerning the project may be obtained from myself or Mr Owen Parry, Research Assistant employed on the project, at the Centre, or from Mr Brian Wilcox, Senior Advisor (Research and Evaluation) at your Education Department.

I hope you will be able to help us.

Yours sincerely,

W. Latham

W. Latham,
Co-ordinator,
Language Development Centre

36 Collegiate Crescent
Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742)

Department of Professional Studies
Head of Department G. Q. Craig, MA.

Please reply to:
Language Development Centre,
37 Clarkehouse Road,
Sheffield,
S10 2LD

WL/JHB

Dear Sir,

As you will see from the attached document, the Polytechnic is undertaking research in the area of the reading requirements of 16 year old school-leavers when they start work.

We have, so far, investigated the requirements of the jobs or training situations in which these leavers find themselves, and are moving on to look at the related area of Trade Union membership.

I wonder if you would be so kind as to provide us with specimens of material your union gives to new starters at work. We are interested in any and all such material, from forms related to joining the union, material describing the nature and function of the union, etc. and, of course, any recruitment materials you may provide.

We would, of course, seek your express permission, should we desire to use this material for other than purely internal purposes.

With many thanks.

Yours faithfully,

W Latham
Co-ordinator

Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742)

Department of Professional Studies
Head of Department G. Q. Craig, MA.

Please reply to:
Language Development Centre,
37 Clarkehouse Road,
Sheffield,
S10 2LD

OP/JHB

Dear Sirs,

I am writing to ask for your help in a research project being undertaken at this Polytechnic.

We are designing a reading test for use with 16 year old school-leavers, basing our questions on reading materials actually used at work by such young people.

I wonder if you would be so kind as to send me specimens of promotional or informational materials of the sort you would typically provide for retailers of your products, or any specifically produced training materials.

We would, of course, seek your express permission, should we desire to use this material for other than purely internal purposes.

With many thanks.

Yours faithfully,

Owen Parry
Research Assistant

36 Collegiate Crescent
Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742)

Department of Professional Studies
Head of Department G. Q. Craig, MA.

Please reply to:
Language Development Centre,
37 Clarkehouse Road,
Sheffield,
S10 2LD

WL/JHB

19 October 1978

Dear

Functional Reading Project

As you will have seen from our recent letter, the Project is now fairly well advanced. We are moving on from the collection of reading materials used by school-leavers, in which you have already been most helpful, to the construction of reading tests based on these materials.

It is with this in mind that I am writing to you again. We are inviting a number of firms to help in assessing the validity of the test questions we have written.

The help we are asking for is this:

a) the scrutinizing of test questions related to the early stages of the employment of * , (to give your opinion on the representative nature of the material we have used, and the validity of the questions we have asked about it); and

b) attendance on a 'content validation' panel for a morning or an afternoon, one day early in November, at this Centre (lunch would be provided).

We would provide the materials for scrutiny in advance of the panel session, for you to look at over a convenient period of time.

Should you feel able to help us in this way, I would be most grateful if you would return the attached form indicating your availability.

Yours sincerely,

W. Latham

W Latham
Co-ordinator
Language Development Centre

* one of : apprentices, clerical staff, sales assistants or operatives/other workers

I. CONTENT VALIDATION PANEL ATTENDANCE

The panels are drawn from the industries employing the category of employees for that panel. For example, the Apprentices panel is drawn from the engineering, tool-making, metal manufacture, construction, catering, hairdressing and motor vehicle repairing industries.

Mr Latham and Mr Parry will explain, as an introduction, the classification of content that the project is using and, hopefully, this will clarify the panel's activities better than this brief description.

The panels will then be asked to pool their opinions on each test item, going through the test 'passages' and questions in turn.

Please bring all the items with you.

II. PROCEDURE FOR SCRUTINY OF TEST ITEMS

Enclosed are copies of test items, comprising of a 'passage' (either text, form, sign etc.) on the left, and a question or questions on the right. The 'passages' are, in the main, taken from reading materials used by school-leavers in the early stages of their employment.

We have grouped the 'passages' according to five types of job or use:

a) Induction (Terms of employment, handbooks, health and safety rules, etc)

b) Apprentices

c) Clerical juniors

d) Distributive Operatives/others

e) Operatives/others

The materials we have sent you lie in categories (a) and

Procedure for Stage A

We should like you first to go through it, to see whether it is representative of 'passages' used in your industry (or in your firm alone, if you feel unable to go further). The 'passages' do not have to be identical, only similar in style, format and structure. For example, a labelled diagram of a capstan lathe is similar to a labelled diagram of a Volkswagen engine; a job description set out in numbered, detailed stages involves similar style, format and structure whether it be for auto-blade machine operator or a sweet-manufacture operative.

If the 'passages' are irrelevant, obscure or different in style-format and/or structure to 'passages' used in your industry for this category of job, please make a note in Part A of the form attached to the test item.

If, after looking at all the 'passages', you can think of anything we have missed out that is important, please note it on the General Form attached to this document. (If you could bring examples of such missed materials to the panel, this would help us enormously).

Finally, if the 'passage' for any item would never be met by a school-leaver, please make the appropriate note.

Procedure for Stage B

Please go through each test item again, this time looking at the questions associated with each 'passage'.

They fall into three types:

a) oral

b) 'Action'

c) multiple-choice

....

We have attempted to make the task which the question sets as near to the real-life reading task in which the 'passage' is used. For example, filing exercises for office junior work, job instructions for operatives, comprehension of training booklets for apprentices are tasks we have tried to include. Often the material has been less than amenable to these processes and various compromises have had to be made, but in the main, this general rule has been followed.

For each question, will you please judge whether it reflects an important task for the category of job (or induction), which you would normally require of a school-leaver.

If the question is irrelevant, trivial or extremely rare, please make a note on Part B of the form attached to the test item.

Any other comments would also be appreciated.

TEST ITEMS : NOS

A. 'PASSAGES'

1. Is 'passage' similar to type used in your industry?

2. Is it irrelevant
   obscure
   dissimilar to type used in your industry?

3. Never met by school-leaver

B. QUESTIONS

	Question No. (Where applicable?)		
	1	2	3
For each question			
Is it important task?			
irrelevant			
trivial			
extremely rare			

Other comments

MATERIALS NOT REPRESENTED HERE

Please list description of material(s) and type of job in which it/they is/are encountered.

Please reply to:
Language Development Centre
37 Clarkehouse Road
Sheffield S10 2LD

36 Collegiate Crescent
Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742) Ext 201

Department of Professional Studies
Head of Department G Q Craig M A

WL/VDA

Dear Sirs,

Functional Reading Project

As part of this Project, details of which are attached, we have been collecting specimens of reading materials used in the initial period of employment by 16 year old school-leavers. Various firms were kind enough to participate in this work and provided us with specimens of reading materials issued by your organisation.

Our purpose has been the development of reading tests related to the materials met in the early stages of employment by 16+ year old school-leavers, the questions for which are based on actual materials we have collected. Some of the questions are based on materials collected from your firm, and I am writing to request your permission to reproduce these materials. The exact form in which they would be used is shown on the specimen sheets attached.

We have removed all identifying characteristics from the material thus preserving complete confidentiality.

If you feel that you can help us in this way, I would be grateful if you would complete and return the attached form in the envelope provided.

Yours sincerely,

W. Latham

W Latham
Co-ordinator
Language Development Centre

effield City Polytechnic

emorandum to

m Owen Parry, Research Assistant,
Language Development Centre

Date 19 November 1979

re: Functional Reading Project

I am writing to ask you for your assistance with the above Project and a few minutes of your time.

The Project is concerned with the identification and classification of job-related reading tasks encountered by school-leavers starting work or training at 16+, and the production of related reading tests. As part of the Project, a computerized question-bank has been developed, containing large amounts of data for every test question we have written. To complement this bank, it is proposed to develop a second bank, that of technical vocabulary occurring in the test passages. It is with this second bank that you are able to assist us.

In order to pick out the technical vocabulary, we are asking members of the English and Communication Studies Departments to go through about five of our test passages, and underline, or otherwise indicate, any word or phrase they consider to be either a technical term or an everyday term used in a technical sense.

It would be of very great help to us if you could undertake this brief task for us, with the attached test passages and return them in the envelope provided. If you feel unable to help, I would be grateful for their return anyway.

If you are able to help, we would be very glad to receive the materials by the end of this term.

Many thanks,

Owen Parry

Owen Parry,
Department of Professional Studies.

Metropolitan District

Heads of:

Education Department
Chief Education Officer
G M A Harrison, MA
PO Box No. 67
Leopold Street
Sheffield
S1 1RJ

Telephone 0742-26341

If telephoning, please ask for extension and then Mr
Your reference
Our reference
Date 9th April 1979

Dear Colleague,

Functional Reading Project: Sheffield City Polytechnic

As you will know, the Authority is supporting a research project concerned with assessing the reading skills demanded of the school leaver, particularly in the world of work. The project is based at the Language Development Centre of the Sheffield City Polytechnic and is being conducted by Mr. W. Latham, Coordinator of the Centre, and Mr. O. Parry, research assistant.

Two schools have taken part earlier in the year in piloting the assessment materials which have been produced. The completed materials now require validation on a larger scale. It is hoped to involve a further eight schools which have been randomly selected in assisting with this next stage of development.

The name of your school has been identified for inclusion in a possible sample. I would be grateful if you would consider taking part in this next phase of the project. The researchers hope to try their materials out on a sample, possibly up to 50, of fifth form pupils during the coming summer term.

The materials are based on authentic examples of reading matter actually encountered by school leavers in their initial period of employment or training. They therefore represent a unique addition to the range of assessment instruments currently available in this country. I hope, if you are willing to take part, you will find the experience of seeing these materials and the results which are produced of interest and value.

Mr. Latham will be in touch with you early next term with a view to discussing the possibility of your involvement and to give further details of what the cooperation would involve. I hope you will be able to meet him.

Yours sincerely,

G M A. Harrison

Chief Education Officer

cc Mr. W. Latham, Language Development Centre
Mr. O. Parry, Language Development Centre

JW

36 Collegiate Crescent
Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742) (Ext 201)

Department of Professional Studies
Head of Department G Q Craig M A

Language Development Centre
37 Clarkehouse Road
Sheffield S10 2LD

OP/VDA

19 October 1979

Dear

re FUNCTIONAL READING PROJECT

I hope you will remember the visit of my research assistant, Mr Parry, and myself to discuss the above project with you, in the Summer term this year.

As you know, in collaboration with Sheffield LEA, we have been developing reading tests based on job-related reading tasks, for use with 16+ school-leavers. We have discussed with you the possibility of administering our test, and a published test for comparison, to some of your pupils. Whilst you were unable to accommodate us last term, you were kind enough to indicate your willingness to participate in the project at a future date.

We are proposing to undertake a second round of testing in the coming Spring Term and would welcome an occasion to discuss the matter with you again. I shall take the opportunity to telephone you in the next few days to see if we may meet before the end of this term.

Yours sincerely,

W Latham
Co-ordinator
Language Development Centre

Schedule 3

Sheffield City Polytechnic Functional Reading Project W

When this young person enters employment or the Youth Opportunities Scheme or has a definite intention to start a full-time FE course in September, please fill in the relevant details below and return this card to [redacted].

Thank you.

Name of person: ______________________________

Name of employer: ______________________________

Name of contact in firm: ____________________

Date of starting ____________ Type of work ____________

If applicable, please tick the appropriate box(es) below:

Work Experience ☐ Full-time college ☐ Other training ☐

Temporary Job ☐ Job not in Sheffield City area ☐

36 Collegiate Crescent
Sheffield S10 2BP
Telephone Sheffield 665274 (STD Code 0742) Ext 201

Department of Professional Studies
Head of Department G Q Craig M A

Please reply to:
Language Development Centre
37 Clarkehouse Road
Sheffield S10 2LD

WL/VDA

Dear

FUNCTIONAL READING PROJECT

As you will see from the attached document, we are, in collaboration with the Education Department of Sheffield Metropolitan District Council, developing a test which can be used to assess how ready are pupils, leaving at 16+, for the reading requirements of the first six weeks of their employment.

Over the past eighteen months, a large number of firms in Sheffield, Industry Training Boards and Colleges of Further Education have assisted us in our work. A large collection of reading materials used by school-leavers has been amassed and used to construct the test described above.

A number of recent school-leavers were given the test before leaving school, and we are now following them up as they get jobs, in order that we can compare their test results with their actual performance at work. This will enable us to calculate the ability of the reading test to predict adequate performance on job-related reading tasks.

One of the school-leavers has recently entered your employment and I am writing to ask for your co-operation in this stage of the project.

Co-operation would involve the young person's immediate supervisor completing a short assessment form, a copy of which is attached, and then returning it to us in the envelope provided.

The young person was informed that a follow-up of some testees would take place, but not specifically in his or her case. Should you wish to consult the young person, this is perfectly acceptable and we would, of course, respect any objections.

The information you provide will be treated as strictly confidential and no mention will be made of your company or individuals in any report we may write.

Please do not hesitate to contact me, or the Research Assistant to the Project, Mr Owen Parry, should you wish to discuss this matter further. We hope this project will be one step in decreasing the gap between education and the needs of industry and commerce and that you will be able to help us.

Yours sincerely,

W. Latham

W Latham
Co-ordinator
Language Development Centre

Schedule 4

Name of subject

Date of rating

On the list attached, there are nine questions and beneath each one are five words or phrases.

Please read each question and then rate the young person, named above, by putting a circle around the word or phrase which, in your opinion, most nearly describes his or her performance in relation to the question.

Here are two examples of completed questions:

3. How often does the young person have to ask for guidance in a task needing reading?

Very often	Quite often	Sometimes	Not very often	Very seldom

7. In a task not involving reading, how satisfied are you with the young person's speed of performance?

Very dissatisfied	Dissatisfied	Fairly Dissatisfied	Satisfied	Very Satisfied

Please read each question carefully.

How well does the young person persevere with tasks not involving reading?

Very well	Quite well	Fairly well	Not very well	Not at all well

---

In a task needing reading, how satisfied are you with the young person's speed of performance?

Very satisfied	Satisfied	Fairly satisfied	Dissatisfied	Very dissatisfied

---

How often does the young person have to ask for guidance in a task needing reading?

Very often	Quite often	Sometimes	Not very often	Very seldom

In a task not involving reading, how satisfied are you with the young person's overall performance?

Very dissatisfied	Dissatisfied	Fairly satisfied	Satisfied	Very satisfied

---

How able is the young person in expressing him or herself, in general?

Very fluent	Quite fluent	Able	Not very fluent	Unable

How well does the young person persevere with tasks involving reading?

Very well	Quite well	Fairly well	Not very well	Not at all well

---

How often does the young person have to ask for guidance in a task not needing reading?

Very seldom	Not very often	Sometimes	Quite often	Very often

In a task not involving reading, how satisfied are you with the young person's speed of performance?

Very dissatisfied	Dissatisfied	Fairly satisfied	Satisfied	Very satisfied

---

In a task involving reading, how satisfied are you with the young person's overall performance?

Very satisfied	Satisfied	Fairly satisfied	Dissatisfied	Very dissatisfied

---

SHEFFIELD OCCUPATIONAL FUNCTIONAL READING PROJECT

IBS

Manual of the Item Banking System

Owen Parry

1980

A number of diagrams and figures included in this work are reproductions and reductions from computer output. The quality of such figures is therefore not always of the standard one would wish and apologies are made to the reader.

CONTENTS

Chapter 1

The Sheffield Occupational Functional Reading Project

Introduction

In the Green Paper, 'Education in Schools - A Consultative Document' (Cmnd 6689, 1977) it was argued that 'schools must prepare their pupils for the transition to adult life and work'. If such preparation is accepted as one of the duties of the school, it seems reasonable to expect secondary schools to prepare their pupils, as far as may be possible, for the functional reading (ie. reading associated with a role or task) which will be required by them when they leave school.

Before schools can prepare their pupils, however, it is necessary for the relevant reading tasks to be identified and means devised for ascertaining the degree to which pupils can cope with the tasks so identified. To identify these tasks and provide means of assessing pupils' attainments in relation to them was the purpose of a research project undertaken by Sheffield City Polytechnic.

Project objectives

In September 1977,Sheffield City Polytechnic, in association with Sheffield Metropolitan District Council, initiated the Sheffield Occupational Functional Reading Project (SOFRP), with the following objectives:

i) The identification and classification of reading tasks, associated with job seeking and employment which a 16 year old could face immediately he/she left school.

ii) The construction and validation of a criterion referenced group test, based on the reading tasks identified in achieving objective (i) above.

iii) The construction of diagnostic tests to be used with individuals who fail to reach the criterion scores on the group test.

(A fourth objective, subject to discussion with Sheffield Schools, was the production of materials to be used in teaching related to occupational functional reading).

The Progress of the Project

In order to identify the reading tasks which might be faced by 16 year old school leavers a large number of firms in Sheffield were visited. The firms were chosen at random from groups of firms representing the main employment areas entered by 16+ school leavers. Each firm gave details of the types of jobs into which leavers were recruited, details of training (where applicable) and induction procedures. Job related reading materials were inspected and collected where possible. Relevant job-seeking materials were obtained from the Careers Service and, where their courses made up the substantial reading requirement of a type of job, from the Further Education colleges..

The reading materials obtained were used to construct test items (an item consisting of a passage to be read and one question related to it). A passage might be purely prose plus a diagram or illustration, a form to be filled in, etc.

Each item was scrutinized by a panel of employers's respresentatives, to ensure that it was representative of job related reading tasks encountered by 16+ school leavers, and also by linguists to ensure that there were no undue complexities in the questions. Items approved by employers' representatives and linguists were then used to construct a possible test. In order to see whether pupils would find any items ambiguous and to try out a proposed pattern of test administration, the test was given to pupils in two schools in Sheffield. Following this, certain items were changed or replaced and, where necessary, re-piloted.

The data associated with each item (eg. the industry and job for which it is relevant, the number of pupils who have taken it, etc) was banked in computer storage. It is the use of that computer data bank that forms the content of this Manual.

A criterion-referenced test of sufficient quality has been developed as a result of the Project. The role of the computer in test development was so intricate and invaluable that it was decided early in the work to develop the data bank for more than just the single test to which the work of the Project was devoted. Hence, the facilities and systems described in this Manual are for the development of occupational functional reading tests of general (such as the one developed) or specific (eg. a test for engineering apprentices) uses. It is assumed that in using this Manual, the test developer is aware of the SOFRP, the meanings of the various data types (see Chapter 3, Table 1),the statistical concepts involved, and the nature of criterion-referenced measurement.

CHAPTER 2

Computerised Item Banking Systems - An overview

An 'item' is a test question. Its format may be of various types: eg. a prose passage and a question; a diagram and a question; or merely a question itself, where the knowledge is assumed, rather than presented in a text. In item-banking systems, this definition becomes rather less clear-cut, as the term 'item' often refers to a test question plus all relevant information pertaining to it. Further, smaller systems - or smaller installations containing the system - frequently refer to an 'item' whilst, in fact, only meaning the data about each question.

The growth of item-banks (in all three senses, above) has been stimulated by the allied growth in assessment in education and other areas. The ability to produce tests of known characteristics in a relatively short period is a valuable asset and, despite initial development costs of the bank, quite likely to be more efficient and cost effective. C J Byrne, in his review of question-banking systems, (Byrne, 1975) indicates the existence of over one hundred different systems. The majority of these are located in the United States and there appears to be a high level of sophistication involved in many of them, with numerous different facilities available. Descriptions of many banks are available in a symposium published in the Education Technology magazine (May, 1973)

Byrne (1975) points out that computerized banks can be classified into three categories: "item-bankers' banks", "test-developers' banks", and "bank-users' banks". The first are more concerned with the computing aspects of the bank and the provision of a wide range of facilities and options; the second strive to produce tests of high technical quality; the third are banks designed for non-experts to devise their tests with the minimum of effort. Sheffield City Polytechnic installed an item-bank for questions in the physical sciences and home economics, (1976), but found that the IBM FIBEL system they were using was, in fact, far too large for easy user access and only a small part was ever used. FIBEL is clearly a bank of the "item-banker" type, even though that part actually used, the SHEAF system, was user-oriented. The Schools and Local Authorities Item Banks, devised by the NFER, are banks of the second, "test-constructors" type, the aim being to produce tailor-made tests of high technical quality - but with limited access by the test user to the banks. A characteristic of both these examples is that the questions themselves are stored in the computer, an aspect with its own pro's and con's, but certainly a factor in bank usage. The Test Development and Research Unit's item-banking system at Cambridge (Massey & Newbould, 1976) is an example of a rather more user-oriented system, and here only the item-related data is stored, with the physical items elsewhere.

The Sheffield Occupational Functional Reading Project (SOFRP) Item-Banking System is designed to be user-oriented. It has the advantage of being a bank of limited capacity and specific usage. The degree of sophistication required is therefore fixed and the weight of effort in development has been to cover as many contingencies as possible whilst maintaining simplicity of use.

The facilities available are discussed below, as are the operating procedures. To simplify the User's tasks, specific files have been established to submit and execute work, and an interactive program written to design the data request statements automatically.

This manual discusses each aspect of the work quite fully, possibly in greater detail than necessary. The careful reader, however, should be able to establish for himself the underlying moves for simplicity of use.

References

Byrne C J (1975) — Computerized Question Banking Systems. Institute of Educational Technology, Open University. January 1975

Massey A J, & Newbould C A (1976) — Computerized Item-Banking System (CIBS) Handbook. Occasional Publication No.3, Test Development & Research Unit, Cambridge

Hammond L, Dean C & Morgan G (1976) — Final Report on the SHEAF Evaluation Project. Sheffield City Polytechnic

CHAPTER 3

Using the SOFRP Item-Banking System

1. Introduction

The SOFRP Item-Banking System (IBS) is a large computer program linked to storage devices, which enables an individual (the 'User') to perform certain operations in relation to a set of data held about test items. Twenty-nine data are held about each item in store, these items comprising the 'Item Bank' itself. Each datum is referred to as a 'variable' or 'Bank variable', so each item has twenty-nine variables. Further, each datum has a label (or keyword) of up to four characters (eg. TEXT, BOXØ, LING) and a variable number (1 to 29) by which it may be considered. Table 1 lists these numbers and labels, with descriptions and the maxima and minima, for each variable.

Various facilities are available within the IBS and these are considered in subsequent chapters. The workings of the IBS are not discussed below; rather, the procedures for operating the system are described.

The remainder of this Chapter deals with certain basic information for using the IBS within the Polytechnic computing facilities.

2. MUSIC

The McGill University System for Interactive Computing (MUSIC) is the operating system of the computers used at Sheffield City Polytechnic. MUSIC operates in the same general fashion as most operating systems, and is fairly flexible and comprehensive for the ordinary User.

2.1 Terminals

A computer terminal is a keyboard attached by a telecommunications line to the computer. Work may be processed interactively from a terminal: ie. the User and computer interact via the keyboard. The User will need to familiarise himself with the operation of a terminal.

2.2 Batch work

For longer pieces of work it is more efficient (in processing terms) to allow the work to be done in 'batch'. Here, the User provides a set of specific commands and data (a 'job') for the computer to process. These are fed into a 'stream' of successive jobs to be processed

Variable No.	Description of datum	Keyword	Minimum value	Maximum value
1.	UNIQUE, IDENTIFYING NUMBER	ID	1	250
2.	INDUSTRIAL CATEGORY (BSIC)	IND	1	38
3.	CONTENT TYPE (STICHT,1974, CLASSIFICATION)	CONT	1	6
4.	JOB TYPE (CAREERS SERVICE CLASSIFICATION)	JOB	1	8
5.	PHYSICAL FORMAT OF MATERIAL (PROSE, DIAGRAMS, ETC)	MAT	1	7
6.	FIRM, ITB, COLLEGE ORIGINALLY SUPPLYING ITEM	FIRM	1	118
7.	SIZE OF FIRM	SIZE	1	6
8.	ITEM QUESTION TYPE (ACTION, MULT. CHOICE, ETC)	IQT	1	3
9.	ANSWERING KEY	KEY	0000	9000
10.	ORDER OF ITEM ON THE PAGE	TEXT	1	4
11.	CODE OF LAST TEST IN WHICH ITEM WAS USED	LAST	1	99
12.	NUMBER OF THE ITEM IN LAST TEST	NOQ	1	99
13.	NUMBER OF TESTEES WHO HAVE TAKEN THE ITEM	NUM	0	9999
14.	NUMBER OF ANSWER BOXES(MAX.=10)	NALT	1	10
15.	NUMBER OF TESTEES OMITTING ITEM	OMIT	0	999
16.	NUMBER MARKING FIRST BOX	BOX1	0	999
17.	NUMBER MARKING SECOND BOX	BOX2	0	999
18.	NUMBER MARKING THIRD BOX	BOX3	0	999
19.	NUMBER MARKING FOURTH BOX	BOX4	0	999
20.	NUMBER MARKING FIFTH BOX	BOX5	0	999
21.	NUMBER MARKING SIXTH BOX	BOX6	0	999
22.	NUMBER MARKING SEVENTH BOX	BOX7	0	999
23.	NUMBER MARKING EIGHTH BOX	BOX8	0	999
24.	NUMBER MARKING NINTH BOX	BOX9	0	999
25.	NUMBER MARKING TENTH BOX	BOXO	0	999
26.	NUMBER CORRECTLY RESPONDING	CORE	0	999
27.	LINGUISTIC TASK TYPE	LING	1	4
28.	ITEM NUMBER OF ITEM REPLACING THIS ONE	REPL	0	250
29.	NUMBER OF DISCREPANCIES IN TEST-RETEST	DISC	0	999

Table 1 Data stored per item

sequentially (in fact, short or small output jobs tend to be processed first) whilst the User goes elsewhere about his business.

Batch jobs have the advantage of output on wide paper (132 print positions across the page compared to 80 positions at the terminal) and that the User does not have to stay near the terminal. The 'turnround' time (from submission to collections) varies from 4/5 hours to overnight, depending on the volume of work to be processed: the longer the queue the longer the wait.

All major SOFRP Item-Bank jobs are processed as batch jobs although they are set up at a terminal (see Chapter 5).

2.3 MUSIC documents

A set of explanatory documents are published by the Computer Services Department and are referred to extensively in this Manual. The User is recommended to familiarise himself with several of the basic ones before using the System. 'MUSIC documents' is abbreviated to 'Mdoc' throughout this Manual.

3. Programming languages

The Main system is written in FORTRAN IV and uses the FORTG1 processor via the Link Editor and overlay system. It is stored as a load module.

All other support and processing programs are also in FORTRAN IV

It is not necessary for the User to have any knowledge of the language to use the SOFRP Item-Banking System.

4. User code and Password

Certain identification and accounting data are required for each job submitted, at terminal or in batch.

4.1 The User Code

There is only one code available to perform operations on the entire system. All storage devices are linked to this code and, as no job can do without those devices, it must be used.

4.2 Password

For work at a terminal, a password is also necessary, to avoid misuse of a User code by unscrupulous persons who might have observed it on the print-out.

4.3 The User code and Password are confidential and may only be obtained on personal application to Mr W Latham, Language Development Centre, Sheffield 665274 ext 201.

5. Punched card jobs

A number of IBS jobs need to be submitted as decks of punched cards. These may be prepared by the User himself, or forms may be coded and given to the Computer Services Department, who will then prepare the decks.

Special forms, specimens of which are given in Appendix III, have been prepared to minimise difficulty. The User enters his data or commands into the appropriate spaces and hands - or send via internal mail - the forms to Job Reception, Heriot House. The cards will then be submitted for processing automatically. The Use should allow a reasonable amount of time for punching and processing (3 days to a week, or longer if work is heavy at the Computing Unit) before telephoning (Pond Street ext 482) and either asking for the materials to be returned or checking if they are ready and collecting them personally.

6. Other aspects

The following aspects are relevant:

ASPECT	REFERENCE
General terminal use	Mdoc 2.2/9
Batch commands	Mdoc 2.2/3
Sample terminal sessions	Mdoc 2.2/12 p.6 ff
Context Editor	Mdoc 2.2/13

Users are advised, at all times, to seek help with the system from Computer Services Department before trying to undertake work.

CHAPTER 4

Control of the System

The Item Banking system comprises a series of subroutines controlled by a central program. One or more subroutines may be used in the course of executing a particular requirement (or 'option'). That requirement is provided by the User in the form of certain specific, sequenced commands. 'Sequenced' is emphasised here, as the system is programmed for a set order of commands, rather than having a command recognition facility.

The operation of the Main Control System is shown diagrammatically in Figure 1.

The information specifying the nature of the Bank and the requirements made upon it are given in the two 'READ' boxes. Control information is provided as 'data' for the system, and may be categorised as:

1. Run Parameters: these specify the number of items in the Bank and the amount of associated data, the initial value for certain options and the number and names of options to be invoked. Some of these are obtained internally by the system itself.

2. Option Parameters: these provide the information required for the operation of the subroutine(s) involved in the execution of an option.

Data types

Data may be of two types: numerical or alphamerical, depending on the information required by the control system at that time.

Numerical data are comprised of integer numbers, of up to 5 digits. Examples are 3Ø, 27, 56103, etc. There are no data which require numbers involving decimal parts or other ways of expressing large numbers.

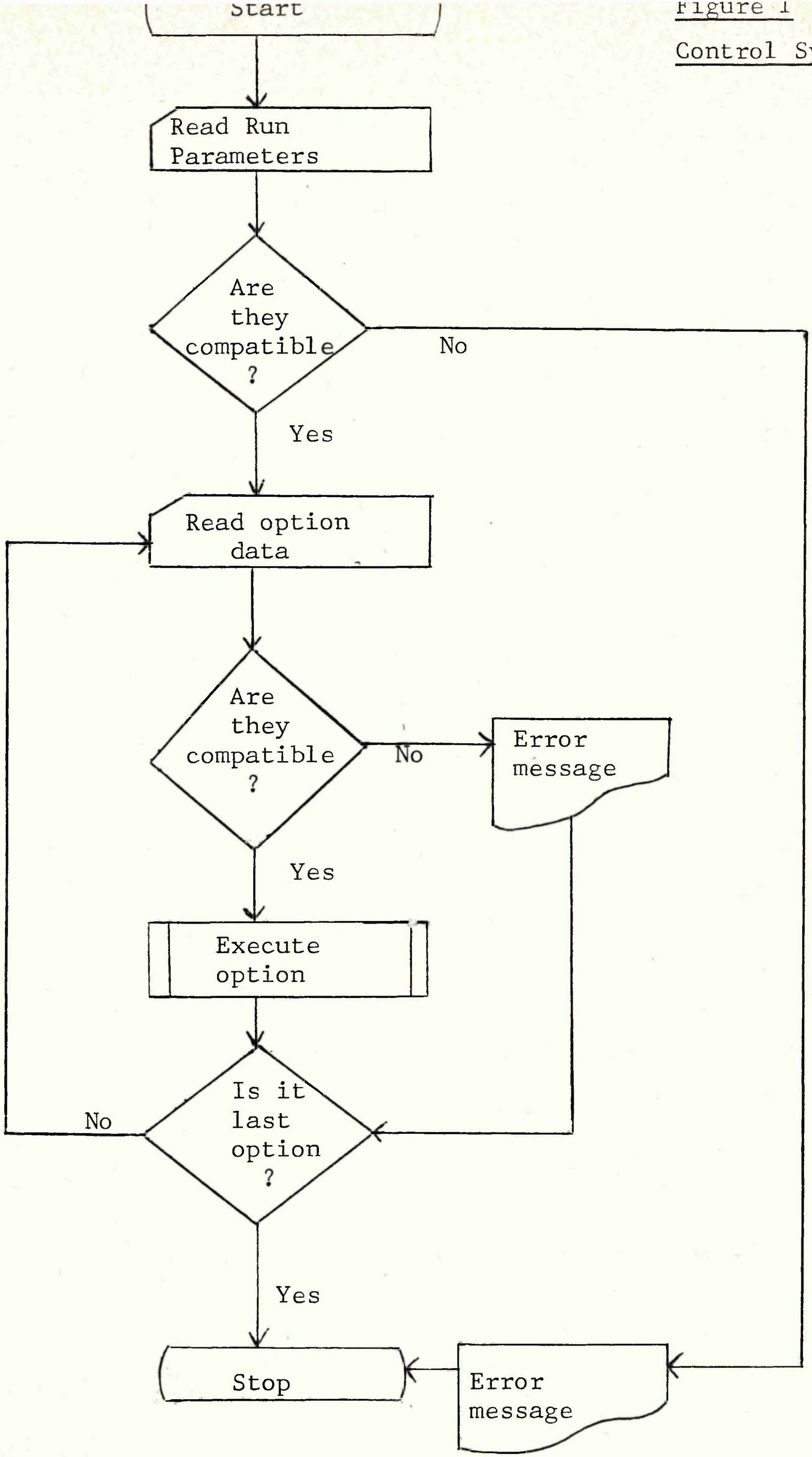
Alphameric data are comprised of certain combinations of letters and numbers which represent certain 'keywords' or identifiers. Examples are BOXØ, LING, UPDATE, etc. They take no numerical value in any computation, but act as signallers for the use of options or procedures or variables.

Data Input

For any run or job involving the system, data must be provided in sequenced order. Such data might be numerical or alphameric, depending on the User's choice of work. See Chapter 5.

Figure 1

Control System



In summary, then, control of the operation of the system is by User-provided data statements. These can be numerical or alphamerical, depending upon the datum required by the input sequence. Numerical values are integer numbers, received in col.21 onwards. Alphamerical data are of character-format and are received in col. 1 onwards. A Comment, when required, takes up the full 80 columns.

An interactive program controls the sequence of command statements by asking for values or alphameric data at the appropriate time. (See Chapter 5)

Error control

It is always possible, bearing in mind the fallibility of human beings, that the User may provide values or keywords that are incompatible with the System's requirements. Values may be mutually inconsistent, exceed maxima, give a null result, etc.

For this reason, all values are checked, insofar as possible, for errors. Messages are output and options or jobs cancelled, depending upon the severity of the errors. It is hoped that this will prevent the production of undetected unreliable results. The principle of'garbage-in, garbage-out' still obtains, however, and the checking routines cannot cope with, for instance, misuse or editing of the User Data Sets (see below) to provide inconsistent results.

CHAPTER 5

Interactive Data Statement Design Program

1. Introduction

The work of the Item Bank is controlled by a series of ordered commands, called a data statement (DS). The design of the DS for a job using the Item Bank is handled interactively (ie. the User sits at a terminal and provides information in response to questions from the computer).

2. Use of the program

At a terminal, the DS program is invoked by the command: /EXEC DATAST. The program then proceeds to write up a series of questions requiring single answers. All numerical values required are integers (i.e. whole numbers with no decimal part) and no decimal point or fractional part should be provided.

The following types of data may be required:-

i) numerical values (the vast majority)
ii) an option name (see the relevant chapters, 6 to 11)
iii) a title or comment (up to 80 characters, which will be used as a title for the relevant option)
iv) a Bank variable name (see Table 1)

The meaning of each data request is explained in the relevant chapters for each option, and for the control system (Chapters 4 and 6 to 11)

3. Functioning of the Program

The program creates a DS and writes it onto a UDS (see Chapter 13) storage file. This file is then used, subsequently, by the main Item Bank system. The DS is in the correct format for the main system (usually an identifying label in cols. 1 to 20 and numerical values in cols. 21 onwards).

4. Changing values

The User may wish to change, delete or add values to the DS. At the end of the DS creation, the program issues a list of the DS and asks if changes are required. If so, the program itself ends, but an editing program (EDITDA) is immediately invoked, using the MUSIC Context Editor (Mdoc 2.2/13).

In response to the output 'EDIT' at the terminal, the DS values may be changed.

It is to be noted that no lines may be added to the file, as this would extend it beyond its end-of-file marker. Instead, the REPLACE facility should be used, rather than the INSERT command.

5. Job submission

Job submission is contingent upon User requirements. The submission will only take place on the User's sayso at the terminal.

If the User does wish for submission, he has the choice of keeping his DS after the job, or not.

Job submission is automatic.

The message:

JOB NUMBER: XXXX

where XXXX is the "remote job number" is output and this number should be used for queries or when collecting work. If uncollected, work will be sent automatically via the internal mail.

Existing DS will be deleted on Wednesday afternoons as part of the usual systems purge.

A User keeping a DS after submission may re-submit it via the SUBMIT system (Mdoc 2.2/7) using the file name 'BRUN5'. If no longer wishing to use it, he should use the command (at a terminal) /EXEC DELDAT

which deletes the DS.

Information about submitted jobs may be obtained by the commands

/EXEC MYRJE or

/QUERY

6. Provision of Control Data

The Interactive Program writes up a series of questions at the User's terminal indicating the data required. The first set of questions concern the overall control of the job: the initial positioning of the file markers and the selection of how many and which options are to be used. These are the Run Parameters.

Figure 2 below gives a sample of this section of the Interactive Program.

```
          SHEFFIELD OCCUPATIONAL FUNCTIONAL READING PROJECT
              ITEM BANKING INTERACTIVE PROGRAM


ENTER THE INITIAL RECORD NUMBER FOR UDS2
?
1

HOW MANY OPTIONS DO YOU WISH TO USE ?
?
6

DO YOU WISH FOR A LIST OF OPTION NAMES ?
?
YES
   SELECT
   STATIST
   SELSTAT
   BANKLIST
   UPDATE
   CROZTABS


PLEASE ENTER AN OPTION NAME
?
UPDATE

PLEASE ENTER AN OPTION NAME
?
STATIST

PLEASE ENTER AN OPTION NAME
?
BANKLIST

PLEASE ENTER AN OPTION NAME
?
CROZTABS

PLEASE ENTER AN OPTION NAME
?
SELECT

PLEASE ENTER AN OPTION NAME
?
SELSTAT
```

FIGURE 2: Initial Output of Interactive Data Statement Program

The following limitations apply to the Run Parameters:

a) Initial Record for UDS 2 (see Chapter 13)

This value should normally be given as 1 except when option UPDATE (see Chapter 11) is to be used, in which case it should be set to the record number of the first record in the set of test data to be analysed.

A zero value may not be provided.

The range of values is 1 to 12000, but there is an upper limit of:

$$\text{Limit} = 12000 - \left[(a \times b) + a + (b \times 15)\right] \qquad (5.1)$$

where a is the number of testees

and b is the number of items in the test

For example, if 150 testees take a 30-item test, Equation 5.1 gives

$$\begin{aligned}\text{Limit} &= 12000 - \left[(150 \text{x} 30) + 150 + (150 \text{x} 30)\right]\\ &= 12000 - 5100\\ &= 6900\end{aligned}$$

So the initial record may not be specified as greater than 6900, as there will be insufficient room in the allotted storage area for the data.

b) Number of options

No more than six (6) options may be specified for any one run of the IBS. Option UPDATE may only be specified once under normal circumstances.

A zero value may not be specified.

c) Option names

The names of options must be given exactly, without mis-spellings or abbreviations. The names are given below, with the relevant Chapter references:

Option name	Description	Chapter
BANKLIST	Frequency tables of data types	6
CROZTABS	Crosstabulated frequency tables	7
SELECT	Selection of items from Bank	8
STATIST	Correlation of two Bank variables	9
SELSTAT	Selection and correlation	10
UPDATE	Analysis of test data	11

Figure 3: Option names and descriptions

Other names will not be accepted and the option ignored.

After the provision of these Run Parameters, control passes to the individual option data inputs, one after the other, for each option selected. These inputs are considered separately in the relevant chapters.

7. Codebook

The User has, during the course of designing his DS, access to the System Codebook. He may request a list of variable numbers, labels and maximum values, similar to Table 1, and he may have descriptions of what specific codes or values mean for certain variables. Examples of codebook references are given in Figures 4 and 5 below.

8. Comments and Titles

Each option used must be preceded by a comment or title, which is output as a heading for the option. This may be up to 72 characters in length. (A blank line may be inserted if required but a title is recommended as a reminder to the User as to the nature of his job).

```
DO YOU WISH TO CONSULT THE CODEBOOK
?
YES

DO YOU WISH FOR A LIST OF THE BANK VARIABLE NUMBERS    & LABELS ?
?
YES

NUMBER          LABEL          MAX. VALUE
   1            ID             NO MAX. VALUE
   2            IND               38
   3            CONT               6
   4            JOB                8
   5            MAT                8
   6            FIRM             118
   7            SIZE               7
   8            IOT                3
   9            KEY            NO MAX. VALUE
  10            TEXT               4
  11            LAST              30
  12            NUM            NO MAX. VALUE
  13            NUM            NO MAX. VALUE
  14            NALT              10
  15            OMIT           NO MAX. VALUE
  16            BOX1           NO MAX. VALUE
  17            BOX2           NO MAX. VALUE
  18            BOX3           NO MAX. VALUE
  19            BOX4           NO MAX. VALUE
  20            BOX5           NO MAX. VALUE
  21            BOX6           NO MAX. VALUE
  22            BOX7           NO MAX. VALUE
  23            BOX8           NO MAX. VALUE
  24            BOX9           NO MAX. VALUE
  25            BOX0           NO MAX. VALUE
  26            CORE           NO MAX. VALUE
  27            LING               4
  28            REPL           NO MAX. VALUE
  29            DISC           NO MAX. VALUE
```

Figure 4: Interactive Program Reference to Codebook: I

```
DO YOU WISH FURTHER INFORMATION ?
?
YES

PLEASE TYPE IN THE NUMBER (BETWEEN 1 AND 30) OF THE BANK VARIABLE
FOR WHICH YOU REQUIRE DETAILS
?
8

PLEASE TYPE IN THE VALUE OR CODE FOR WHICH YOU REQUIRE DETAILS
?
3

                    VARIABLE INT   ,NUMBER  8
                    ************************

       CODE OR VALUE                 DESCRIPTOR
       *************                 **********
               3          ACTION-TYPE

DO YOU WISH FOR DETAILS OF OTHER CODES ?
?
YES

PLEASE TYPE IN THE NUMBER (BETWEEN 1 AND 30) OF THE BANK VARIABLE
FOR WHICH YOU REQUIRE DETAILS
?
12

PLEASE TYPE IN THE VALUE OR CODE FOR WHICH YOU REQUIRE DETAILS
?
13

                    VARIABLE NOQ   ,NUMBER 12
                    ************************

       CODE OR VALUE                 DESCRIPTOR
       *************                 **********
              13          THIS IS THE QUESTION NUMBER IN LAST TEST

DO YOU WISH FOR DETAILS OF OTHER CODES ?
?
NO

 END OF THIS REFERENCE TO CODEBOOK
```

Figure 5: Interactive Program Reference to Codebook: II

Option BANKLIST : summary tables of the frequency of item-types

Introduction

BANKLIST is a facility whereby a list of each value or code on certain variables is produced, with the number of items with that code beside it. This is obviously of great use in the initial preparation stages of test design. It would be of little use designing a test for which the Bank contained no items of the required types!

The values or codes for ten (1Ø) variables are output. These are default variables, representing the ten most important classification types, already built in to the system. These are:

Variable no.	Label	Description
2	IND	Industrial category
3	CONT	Content-type
4	JOB	Job-type
5	MAT	Format of material
7	SIZE	Size of originating firm
8	IQT	Item question type
1Ø	TEXT	Order of item on page
11	LAST	Code of last test
14	NALT	No. of answer boxes
27	LING	Linguistic task

These variables can, however, be replaced by the User for a single run with others in which he may be more interested (see below). It is not appropriate to provide a continuous variable (ie. any of NUM to CORE inclusive) for this option, as the list would be exceedingly long and largely comprising zeroes, with frequency counts for each integer value up to the maximum provided.

No more than ten variables can be provided for overriding the default values. If less than ten are provided, then only the default variables up to and including that value will be overridden. Subsequent default variables will continue to output.

Option data requirements

It is clear that the data requirements for this option are very small. The User must specify whether any variables are to be overridden and, if so, how many. If they are to be overridden, a list of the relevant variables and their respective maximum values is required by the interactive system.

Example of the Interactive System for this option is given in Figures 6 and 7 below

```
OPTION NUMBER 5 : BANKLIST

PLEASE PROVIDE A TITLE OR COMMENT
?
THIS INVOLVES THE SUMMATION AND OUTPUT OF FREQUENCIES

HOW MANY DEFAULT VALUES DO YOU WISH TO REPLACE ?
TYPE Ø FOR DEFAULT TO OPERATE
?
Ø
```

FIGURE6: Interactive Data Statement Program BANKLIST Default

```
OPTION NUMBER 3 : BANKLIST

PLEASE PROVIDE A TITLE OR COMMENT
?
   THIS OPTION CALCULATES FREQUENCIES OF NUMBER OF ITEMS OF EACH TYPE

HOW MANY DEFAULT VALUES DO YOU WISH TO REPLACE ?
TYPE Ø FOR DEFAULT TO OPERATE
?
2

PLEASE BE READY TO ENTER VARIABLE AND MAXIMUM  VALUES

PLEASE ENTER VARIABLE NUMBER
?
28

PLEASE ENTER VALUE
?
173

PLEASE ENTER VARIABLE NUMBER
?
16

PLEASE ENTER VALUE
?
1Ø
```

FIGURE 7: Interactive Data Statement Program
BANKLIST Override

Option results output

Sample outputs from option BANKLIST are shown in Figures 8 and 9.

```
                         OPTION NUMBER 1
                         ******************

                         BANK LISTINGS


TITLE            SAMPLE OUTPUT FROM OPTION BANKLIST
*******   ********************************************************************************


               DEFAULT VALUES HAVE NOT BEEN OVERIDDEN
                           THESE ARE

               VARIABLE NUMBER     LABEL     MAXIMUM VALUE

                      2            IND             38
                      3            CONT             8
                      4            JOB              7
                      5            MAT              7
                      7            SIZE             6
                      8            IQT              3
                     10            TEXT             4
                     11            LAST            10
                     14            NALT            10
                     27            LING             4
```

Figure 8 :Sample output from Option BANKLIST showing default variables

VARIABLE NAME : CONT

VALUE OR CODE	NUMBER OF ITEMS
1	0
2	62
3	77
4	17
5	7
6	10
7	0
8	0

VARIABLE NAME : JOB

VALUE OR CODE	NUMBER OF ITEMS
1	72
2	0
3	32
4	25
5	19
6	22
7	3

Figure 9 : Sample output from Option BANKLIST

CHAPTER 7

Option CROZTABS : crosstabulation of discrete Bank variables

Introduction

Crosstabulation is essentially a two-dimensional form of BANKLISTing. Whilst enumerating items with each code or value of a given variable, it enumerates also for the corresponding codes or values of a second variable. Users of SPSS will be familiar with the CROSSTABS procedure, of which this is a much simpler version.

The use of this option is indicated for more specific data combination requests in the construction of tests. That is, should the User require more complex combinations of characteristics than are detailed by BANKLIST, several CROZTABS operations will give him frequency counts of items matching his requirements. This is less time-consuming for the User than a SELECT option run and the attendant screening of items afterwards.

Option data requirements and formats

Simplest of the options, CROZTABS requires only the specification of the variable labels as alphamerics to operate.

Whilst there is no specific restriction on which variables may be specified, variables having in excess of 39 codes or values will cause a FORTRAN execution error. This option is not considered useful for continuous data, it is recommended for the classificatory data of variables 2 to 11.

An example of the Interactive System for this Option is given in Figure 10 below.

```
OPTION NUMBER 6 : CROZTABS

PLEASE PROVIDE A TITLE OR COMMENT
?
THIS OPTION OUTPUTS TABLES OF CROSSTABULATED VARIABLES

ON REQUEST,ENTER TWO VARIABLE LABELS

ENTER A LABEL
?
CONT

ENTER A LABEL
?
JOB
```

FIGURE 10: Interactive Data Statement Program
Option CROZTABS input

Option results output

Whereas the order of the variables is unimportant for the operation of the option, it does have some effect on output formats.

The first variable specified forms the row variable; the second the column variable. Should the row variable have in excess of nine (9) codes or values, cell size is reduced and nineteen codes plus row totals are printed per row. Should this still be insufficient, a continuation table is printed on the next page.

The column variable may be of any length (up to 39), but it is recommended for clarity of presentation that the larger variable is input second, to give larger cell size if possible.

Portions of sample output are given in Figures 11 and 12.

TITLE THIS IS A RUN FOR SAMPLE OUTPUT FROM CROZTABS

CROSSTABULATIONS

THE FOLLOWING VARIABLES HAVE BEEN SUPPLIED :

ROWS : LING COLUMNS : JOB

VARIABLE NAME : JOB

	1	2	3	4	5	6	7	8	9
1	33	0	27	10	11	5	2	0	88
2	28	0	1	15	2	16	1	0	63
3	0	0	0	0	0	0	0	0	0
4	11	0	4	0	6	1	0	0	22
5	72	0	32	25	19	22	3	0	173

Figure 11 : Sample output from Option CROZTABS (Large Cell)

CROSSTABULATIONS

THE FOLLOWING VARIABLES HAVE BEEN SUPPLIED :

ROWS : JOB COLUMNS : IND

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	10	21	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	21	1
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	4	0	0	2	0	0	0	11	0	0
4	0	0	0	0	0	3	0	0	0	0	0	0	0	4	3	12	0	0	0	0
5	0	2	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	9	0	3	1	0	0	0	0	0	0	0	0	0	5	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	12	30	3	3	5	0	0	0	0	10	0	0	6	3	17	0	11	21	1

Figure 12 : Sample output from Option CROZTABS (Small Cell)

OPTION SELECT : the selection of items on the basis of given characteristics

Introduction

To choose a set of items from the Bank means that the User must specify a set of characteristics to which those items must conform. Essentially, a 'characteristic' is a given value or code on a given variable. For example, one characteristic of the required items might be that they were derived from material used in the Engineering industry; ie. that the code on variable 2 (Industry) is 3 (Engineering).

There are four types of characteristic: that the given value equals that of the corresponding variable for the Bank item; the given value is less than the value for the Bank item; that it is greater than that value; or it is not equal to that value.

Take, for example, the number of pupils correctly responding to each item (variable 26). We may require items with, say, 34 persons correctly responding to them: ie. our given value, 34, must equal the value for variable 26 on any item for it to be selected (it must be 34, as well). Perhaps we are looking for more difficult items and require items whose value for variable 26 is less than 34; or easy items, where the value on variable 26 is greater than 34. Or perhaps 34 correct occurs very often, and we wish to look at other items, so the value on variable 26 must not equal 34.

This, then, is the rationale of the selection option. There is, however, one more simple thing we may require. It may be the case that we are not too sure about the types of items we want, or we know, via BANKLIST or CROZTABS, that there are not enough items with all our characteristics. We must 'weaken' our criteria, therefore. We do this by accepting items with, say, two out of three characteristics, ie. they only have to fulfill some criteria. This we can do for 'equals', 'less than', 'greater than' and 'not equal' criteria.

Option data requirements

The User may specify the following data:

1. A 'REQUIREMENT', of a certain number of items from Ø to the Bank maximum (see 'Run Parameters'). Items will be output up to the value of 'REQUIREMENT', or, if it is Ø, all conforming items will be listed.

2. 'EQUALS CRITERIA', that is, the number of variables and associated values that will be specified. More than one use of a variable counts as a separate criteria. A maximum of 30 criteria is allowed.

3. 'LESS THAN CRITERIA', as above.

4. 'GREATER THAN CRITERIA', as above.

5. 'RESTRICTIONS', as above.

6. 'FULFILL' values : for each of 2 to 5 inclusive, above, the number of criteria which must be fulfilled for the item to be selected. To be omitted if the relevant criteria value is Ø.

7. 'FULFILL' subsequent levels : here, an added facility is that, for 'EQUALS' criteria alone, where REQUIREMENT is not equal to Ø, the User may specify even weaker fulfillment levels. If insufficient items of the first fulfillment level are found, second level items are output until the requirement is met or the Bank is exhausted. The items conforming to the third fulfillment level are output, if the requirement is still not met.

8. 'OUTPUT' : the number of lines to be output per item.

9. Variables and associated values: for each criterion specified above, a list of the variable numbers and associated values.

Option data formats

All SELECT data are numerical.

The following restrictions also apply:

i) If any criteria value is Ø, its FULFILL value is omitted;

ii) No FULFILL value may equal Ø, as this is either meaningless or results in whole-Bank output, depending on the other criteria.

iii) If REQUIREMENT is Ø, the FULFILL subsequent level values(7 above) are omitted. If it is not Ø, they must be included, even if their FULFILL values are the same as FULFILL 1st Level.

iv) Fulfillment is combinatorial; that is, to be selected, an item must satisfy all sets of selection criteria to the relevant FULFILL values. Just conforming to EQUALS and not LESS THAN criteria will not cause the item to be selected, for example.

v) No FULFILL value may exceed that of its related criterion type.

An example of the Interactive System for this option is given in Figure 13 below.

Option results output

The following information is printed out:

1. Title, as specified by COMMENT

2. A list of the relevant selection criteria and variables with associated values;

3. Selected items, in sets of five, in columnar format, with labels on the left-hand side.

4. If REQUIREMENT is not equal to Ø; after the first level items, if the REQUIREMENT is not fulfilled, a list of the items available at all three fulfillment levels.

5. This is followed by output up to requirement or until no further items can be found.

6. A list of the number of items actually output at each level is then printed.

Sample outputs are given in Figures 14 and 15.

```
PLEASE PROVIDE A TITLE OR COMMENT
?
THIS IS SAMPLE INPUT FOR OPTION SELECT

DO YOU WISH TO CONSULT THE CODEBOOK
?
NO

HOW MANY ITEMS DO YOU REQUIRE ?
PLEASE ENTER A SPECIFIC VALUE OR ZERO FOR A GENERAL SEARCH
?
16

HOW MANY EQUALS CRITERIA WILL BE USED ?
?
3

FULFILLMENT OF HOW MANY IS REQUIRED ?
?
3

AND AT SECOND LEVEL ?
?
2

AND AT THIRD LEVEL ?
?
1

HOW MANY GREATER THAN CRITERIA WILL BE USED ?
?
2

FULFILLMENT OF HOW MANY IS REQUIRED ?
?
2

HOW MANY LESS THAN CRITERIA WILL BE USED ?
?
1

FULFILLMENT OF HOW MANY IS REQUIRED ?
?
1

HOW MANY RESTRICTIONS WILL BE USED ?
?
1

FULFILLMENT OF HOW MANY IS REQUIRED ?
?
1

HOW MANY LINES OF OUTPUT ARE REQUIRED PER ITEM
?
29

?
5

PLEASE ENTER VALUE
?
2

PLEASE ENTER VARIABLE NUMBER
?
6

PLEASE ENTER VALUE
?
1

PLEASE ENTER VARIABLE NUMBER
?
27

PLEASE ENTER VALUE
?
1

GREATER THAN CRITERIA

PLEASE ENTER VARIABLE NUMBER
?
16

PLEASE ENTER VALUE
?
23

PLEASE ENTER VARIABLE NUMBER
?
26

PLEASE ENTER VALUE
?
5

LESS THAN CRITERIA

PLEASE ENTER VARIABLE NUMBER
?
15

PLEASE ENTER VALUE
?
10

RESTRICTIONS

PLEASE ENTER VARIABLE NUMBER
?
2

PLEASE ENTER VALUE
?
3
```

Figure 13. Sample of Interactive System input for Option SELECT

```
TITLE      THIS IS A RUN TO GIVE SAMPLE OUTPUT FOR OPTION SELECT
*******    *******************************************************************************

                        NUMBER OF ITEMS REQUIRED                                10

                        NUMBER OF EQUALS CRITERIA                                2

                        NUMBER OF CRITERIA TO BE FULFILLED (1ST LEVEL)           2

                        NUMBER OF CRITERIA TO BE FULFILLED (2ND LEVEL)           1

                        NUMBER OF CRITERIA TO BE FULFILLED (3RD LEVEL)           1

                             VARIABLE          CODE OR VALUE
                               IND                   2
                               JOB                   1

                        NUMBER OF GREATER THAN CRITERIA                          1

                        NUMBER OF CRITERIA TO BE FULFILLED                       1

                             VARIABLE          CODE OR VALUE
                               OMIT                  2

                        NUMBER OF RESTRICTIONS                                   1

                        NUMBER OF CRITERIA TO BE FULFILLED                       1

                             VARIABLE          CODE OR VALUE
                               LING                  1
```

Figure 14 Sample output from Option SELECT

VARIABLE LABEL	FIRST ITEM	SECOND ITEM	THIRD ITEM	FOURTH ITEM	FIFTH ITEM
ID	36	37	38	39	43
INC	33	16	16	16	38
CONT	4	4	4	4	4
JOB	6	4	4	4	6
MAT	1	2	2	1	1
FIRM	1	96	96	112	1
SIZE	6	4	4	5	6
IQT	1	1	1	1	1
KEY	2000	5000	1000	3000	5000
TEXT	1	1	2	3	2
LAST	13	12	12	12	13
NOQ	10	8	9	10	11
NUM	30	10	10	10	30
NALT	5	5	5	5	5
OMIT	2	0	0	2	2
BOX1	0	1	9	1	0
BOX2	28	0	0	2	0
BOX3	0	0	1	6	0
BOX4	0	3	0	0	0
BOX5	0	6	0	0	28
BOX6	0	0	0	0	0
BOX7	0	0	0	0	0
BOX8	0	0	0	0	0
BOX9	0	0	0	0	0
BOX0	0	0	0	0	0
CORE	28	6	9	6	28
LING	2	2	2	2	2
REPL	0	0	0	0	0
DISC	0	0	0	0	0

Figure 15 : Sample output from Option SELECT

CHAPTER 9

OPTION STATIST : the intercorrelation of Item Bank variables

Introduction

Relationships within the Item Bank are usually worthy of investigation, as an indication of the effects of different items on performance, or intra-performance measures.

Correlations, of course, are best done with continuous rather than discrete variables and with higher levels of measurement than nominal (preferably higher than ordinal). It is recommended, therefore, that certain variables are not the subject of this option; in particular the classification data in variables 2 to 8 inclusive. For these, the use of CROZTABS and hand-calculation of relevant statistics is suggested.

Option STATIST correlates the values of two specified variables for all items in the Bank, using the product-moment coefficient, r_{AB}, calculated by

$$r_{AB} = \frac{[(N.\Sigma(A.B)) - (\Sigma A.\Sigma B)]}{\sqrt{[(N.\Sigma A^2) - (\Sigma A)^2].[(N.\Sigma B^2) - (\Sigma B)^2]}}$$

Option data requirements

The data required for this option are simply the specification of the two Bank variables to be correlated. The same variable may not, however, be correlated with itself, as this would be meaningless.

An example of the Interactive System for this option is given in Figure 16, below.

```
OPTION NUMBER 4 : STATIST

PLEASE PROVIDE A TITLE OR COMMENT
?
THIS OPTION CORRELATES TWO VARIABLES

WHICH VARIABLE NUMBERS DO YOU REQUIRE ?
PLEAE TYPE IN THE FIRST NUMBER
?
13

AND THE SECOND NUMBER ?
?
16
```

FIGURE 16: Interactive Data Statement Program
Option STATIST input

Figure 17 shows a typical output from Option STATIST

```
                                                    OPTION NUMBER 1
                                                    ******************


TITLE      THIS IS A RUN TO PROVIDE SAMPLE OUTPUT FROM STATIST
*******    ********************************************************************************


                                                    STATISTICS

                                  CORRELATION OF VARIABLES WITHIN THE ITEM BANK

                                     FIRST VARIABLE            SECOND VARIABLE

                                         OMIT                        BOX1

A VALUE OF 9.9 WILL BE PRINTED WHERE A COEFFICIENT CANNOT BE CALCULATED

                                                          *
                                        VARIABLE ONE      *       VARIABLE TWO
                                        **********************************
          SUM OF VALUES                       264.00      *          1479.00
          MEAN VALUE                            1.53      *             8.55
          STANDARD DEVIATION                    3.79      *            26.05

          NUMBER OF ITEMS                 173
          SUM OF SQUARES                 6568.00
          CORRELATION                       0.25
```

Figure 17: Sample output from Option STATIST

CHAPTER 10

Option SELSTAT : selection of items and the intercorrelation of given variables for those items

Introduction

This option is a combination of options SELECT and STATIST (q.v.) and allows the User to perform correlations on the data of items conforming to given characteristics.

This option is particularly useful for examining relationships for, say, items used in a test, over several testing sessions. The normal output from UPDATE (q.v.) would not give as full a picture as the updated Bank intercorrelations on various variables.

Also, items which have been replaced, or which have certain undesirable characteristics can be eliminated before calculation.

Option data requirements

The data requirements for this option are a combination of those of its parent options, with the following changes:

1. There is no REQUIREMENTS value

2. There are no FULFILL subsequent levels for EQUALS CRITERIA

3. There is no OUTPUT value

4. The variable numbers for intercorrelation come <u>after</u> the variable and associated value requests, and are <u>not</u> preceded by a COMMENT

All previous restrictions on size apply.

An example of the Interactive System for this option is given in Figure 18 overleaf.

```
OPTION NUMBER 2 : SELSTAT

PLEASE PROVIDE A TITLE OR COMMENT
?
    THIS OPTION SELECTS GIVEN ITEMS & THEN CORRELATES TWO SETS OF VALUES

HOW MANY EQUALS CRITERIA WILL BE USED ?
?
1

FULFILLMENT OF HOW MANY IS REQUIRED ?
?
1

HOW MANY GREATER THAN CRITERIA WILL BE USED ?
?
0

HOW MANY LESS THAN CRITERIA WILL BE USED ?
?
0

HOW MANY RESTRICTIONS WILL BE USED ?
?
1

FULFILLMENT OF HOW MANY IS REQUIRED ?
?
1

EQUALS CRITERIA

PLEASE ENTER VARIABLE NUMBER
?
4

PLEASE ENTER VALUE
?
1

RESTRICTIONS

PLEASE ENTER VARIABLE NUMBER
?
2

PLEASE ENTER VALUE
?
3

WHICH VARIABLE NUMBERS DO YOU REQUIRE ?
PLEASE TYPE IN THE FIRST NUMBER
?
15

AND THE SECOND NUMBER ?
?
25
```

FIGURE 18: Interactive Data Statement Program
Option SELSTAT input

Sample output from this option is given in Figure 19.

```
NUMBER OF EQUALS CRITERIA                                  1

NUMBER OF CRITERIA TO BE FULFILLED (1ST LEVEL)             1

          VARIABLE               CODE OR VALUE
            IND                       11

        FIRST VARIABLE                SECOND VARIABLE

           OMIT                           CORE


                                                  VALUE FOR
  COUNTER          ITEM NUMBER          VARIABLE 1        VARIABLE 2
     1                 7                    8                 90
     2                14                    0                 10
     3                15                    0                 10
     4                16                    0                 10
     5                81                    1                  3
     6                82                    1                  4
     7                83                    1                  3
     8               146                    0                  8
     9               147                    1                  7
    10               148                    0                  8

                                    *
                 VARIABLE ONE       *        VARIABLE TWO
                 *************************************
SUM OF VALUES              12.00    *         153.00
MEAN VALUE                  1.20    *          15.30
STANDARD DEVIATION          2.44    *          26.39

NUMBER OF ITEMS            10
SUM OF SQUARES            737.00
CORRELATION                 0.95
```

Figure 19: Sample output from Option SELSTAT

CHAPTER 11

Option UPDATE : the scoring of tests and related statistical functions

Introduction

This is by far the most complex of the options, with several optional routines available.

The task of the routine is to read pupil-item-responses from a User Data Set (UDS) and produce test scores for each pupil and report on the performance of each item and the whole test. As items are not of common format nor the test a standardised one, the facilities available under this option are overinclusive in terms of the data the User may specify. That is, the criteria for the acceptability of a test of this type are by no means universally agreed and so output reflects several differing views in the field.

Sub-options

There are thirteen (13) sub-options in UPDATE, some of which are independent, others requiring a precursor sub-option to operate.

1. UPDATE : this sub-option causes the relevant item-data to be added to the Bank itself, including the number taking the test, responses for each answer box and the number of correct responses. It is recommended that a trial run of any data is performed before including this sub-option, as an extra check to prevent corruption of the Bank. In fact, UPDATE is performed last of the sub-options, so that any errors cause option rejection before updating.

2. RETEST : a second set of pupil-item-responses for subjects taking the same test on more than one occasion is read from the UDS.

3. SCORES : data for each pupil for the test (and retest if invoked) to be output, with total score, scores on (up to) 18 subtests and the percentages of the whole for each of these.

4. GROUP : total raw scores for test (and retest if invoked) to be printed out in a table with mean(s) and standard deviation(s).

5. ITEMDATA : a table of item-performance is printed, showing the number taking the item, those omitting, those selecting each answer box and those answering correctly (and total discrepancies, if RETEST is invoked).

6. QUALITY : this sub-option causes statements of item quality to be output with the ITEMDATA sub-option, which must have been previously specified. High levels of omission ($>$ 5%) or distractor selection ($>$ 20%) are signalled for Multiple-choice items only.

7. GRPNAMES : the names or labels of subjects are read from UDS and are output with sub-options SCORES and GROUP, if previously specified.

8. TESTCHAR : a table showing the number of items in the test and each of the subtests is printed out.

9. DISCREPS : this sub-option causes a table of discrepancy values for each item to be printed. A discrepancy is the answering of an item in a different way on two occasions. A '+ve' discrepancy is wrong first time, right second; a '-ve' discrepancy is right first, wrong second; an 'incorrect' discrepancy is wrong both times; a 'correct' discrepancy is right both times. This sub-option may only be invoked if RETEST has been specified. A table giving total percentages of the types of discrepancy is also output.

10. RESPONSE : a table of response patterns is listed for each item-response-pattern, with the frequencies of each. This is principally of use for 'action-type' items, where QUALITY is of no use, but also indicates the patterns for multiply-answered multiple-choice items.

11. MATRICES : the intercorrelation matrix of scores and sub-scores is printed (with a similar table for RETEST if invoked and the test-retest matrix).

Calculations are based on

$$r_{AB} = \frac{[(N.\Sigma(A.B)) - (\Sigma A.\Sigma B)]}{\sqrt{[(N.\Sigma A^2) - (\Sigma A)^2].[(N.\Sigma B^2) - (\Sigma B)^2]}}$$

12. DISCRIMS : discrimination indices are output for each item, both for the whole test and for each subtest. Point-biserial correlation is used:

$$r_{pbis} = \frac{\mu_{corr} - \mu_{incor}}{\sigma_T . \sqrt{p.q}}$$

item

μ_{incor} = mean score of those incorrectly answering item

σ_T = standard deviation of test scores

p = proportion answering correctly

q = proportion answering incorrectly

In criterion-referenced measurement, the discrimination of an item is of little importance. A negative discriminator, particularly within a subtest, is an indicator of a poor item, however, as it is one that is more often correctly answered by the less able than the more able, in terms of the total test score. DISCRIMS is available for RETEST, if invoked.

13. KAPPA : the coefficient κ (Cohen, 1960) can be used to establish the reliability of decisions made on the basis of test score or sub-score. Its use in this kind of measurement has been delineated by Swamanithan et al, (1974),and, although there remains some doubt as to its efficacy, it is available here. The coefficient measures the degree of agreement between total test scores on two occasions, adjusting, via cross-marginal proportions, for chance.

For a full exposition see Swamanithan et al (1974). The calculation of κ is by

$$\kappa = (prop_o - prop_c) / (1.0 - prop_c)$$

where $prop_o$ = the observed proportions of agreement and $prop_c$ = the proportion of agreement expected by chance. This coefficient has a standard error, σ , which is also reported:

$$\sigma = \frac{\sqrt{[prop_o \, . \, (1.0 - prop_c)]}}{[N \, . \, (1.0 - prop_c)]}$$

The related coefficient, ϕ , is also output, based on

$$\phi = \frac{MM.NN - MN.NM}{(MM + NN)(MM + NM)(NM + NN)(MN + NN)}$$

Where MM is the proportion at or above cutting score on both occasions;

NN is the proportion below on both occasions;

MN is the proportion at or above on the first occasion and below on the second and

Option data requirement

The two most obvious requirements are the number of testees and the number of i5ems in the test. The maximum for these are 25Ø and 35 respectively.

Following these data, the system needs information on how many and which sub-options are to be used. Further, for UPDATING purposes, an identifier is needed, 'TEST ID'. If RETEST is to be used, the starting point on UDS for the second set of pupil-item-responses mustrbe specified.

This starting point - the record number of the first record of the RETEST data - must clearly not overlap with the first set of data. It therefore may not be less than the last record of that set. Hence, a lower limit may be defined as:

$$\text{Lower Limit} = a.b + a + 15b + 1$$

where a is the number of subjects and
b is the number of items in the test.

A maximum of 50 subjects is allowed for a RETEST analysis and the size of the UDS should be quite sufficient to hold both test and retest data. If there are several sets of data on the UDS, however, it is necessary to specify an upper limit for the RETEST initial record number. This is

$$\text{Upper Limit} = 12000 - \text{Lower Limit}$$

This prevents the IBS attempting to analyse data beyond the extent of the UDS.

In fact, these restrictions are included only as a reminder to the User to be careful in his presentation of data. It is clearly impossible for data to have been placed on the UDS beyond its extent, for the IBS to analyse; but it is worthwhile that the User be informed of the above limits.

Option data format

Data for this option are a mixture of alphamerical and numerical.

The number of subjects and items are numerical and specified immediately after the COMMENT card.

The number of sub-options is alphamerical and may take values from ONE to TWELVE, or ALL for thirteen sub-options. Following this, the sub-option names are given alphamerically.

TEST ID is numerical as is the UDS RETEST record number.

An example of the Interactive System for this option is given in Figure 20, below.

```
DATA REQUIREMENTS FOR OPTIONS

OPTION NUMBER 1 : UPDATE

PLEASE PROVIDE A TITLE OR COMMENT
?
    THIS OPTION ANALYSES TEST DATA

PLEASE ENTER NUMBER OF SUBJECTS
?
20

AND THE NUMBER OF ITEMS PER TEST
?
31

PLEASE ENTER,AS A WORD,THE NUMBER OF SUBOPTIONS YOU REQUIRE
"ALL" IS ACCEPTABLE FOR ALL THIRTEEN
?
THREE

ON REQUEST,ENTER DESIRED SUBOPTION NAMES

ENTER OPTION NAME
?
SCORES

ENTER OPTION NAME
?
GROUP

ENTER OPTION NAME
?
ITEMDATA

ENTER TEST ID NUMBER
?
13
```

FIGURE 20: Interactive Data Statement Program
Option UPDATE input

Option results output

Sample outputs are given below, in figures 22 to 31.
They show different combinations of sub-options to demonstrate in particular the mutually dependent sets, such as QUALITY, GRPNAMES etc.

References

Cohen J (1960)	A Coefficient of agreement for nominal scales Educational & Psychological Measurement Vol XX, No.1 pp 37-46
Swaminathan H, Hambleton R K & Algina J (1974)	Reliability of Criterion-Referenced Tests: a Decision-Theoretic Formulation. Journal of Educational Measurement Vol.11, No.4, pp 263-267

TEST CHARACTERISTICS
****

TEST CODE NUMBER : 13

TOTAL NUMBER OF ITEMS : 31

CONTENT TYPE SUBTESTS

CODE	NUMBER
1	0
2	5
3	10
4	6
5	4
6	6

JOB TYPE SUBTESTS

CODE	NUMBER
1	14
2	0
3	5
4	4
5	2
6	6
7	0
8	0

LINGUISTIC TASKS

CODE	NUMBER
1	17
2	13
3	0
4	1

Figure 21 : **Output using suboption TESTCHAR**

```
                         RESULTS FOR TESTEE NUMBER    300
                         ********************************
1.TEST

     NO. OF ITEMS TAKEN       31
     NO. OF ITEMS CORRECT     21
     PERCENTAGE CORRECT       67.74

          SUBTESTS
              1. CONTENT TYPE
                   CODE        SCORE        PERCENTAGE
                    2            4            80.00
                    3            7            70.00
                    4            4            66.67
                    5            1            25.00
                    6            5            83.33

              2. JOB TYPE
                   CODE        SCORE        PERCENTAGE
                    1           10            71.43
                    3            2            40.00
                    4            2            50.00
                    5            2           100.00
                    6            5            83.33

              3. LINGUISTIC TASK
                   CODE        SCORE        PERCENTAGE
                    1           13            76.47
                    2            8            61.54
                    4            0             0.0
```

Figure 22 **: OUTPUT using suboptions SCORES**

```
***********************************
1.TEST

        NO. OF ITEMS TAKEN        30
        NO. OF ITEMS CORRECT      24
        PERCENTAGE CORRECT        80.00

                SUBTESTS
                        1. CONTENT TYPE
                            CODE          SCORE          PERCENTAGE
                             2             11              91.67
                             3             10              71.43
                             4              3             100.00
                             6              0               0.0

                        2. JOB TYPE
                            CODE          SCORE          PERCENTAGE
                             1              1              33.33
                             2              2             100.00
                             3              1              50.00
                             4              9              90.00
                             5              7              77.78
                             6              4             100.00

                        3. LINGUISTIC TASK
                            CODE          SCORE          PERCENTAGE
                             1             10              71.43
                             2              9             100.00
                             4              5              71.43

 2. RETEST

        NUMBER OF ITEMS TAKEN     30
        NUMBER OF ITEMS CORRECT   27
        PERCENTAGE CORRECT        90.00

                SUBTESTS
                        1. CONTENT TYPE
                            CODE          SCORE          PERCENTAGE
                             2             12             100.00
                             3             11              78.57
                             4              3             100.00
                             6              1             100.00

                        2. JOB TYPE
                            CODE          SCORE          PERCENTAGE
                             1              2              66.67
                             2              2             100.00
                             3              2             100.00
                             4              9              90.00
                             5              8              88.89
                             6              4             100.00

                        3. LINGUISTIC TASK
                            CODE          SCORE          PERCENTAGE
                             1             12              85.71
                             2              9             100.00
                             4              6              85.71
```

Figure 23 :Output using suboptions SCORES and RETEST

```
                    RESULTS FOR     SMITH,JONATHON : TESTEE NUMBER   1300
                    ************************************************************
1. TEST

      NO. OF ITEMS TAKEN       31
      NO. OF ITEMS CORRECT     16
      PERCENTAGE CORRECT       51.61

              SUBTESTS
                   1. CONTENT TYPE
                      CODE       SCORE       PERCENTAGE
                        2          3           60.00
                        3          7           70.00
                        4          3           50.00
                        5          0            0.0
                        6          3           50.00

                   2. JOB TYPE
                      CODE       SCORE       PERCENTAGE
                        1          5           35.71
                        3          3           60.00
                        4          3           75.00
                        5          2          100.00
                        6          3           50.00

                   3. LINGUISTIC TASK
                      CODE       SCORE       PERCENTAGE
                        1         11           64.71
                        2          5           38.46
                        4          0            0.0
```

Figure 24 **: Output using suboptions SCORES and GRPNAMES**

GROUP SCORES FOR TEST AND RETEST
****

SUBJECT	TEST	RETEST
103	24	27
104	22	26
106	19	21
107	20	25
108	27	24
110	26	26
111	26	26
112	21	28
113	21	27
114	22	24
115	25	29
116	24	27
117	23	25
118	24	25
119	25	23
121	23	23
122	23	25
123	24	27
124	26	26
125	26	22
126	27	26
127	25	25
128	28	25
129	27	29
130	26	29
131	23	25
132	28	30
133	27	27
134	23	25
135	27	24
136	22	24
137	23	25
138	22	21
141	27	26
142	23	24
143	25	26
144	27	29
145	29	30
146	28	29
147	23	21
148	25	25
149	22	26
150	26	28
152	26	27

MEAN OF TEST = 24.55 STANDARD DEVIATION = 2.38

MEAN OF RETEST = 25.32 STANDARD DEVIATION = 2.35

Figure 25 **: Output using suboptions GROUP and RETEST**

DISCRIMINATION INDICES
****

A VALUUE OF 9.9 WILL BE PRINTED WHERE A COEFFICIENT CANNOT BE CALCULATED
T-TEST OF ZERO CORRELATIONS SHOULD USE DF = 60.0

WHOLE TEST VALUES

ITEM NUMBER	INDEX	T-TEST VALUE
72	0.32	2.60
165	0.43	3.73
141	0.64	6.44
74	0.31	2.53
7	0.48	4.21
73	0.49	4.36
1	0.46	4.00
37	0.24	1.88
38	0.12	0.97
39	0.36	3.03
108	0.14	1.10
109	0.39	3.28
110	0.55	5.13
94	9.90	9.90
95	0.29	2.39
96	0.28	2.29
47	0.24	1.91
48	0.34	2.79
49	0.35	2.94
2	0.24	1.91
3	0.35	2.91
4	0.11	0.84
104	0.27	2.15
105	0.31	2.56
106	0.38	3.21
130	0.29	2.33
131	0.65	6.55
132	0.34	2.76
127	0.69	7.31
128	0.48	4.26
129	0.34	2.79

SUBTESTS

CONTENT-TYPE 2

ITEM NUMBER	INDEX	T-TEST VALUE
72	0.54	4.91
74	0.55	5.13
104	0.27	2.13
105	0.53	4.83
106	0.59	5.71

Figure 26 **Part of output from suboption DISCRIMS**

CORRELATION MATRIX OF SCORES
****

A VALUE OF 9.9 WILL BE PRINTED WHERE A COEFFICIENT CANNOT BE CALCULATED

MEAN OF SCORES = 18.71 STANDARD DEVIATION = 4.79

		CONTENT CATEGORIES						JOB CATEGORIES								LINGUISTIC TASK			
	TOTAL	1	2	3	4	5	6	1	2	3	4	5	6	7	8	1	2	3	4
TOTAL	1.00	9.90	0.71	0.80	0.71	0.83	0.62	0.93	9.90	0.63	0.57	0.49	0.59	9.90	9.90	0.88	0.92	9.90	0.34
C 1	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90
O 2	0.71	9.90	1.00	0.66	0.30	0.42	0.24	0.57	9.90	0.35	0.21	0.61	0.75	9.90	9.90	0.68	0.62	9.90	0.17
N 3	0.80	9.90	0.66	1.00	0.32	0.58	0.33	0.66	9.90	0.74	0.28	0.58	0.53	9.90	9.90	0.88	0.58	9.90	0.28
T 4	0.71	9.90	0.30	0.32	1.00	0.56	0.32	0.68	9.90	0.39	0.78	0.17	0.18	9.90	9.90	0.37	0.87	9.90	0.26
E 5	0.83	9.90	0.42	0.58	0.56	1.00	0.51	0.85	9.90	0.55	0.49	0.19	0.37	9.90	9.90	0.65	0.80	9.90	0.50
N 6	0.62	9.90	0.24	0.33	0.32	0.51	1.00	0.67	9.90	0.20	0.28	0.21	0.38	9.90	9.90	0.70	0.46	9.90	0.05
T 1	0.93	9.90	0.57	0.66	0.68	0.85	0.67	1.00	9.90	0.46	0.43	0.25	0.48	9.90	9.90	0.78	0.89	9.90	0.32
2	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90
3	0.63	9.90	0.35	0.74	0.39	0.55	0.20	0.46	9.90	1.00	0.31	0.24	0.24	9.90	9.90	0.61	0.51	9.90	0.34
4	0.57	9.90	0.21	0.28	0.78	0.49	0.28	0.43	9.90	0.31	1.00	0.30	0.02	9.90	9.90	0.35	0.68	9.90	0.07
J 5	0.49	9.90	0.61	0.58	0.17	0.19	0.21	0.25	9.90	0.24	0.30	1.00	0.30	9.90	9.90	0.64	0.28	9.90	0.08
O 6	0.59	9.90	0.75	0.53	0.18	0.37	0.38	0.48	9.90	0.24	0.02	0.30	1.00	9.90	9.90	0.58	0.48	9.90	0.28
B 7	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90
8	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90
L 1	0.88	9.90	0.68	0.88	0.37	0.65	0.70	0.78	9.90	0.61	0.35	0.64	0.58	9.90	9.90	1.00	0.64	9.90	0.20
I 2	0.92	9.90	0.62	0.58	0.87	0.80	0.46	0.89	9.90	0.51	0.68	0.28	0.48	9.90	9.90	0.64	1.00	9.90	0.27
N 3	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90	9.90
G 4	0.34	9.90	0.17	0.28	0.26	0.50	0.05	0.32	9.90	0.34	0.07	0.08	0.28	9.90	9.90	0.20	0.27	9.90	1.00

Figure 27: Output using suboption MATRICES

DISCREPANCY VALUES
****

ITEM NUMBER	DISCREPANCIES +VE	CORR	INCR	-VE
107	6	23	11	4
126	10	4	27	3
160	4	28	7	5
167	2	41	0	1
168	7	36	1	0
169	3	36	5	0
133	7	34	1	2
134	3	40	0	1
135	2	40	1	1
36	1	42	0	1
43	1	42	0	1
68	3	40	0	1
8	0	44	0	0
9	8	20	12	4
10	7	35	1	1
33	0	44	0	0
34	2	42	0	0
35	2	42	0	0
84	6	9	26	3
85	2	42	0	0
86	8	18	9	9
26	0	44	0	0
27	1	42	0	1
28	0	44	0	0
40	1	42	1	0
142	7	19	14	4
42	5	31	3	5
122	2	35	5	2
123	7	31	3	3
124	4	35	2	3

Figure 28: Output using suboption DISCREPS

CORRELATION MATRIX OF SCORES
****

A VALUE OF 9.9 WILL BE PRINTED WHERE A COEFFICIENT CANNOT BE CALCULATED

MEAN OF SCORES = 21.92 STANDARD DEVIATION = 5.55

SUBTESTS

CONTENT TYPES

TYPE	MEAN	S.DEV
1	0.0	0.0
2	3.30	1.25
3	6.94	1.86
4	4.30	1.47
5	2.12	1.27
6	5.26	1.23

JOB TYPES

TYPE	MEAN	S.DEV
1	10.80	3.14
2	0.0	0.0
3	2.92	1.18
4	2.70	0.99
5	0.70	0.46
6	4.80	1.23
7	0.0	0.0
8	0.0	0.0

LINGUISTIC TASKS

TYPE	MEAN	S.DEV
1	12.90	3.02
2	8.54	2.85
3	0.0	0.0
4	0.48	0.50

Figure 29 : Part of output from suboption MATRICES

KAPPA COEFFICIENTS

A VALUE OF 9.99 WILL BE PRINTED WHERE A COEFFICIENT CANNOT BE CALCULATED

1. WHOLE TEST

CUTTING SCORE	KAPPA	STANDARD ERROR	PHI
1	9.99	9.99	9.99
2	9.99	9.99	9.99
3	0.0	0.99	9.99
4	0.0	0.99	9.99
5	0.0	0.99	9.99
6	0.0	0.99	9.99
7	0.0	0.69	9.99
8	0.66	0.34	0.70
9	0.66	0.34	0.70
10	1.00	0.00	1.00
11	1.00	0.00	1.00
12	0.64	0.25	0.64
13	0.64	0.25	0.64
14	0.64	0.25	0.64
15	0.63	0.21	0.64
16	0.81	0.13	0.81
17	0.81	0.13	0.81
18	0.76	0.13	0.76
19	0.85	0.10	0.86
20	0.63	0.14	0.68
21	0.65	0.13	0.70
22	0.56	0.13	0.57
23	0.51	0.13	0.52
24	0.66	0.11	0.65
25	0.61	0.12	0.61
26	0.47	0.15	0.48
27	0.46	0.17	0.51
28	0.61	0.16	0.66
29	0.41	0.25	0.50
30	-.02	0.71	-.02
31	0.0	0.99	9.99

CONTENT-TYPE 2

CUTTING SCORE	KAPPA	STANDARD ERROR	PHI
1	0.0	0.56	9.99
2	0.23	0.33	0.23
3	0.41	0.18	0.42
4	0.37	0.14	0.35
5	0.29	0.19	0.30

CONTENT-TYPE 3

CUTTING SCORE	KAPPA	STANDARD ERROR	PHI
1	9.99	9.99	9.99
2	0.0	0.99	9.99
3	0.66	0.34	0.70
4	0.79	0.21	0.81
5	0.85	0.15	0.86
6	0.56	0.19	0.58
7	0.43	0.14	0.47
8	0.32	0.12	0.52
9	0.47	0.13	0.48
10	0.41	0.25	0.50

Figure 30 : Sample output from suboption KAPPA

FREQUENCIES OF SCORES
****

SCORE *****	FREQUENCY *********	PERCENTAGE **********
0	0	0.0
1	0	0.0
2	0	0.0
3	0	0.0
4	0	0.0
5	0	0.0
6	0	0.0
7	0	0.0
8	1	3.23
9	0	0.0
10	0	0.0
11	0	0.0
12	0	0.0
13	1	3.23
14	0	0.0
15	0	0.0
16	1	3.23
17	0	0.0
18	2	6.45
19	4	12.90
20	3	9.68
21	6	19.35
22	3	9.68
23	3	9.68
24	2	6.45
25	1	3.23
26	2	6.45
27	0	0.0
28	1	3.23
29	1	3.23
30	0	0.0
31	0	0.0

SUBTEST FREQUENCIES
****

CONTENT SUBTEST: 2

SCORE *****	FREQUENCY *********	PERCENTAGE **********
0	0	0.0
1	0	0.0
2	6	19.35
3	9	29.03
4	10	32.26
5	6	19.35

CONTENT SUBTEST: 3

SCORE *****	FREQUENCY *********	PERCENTAGE **********
0	0	0.0
1	0	0.0
2	1	3.23
3	0	0.0
4	1	3.23
5	3	9.68

Figure 31 : Sample output of frequency tables (suboption GROUP)

CHAPTER 12

Creation and Deletion of MUSIC Save files

1. MUSIC Save files

   A Save File is a file of data (or a program) placed in storage by the main Polytechnic computer system. It is usually convenient for the User to create Save Files to store data from testing before transfer to UDS 2 (see Chapter 13), or for adding items to the Bank.

   All Save Files that the User needs for use of the System are saved directly from his submission of forms for card punching. That is, when handing in the forms, he is not processing data, but saving it for processing via the interactive submission program.

2. Save Files require a unique 'name' of up to 6 characters, the first of which must be alphabetic, and which must not contain any 'special' (eg punctuation) characters.

3. Processing of all Save Files is undertaken from a terminal, using the SUBMIT routine (Mdoc 2.2/7). The User types

   /EXEC SUBMIT

   to which he received the response:

   SURNAM = '   ', FILE = '   ', CLASS = 'A',
   MINS = Ø, SECS = 2Ø, PAGES = 2Ø

   ENTER SURNAME, FILENAME, AND ANY OTHER CHANGES
   ?

   The User's typical response will be:

   SURNAM = 'SMITH', FILE = 'BRUNS', CLASS = 'T', MINS = 3,
   PAGES = 25Ø

   where FILE = '   ' should contain the name of the work to be processed.

   All SOFRP IBS jobs are class 'T' and a message will appear:

   ENTER OPERATOR MESSAGE (MAX. 40 CHARS)

   to which the User should respond

   DISK = MUSIC7

   The routine returns

JOB NUMBER Rnnn SUBMITTED AT q ON z

where Rnnn is the 'remote', 3 digit job-number, q is the time and z the date.

All submitted files will be returned via the internal mail.

4. Deletion of Save Files

The command at a terminal:

/PURGE nnnnnn

will wipe out a file named: nnnnnn

CHAPTER 13

User Data Sets

1. Introduction

User Data Sets (UDS) are storage facilities physically located on a magnetic disc. The hardware aspects need not concern the Users. Suffice it to say that UDS are involved in System operation, all of which are located on one of two discs: MUSIC2, which is permanently mounted, or MUSIC7 which must be mounted for each system job. This is undertaken at the Computer Services Department and should not concern the User. Jobs will occasionally be delayed or fail when no disc drives are available for MUSIC7. Such work should be resubmitted. It is to be noted that drives are more usually free in the afternoon.

Four UDS are used in each System job. They are

UDS 1 : Item Bank Data File
UDS 2 : Pupil-Item Test Responses
UDS 3 : Data Statement
UDS 4 : Operating program (Load module)

These are discussed in detail below.

2. UDS File 1 : Item Bank Data File

This UDS is used to store the actual numerical sets of data available for each item. 80 columns are set aside for this and up to 250 item data sets may be stored.

This file is security-coded (ie. it may only be used, created or deleted by a User with the correct User Code....(see Chapter 3)). It must not, <u>under any circumstances</u>, be deleted without prior consultation. It exists permanently on MUSIC7 disc.

3. UDS File 2 : Pupil-Item Test Responses

This file is used by Option UPDATE and stores the pupils' names, the response to each item and the response-patterns for multiply-answered items (suboptions GRPNAMES, RESPONSE, SCORES etc).

Even if the Option is not called, it is still necessary for this file to exist. Section 6 deals with the placement of data onto this UDS. Retest data is also held on this UDS. It exists permanently on MUSIC7 disc.

4. <u>UDS File 3 : Data Statement</u>

Chapter 5 has dealt with the creation of this UDS. It holds the sequenced commands for the operation of the system and exists temporarily on MUSIC2 disc.

5. <u>UDS File 4 : Load module</u>

This file holds the System itself, as a set of object decks controlled on the MUSIC overlay system with the Link Editor. Its contents and operation need not concern the ordinary User. It must not, <u>under any circumstances</u>, be deleted without prior consultation. It exists permanently on MUSIC7 disc.

6. <u>Placement of data onto UDS 2</u>

As Chapter 12 Section 5 suggests, special forms are used to make this job as simple as possible. The program 'MARKER' is used to read the responses and transfer them to the UDS file in the format required by the system.

The data required to operate the program are:-

1. number of subjects (max. 25Ø)
2. items in the test (max. 35)
3. status of data (TEST or RETEST)
4. Save File name of job

and then, for each subject

1. subject name and number
2. onwards pupil item responses

The following forms need to be completed:

1. To set up the job: Form CS1 or CS1a
2. For each subject except the last: Form CS2
3. For the last pupil: Form CS3

The program writes data onto UDS 2, but this is of limited size (12000 records). It is therefore assumed that the User will normally <u>only require the UDS to hold one set of test data</u> or one set and its RETEST data. The UDS is therefore <u>blanked</u> if the status of the input data is given as 'TEST' on Form CS1. If more data is to be placed on the UDS <u>without</u> blanking, its status should be given as 'RETEST'. This applies even if the additional set is <u>not</u> the Retest data of the original set. The default status is TEST.

Notes on the completion of the special forms are given in Appendix III

7. Other User Data Sets

Four other UDS are kept. The first has been mentioned above, and it holds certain status records:

i) the number of items in the Bank
ii) the number of data per item
iii) the number of valid test data records on UDS 2

A list of the system status can be obtained via the command, at a terminal,

/EXEC STATUS

The relevant output is shown in Figure 32.

A further UDS holds the program used in creating the Data Statement. It serves to speed up the time of operation. The third extra UDS holds the code descriptors used by the Codebook (see Appendix II or Chapter 5 Section 7). The fourth UDS exists temporarily to hold editing data (Chapter 14).

```
        STATUS OF SOFRP ITEM & VOCAB. BANKING SYSTEMS
        ---------------------------------------------

NUMBER OF ITEMS IN BANK          = 173

DATA PER ITEM                    = 29

NUMBER OF VALID TEST RECORDS     =  2956

NUMBER OF VOCABULARY ITEMS       =   693
```

Figure 32 Output from program STATUS

The first and second lines are used by DATAST program as run parameters

The third line is used by program MARKER to assign an initial record to retest data on UDS

The fourth line refers to the SOFRP Vocabulary Banking System (see Parry, 1980)

Reference:

Parry, O (1980) Manual of the Vocabulary Banking System
Language Development Centre,
Sheffield City Polytechnic.

Chapter 14

Editing the Item Bank

Introduction

As has been previously mentioned, Users are falliable and mistakes are often made. Provision for changing values in the Item Bank is therefore necessary. The User may also wish to recode certain variables on a new or extended set of categories. He may wish to add new items to the Bank, or overwrite old and superceded items (NB. items may not be deleted as the access programs use the fact that item identity number and file position are identical).

Adding Items to Bank

Users should code the data onto Form IB1 (Appendix III) and submit the job for card-punching and execution at Computer Services.

The format of the data is clearly important and Users should pay careful attention to the columns in which they specify the codes or values.

Identifying numbers must be in the correct sequence, following directly on from the last number used. Users should execute the status program (STATUS) for information about the last number used, which will be the same as the number of items in the Bank.

This job uses program BANDAT.

Changing values

To change existing values within the Item Bank an editing program EDTRIB is used in batch submission. This is called up and the job submitted via an interactive program EDITBN.

The User types

/EXEC EDITBN

and the program will ask for

i) the identifying item number
ii) the variable number, and
iii) the new value to replace the existing value.

When the User has no more values, he types in

END

The list of edit commands is then printed out and changes can be made.

A maximum of 50 edits may be made in one editing run.

The program will then ask the User if the edit job is to be submitted for batch processing. If so, the usual procedures apply (see eg. the job submission section of Chapter 5).

The edit program uses a temporary UDS on MUSIC2 disk.

Appendix I

SOFRP Item Bank Error Messages

1. 'Check' messages

On encountering an error in the data input, the System produces an error message, specifying the type of error found. This is followed (after any other messages of the same type) by a message on the far right side of output of the type

"Check n completed, Y errors found"

where n is the check number and Y the number of errors found.

Error messages are listed by check number below.

2. Check 1 errors : Major run parameters

2.1 Bank size parameter:

2.1.1 Message: ERROR IN BANK SIZE PARAMETER. NO. OF ITEMS GIVEN AS 0

Here a major system error may have occurred. The data normally available on the status UDS has been blanked, or UDS File 3 has been inaccurately edited. If former, consult Computer Services Department.

2.1.2 Message: ERROR IN BANK SIZE PARAMETER. n EXCEEDS MAXIMUM ALLOWED

where n is the number of items provided by the status UDS. This is a major systems error and indicates that the Bank addition program BANDAT needs alteration, or UDS File 3 has been inaccurately edited. If former, consult Computer Services Department.

2.2 Item Data Parameter

2.2.1 Message: ERROR IN ITEM SIZE PARAMETER. n IS TOO LARGE

where n is the number of data per item provided by the status UDS. Major systems error is indicated, or inaccurate editing of UDS File 3. If former, consult Computer Services Department.

2.2.2 Message: ERROR IN ITEM SIZE PARAMETER: 0 IS NOT ALLOWED

The data on the status UDS has been deleted or UDS File 3 has been inaccurately edited. If former, consult Computer Services Department.

2.3 User Data Set Parameters

The disk storage files or "User Data Sets"(UDS) are read into the program under the control of a variable called "the associated variable". The associated varibale is the record number to be read next.

2.3.1 Message: "ZERO VALUE OF ASSOCIATED VARIABLE NOT ALLOWED. TRY 1" There cannot be a record numbered zero, so an attempt to read using this value will lead to a "direct access" error (see "D-A errors, Appendix 3). The job aborts, giving the message "Try 1" as a suggested initial value that will work.

2.3.2 Message: "n EXCEEDS MAXIMUM FOR ASSOCIATED VARIABLE FOR UDS 2", where n is the value specified. Refer to Chapter 5 for maximum values. This job must abort or lead to an execution error under FORTRAN (IHN2321. see Mdoc 3.2/10)

2.4 Option Parameter

2.4.1 Message: "NO OPTIONS HAVE BEEN SELECTED. JOB WILL ABORT". Simply, zero has been specified. If allowed to proceed, the SELECT option data input will be invoked, as this is encountered in the program first. Errors would be caused, however, so the job aborts.

2.4.2 Message: "n EXCEEDS THE NUMBER OF OPTIONS AVAILABLE ON ONE RUN", where n is the number specified. Although this job could proceed to do the first six options, the User has misread the documentation and other errors may be considered likely. Refer to Chapter 5.

3. Check 2 Errors : Option Keywords

3.1 Message: "name IS NOT ONE OF THE NAMES OF THE OPTIONS. THESE ARE" (List of option names), where 'name' is the keyword received by the System. User has probably mis-typed. This option will be missed out.

4. Check 3 Errors:Selection Parameter

4.1 Message: "NUMBER OF ITEMS REQUIRED IN SELECTION OPTION EXCEEDS NUMBER SPECIFIED FOR BANK SIZE"

This is self-explanatory. The Run Parameter for Bank Size is less than the number of items required to be selected. As with all errors from now on, this will cause the option to be bypassed.

4.2 Messages of general type:

"NUMBER GIVEN FOR - type - CRITERIA, n, EXCEEDS MAXIMUM ALLOWED (3Ø)"

where - type - is one of EQUALS, GREATER THAN, LESS THAN or RESTRICTIONS and n is the number supplied.

Here again the message is self-explanatory. Chapter 8 gives a maximum of 30 for each of these.

4.3 Messages of general type:

"FULFILLMENT VALUE FOR - type - SPECIFICATION EXCEEDS MAXIMUM ALLOWED(3Ø)"

Similarly, the fulfillment values for the criteria type has exceeded its maximum. Messages also are output for the second and third level fulfillments of equals criteria if REQUIREMENT is specified as non-zero.

4.4 Messages of general type:

"FULFILLMENT VALUE EXCEEDS NUMBER OF - type - CRITERIA"

Here, no level of fulfillment may be more than the number actually achieveable (if it were allowed, no items could be selected, giving later abort messages, in SELSTAT or a null output in SELECT).

4.5 Message: "SECOND LEVEL EXCEEDS FIRST LEVEL FULFILLMENT"
or THIRD or SECOND

The provision of higher levels of selection under the REQUIREMENT specification is not permitted, as it is assumed the User will specify the highest selection level first.

5. Check 4 : Selection Arrays

5.1 Message: " - type - CRITERIA: VARIABLE n OUTSIDE SPECIFICATIONS", where - type - is one of EQUALS, GREATER THAN, LESS THAN, or RESTRICTIONS and n is the number specified. This error is produced when a variable number greater than the Item data parameter is provided.

5.2 Message: " - type - VARIABLE n, VALUE NUMBER p EXCEEDS MAXIMUM ALLOWED (q)", where - type - is one of the types listed above, n is the variable number, p is the value provided and q is the maximum value or code allowed for that variable. Maximum values are listed in Table 1, Chapter 5 for all Bank variables, and correction should be made.

6. Check 5 : Statistics parameters

Statistics parameters are very simple. Any non-zero value up to the number specified as the Item data parameter may be specified for either variable, (although the correlation of nominal variables leaves an interpretation problem).

6.1 Message: "FIRST (or SECOND) VARIABLE VALUE FOR CORRELATION EXCEEDS SPECIFICATIONS"

This message directly relates to the upper limit imposed by the Item data parameter.

6.2 Message: "VARIABLES IN CORRELATION ARE EQUAL. THIS IS NOT ALLOWED".

Here, either the variables are equal (resulting in 1.00 as the coefficient and therefore a worthless option) or the data has been entered in the wrong columns.

7. Check 6 : BANKLIST Parameter

Message: "BANK LISTINGS PARAMETER EXCEEDS SPECIFICATION"

The maximum override value for this option is 1∅; this message indicates a value provided in excess of this.

8. Check 7 : BANKLIST Values

8.1 Message: "BANK LIST VARIABLE: n, EXCEEDS SPECIFICATION"

This message indicates that the variable number provided is in excess of that given as the Item data parameter

8.2 Message: "BANKLIST VARIABLE n, VALUE p, EXCEEDS MAXIMUM ALLOWED (q)"

This message indicates that the value p on variable n is in excess of that variable's maximum value or code (Chapter 5, Table 1)

9. <u>Check 8 : UPDATE Parameters</u>

Most of the UPDATE option is completely automatic despite its complexity. Data is drawn largely from disc storage and the possibility for error at System data input is smaller.

9.1 <u>Message</u>: "NUMBER OF TESTEES EXCEEDS SPECIFICATIONS"
There is an upper limit of 250 due to main System storage limits under MUSIC. This message indicates the provision of a parameter in excess of 250.

9.2 <u>Message</u>: "NUMBER OF ITEMS IN TEST EXCEEDS SPECIFICATIONS"
As in 9.1 above, the upper limit for items is 35 items.

9.3 <u>Message</u>: "- name - IS NOT AN ALLOWABLE OPTION PARAMETER", where - name - is the keyword provided. The alphameric keyword provided is not one of the twelve (ONE to TWELVE or ALL), either through mistyping or other causes.

9.4 <u>Message</u>: "- name - IS NOT AN ALLOWABLE OPTION NAME", where - name - is the keyword provided. The alphameric keyword provided is not one of the thirteen option names. allowed.

9.5 <u>Message</u>: "INITIAL RECORD NUMBER FOR UDS 2 LEAVES INSUFFICIENT SPACE FOR CALCULATIONS (n)", where n is the number of records required to hold the specified amount of data.

Here, the UDS 2 initial record value is such that ((N of subjects x no. of items) + (no. of items x 15) + N of subjects) is so large that it requires more room than (12000 - initial record number).

To clarify, records are required for the following reasons:
i) to hold each subject's data (subjects x items)
ii) to hold response patterns (estimated 15 per item)
iii) to hold subject names.

If, therefore, the UDS initial record number means that from that value to the maximum (12000 records) is <u>less</u> than the number of records needed to hold all the data in i) to iii) above, then some mistake must have been made, either in the UDS parameter, the number of subjects or the number of items.

10. <u>Check 9 : Consistency of Update Options</u>

10.1 <u>Message</u>: "DISCREPANCIES (or KAPPA) CANNOT BE CALCULATED. RETEST OPTION HAS NOT BEEN INVOKED" or

Message: "QUALITY (or RESPONSE) CANNOT BE INVOKED. ITEMDATA OPTION HAS NOT BEEN INVOKED"

These messages are self-explanatory in that certain sub-options require the data provided by other sub-options.

10.2 Message: "NUMBER OF TESTEES FOR RETEST ANALYSIS EXCEEDS SPECIFICATIONS"

No more than 50 subjects are allowed for a test-retest analysis.

10.3 Message: "ASSOCIATED VARIABLE FOR RETEST DATA OVERLAPS TEST DATA AREA"

In the use of the RETEST suboption, the starting position of the retest data must be specified. If this is less than the number of records calculated for the test data (9.5 above) then it is rejected as overlapping the storage of test data.

10.4 Message: "RETEST DATA POSITION SPECIFIED BEYOND SIZE OF UDS"

Here, the retest data has been signalled to start beyond the 12000th record ie. the maximum record number allowable.

11. Check 10 : CROZTABS Variables

Message: "- name - IS NOT AN ALLOWABLE VARIABLE NAME"

Again a self-explanatory message, where - name - is a Bank variable label mistyped or incorrectly provided.

12. User Data Set Error

Message: "ERROR ENCOUNTERED ON USER DATA SETS. DATA WRONGLY ORDERED. THIS OPTION TERMINATES SUBJECT NUMBER n TEST ITEM p RETEST ITEM q"

To avoid unreliable data output, all calculations are made but not used (or the Bank updated) until the entire data input is complete. As mismatch occurs between the test data set and the UDS retest data set the above message is output and the option aborted. A mismatch occurs when a subject's data is misplaced in the data set, or item numbers are mispunched or cards misplaced. In the message above, n is the subject number, p and q are the item numbers (provided) from each set respectively. If these latter are equal, then the subject numbers do not coincide, often indicating a whole data set for a pupil is out of place.

Appendix:II IBS Codebook

All codes and variables stored by the Item Bank are numerical and it will be useful for the User to be able to interrogate a terminal to discover the descriptions of the various values. Such a device is the Codebook. A UDS (see Chapter 13) holds a descriptor for every value of every Bank variable. The program is invoked by the command

/EXEC CDBKIB

and Figure 33 below is an example of the system.

```
DO YOU WISH FOR A LISTING OF THE WHOLE CODEBOOK OR JUST
A FEW CODES ?

PLEASE TYPE "BOOK" FOR THE FORMER OR "CODES" FOR THE LATTER
?
codes

PLEASE TYPE IN THE NUMBER (BETWEEN 1 AND 30) OF THE BANK VARIABLE
FOR WHICH YOU REQUIRE DETAILS,AS A 2-DIGIT NUMBER
?
12

PLEASE TYPE IN THE VALUE OR CODE FOR WHICH YOU REQUIRE DETAILS
AS A 4-DIGIT NUMBER
?
  11

                    VARIABLE NOQ  ,NUMBER 12
                    ************************
       CODE OR VALUE                 DESCRIPTOR
       *************                 **********
              11        THIS IS THE QUESTION NUMBER IN LAST TEST

DO YOU WISH FOR DETAILS OF OTHER CODES ?
?
yes

PLEASE TYPE IN THE NUMBER (BETWEEN 1 AND 30) OF THE BANK VARIABLE
FOR WHICH YOU REQUIRE DETAILS,AS A 2-DIGIT NUMBER
?
 3

PLEASE TYPE IN THE VALUE OR CODE FOR WHICH YOU REQUIRE DETAILS
AS A 4-DIGIT NUMBER
?
  3

                    VARIABLE CONT ,NUMBER  3
                    ************************
       CODE OR VALUE                 DESCRIPTOR
       *************                 **********
              3         IDENTIFICATION & PHYSICAL DESCRIPTION
```

APPENDIX III

Special Forms

Form IB 1: coding form for new items

A specimen IB 1 is given overleaf. Data should be coded as given in Table A III.1 below:

Column(s)	Datum	Maximum size
1 - 3 inc	Identification number	25Ø
4 + 5	Industry code	38
6	Content code	6
7	Job code	8
8	Format	7
9 - 11 inc	Firm/organisation code	999
12	Size of Firm	6
13	Question type	3
14 - 17 inc	Key	9000
18	Order of item on page	4
19 + 20	Test code	99
21 + 22	Question number in last test	99
23 - 26 inc	Number taken item	9999
27 + 28	Number of Answer Boxes	10
29 - 31 inc	Number omitting item	999
32 - 34 inc	Number selecting Box 1	"
35 - 37 inc	" " " 2	"
38 - 40 inc	" " " 3	"
41 - 43 inc	" " " 4	"
44 - 46 inc	" " " 5	"
47 - 49 inc	" " " 6	"
50 - 52 inc	" " " 7	"
53 - 55 inc	" " " 8	"
56 - 58 inc	" " " 9	"
59 - 61 inc	" " " 10	"
62 - 64 inc	Number correctly responding	"
65	Linguistic task	4
66 - 68 inc	Number replacing this	250
69 - 71 inc	Discrepancies	999

Forms CS 1 and CS 1a: Pupil response Control Cards

The two versions of this form apply to pupil-responses codes on forms CS 2 and CS 3 (see below). With Form CS 1, the form is the first sheet and precedes the data sheets. Its function is to create a Music Save File with all control cards and data ready for execution. It does not execute the job or place data on a UDS.

Coding is as follows, referring to the form overleaf:

line 1 : /ID is followed by the User's surname, beginning in column 5. nnnn nnn is the User code.

line 2 : aaaaaa is a six-character name for the Save File and it is this name that will be used to execute the job. The name must begin with an alphabetic letter and may not contain punctuation or special characters.

line 4 : bbb is the number of subjects. It is right justified (ie if less than 100, column 21 is left blank, if less than 10, columns 21 and 22 are left blank).

line 5 : cc is the number of items in the test - right justified.

line 6 : dddddd is either TEST or RETEST, left justified.

line 7 : eeeeee is the same name as aaaaaa.

Thereafter, CS 2 is completed per pupil until the last, when CS 3 is completed. These forms are then given in to Computer Services Department for punching. To execute the Save File aaaaaa and so place the data on a UDS, the file may be submitted at a terminal using the SUBMIT routine.

Form CS 1a gives the relevant control cards for a Save File of data eeeeee, which may not contain control cards from CS 1. Form CS 1a may be typed into a terminal and saved for submission.

Column	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	/	I	D																		n	n	n	n		n	n	n		
	/	S	A	V	E		a	a	a	a	a	a																		
	/	I	N	C	L	U	D	E		M	A	R	K	E	R															
	N	O	.		O	F		S	U	B	J	E	C	T	S						b	b	b							
	I	T	E	M	S		P	E	R		T	E	S	T							c	c								
	S	T	A	T	U	S															d	d	d	d	d	d				
	S	A	V	E		F	I	L	E		N	A	M	E							e	e	e	e	e	e				
				d	a	t	a																							
					o	n																								
				F	o	r	m	(s)																						
				C	S		2																							
		[		F	o	r	m		C	S		3		]																

Column	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	/	I	D																		a	a	a	a		a	a	a		
	/	I	N	C	L	U	D	E		M	A	R	K	E	R															
	N	O	.		O	F		S	U	B	J	E	C	T	S						b	b	b							
	I	T	E	M	S		P	E	R		T	E	S	T							c	c								
	S	T	A	T	U	S															d	d	d	d	d	d				
	S	A	V	E		F	I	L	E		N	A	M	E							e	e	e	e	e	e				
	/	I	N	C	L	U	D	E		e	e	e	e	e	e															
						o	t	h	e	r																				
						d	a	t	a																					
						f	i	l	e	s																				
	/	E	N	D																										

Form CS 2: Pupil Responses

The example given overleaf should be self-explanatory. A master-form should be completed and then reproduced to avoid continual coding. In the example, the Pupil name may be up to 20 characters in length and contain any characters. 'nnnn' is a four digit pupil identification number - it may be left blank if desired, but the IBS marking system will output a zero. On each of the following lines, two items may be coded, where iii is the three digit item number. The next four columns are the first four boxes marked and then any other markings. The second item per line starts at column 14. For completion items scored by the tester, code Ø in the first column after the item number for correct, 8888 in the first four after the item number if incorrect. For all items, code 9999 for omission. For ease of marking or any other reason, iii may be given as ØØØ, in which case it will be ignored by the system.

Form CS 3: Last Pupil Response

This form differs solely by the addition of /END on the last line, signalling the end of data.

Column	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	←					P	U	P	I	L		N	A	M	E					→	/	/	n	n	n	n				
	i	i	i											i	i	i														

APPENDIX V

MANUAL OF THE VOCABULARY BANKING SYSTEM

SHEFFIELD OCCUPATIONAL FUNCTIONAL READING PROJECT

VBS

Manual of the Vocabulary Banking System

Owen Parry

1980

THE SHEFFIELD OCCUPATIONAL FUNCTIONAL READING PROJECT

Introduction

In the Green Paper, 'Education in Schools - a Consultative Document' (Cmnd 6689, 1977) it was argued that 'schools must prepare their pupils for the transition to adult life and work'. If such preparation is accepted as one of the duties of the school, it seems reasonable to expect secondary schools to prepare their pupils, as far as may be possible. for the functional reading (i.e. reading associated with a role or task) which will be required by them when they leave school.

Before schools can prepare their pupils, however, it is necessary for the relevant reading tasks to be identified and means devised for ascertaining the degree to which pupils can cope with the tasks so identified. To identify these tasks and provide means of assessing pupils' attainments in relation to them was the purpose of a research project undertaken by Sheffield City Polytechnic.

Project objectives

In September 1977, Sheffield City Polytechnic, in association with Sheffield Metropolitan District Council, initiated the Sheffield Occupational Functional Reading Project (SOFRP), with the following objectives:

i) The identification and classification of reading tasks, associated with job seeking and employment which a 16 year old could face immediately he/she left school.

ii) The construction and validation of a criterion referenced group test, based on the reading tasks identified in achieving objective (i) above.

iii) The construction of diagnostic tests to be used with individuals who fail to reach the criterion scores on the group test.

(A fourth objective, subject to discussion with Sheffield Schools, was the production of materials to be used in teaching related to occupational functional reading).

The Progress of the Project

In order to identify the reading tasks which might be faced by 16 year old school leavers a large number of firms in Sheffield were visited. The firms were chosen at random from groups of firms representing the main employment areas entered by 16+ school leavers. Each firm gave details of the types of jobs into which leavers were recruited, details of training (where applicable) and induction procedures. Job related reading materials were inspected and collected where possible. Relevant job-seeking materials were obtained from the Careers Service and, where their courses made up the substantial reading requirement of a type of job, from the Further Education colleges.

The reading materials obtained were used to construct test items (an item consisting of a passage to be read and one question related to it). A passage might be purely prose plus a diagram or illustration, a form to be filled in, etc.

Each item was scrtuinized by a panel of employers' representatives, to ensure that it was representative of job related reading tasks encountered by 16+ school leavers, and also by linguists to ensure that there were no undue complexities in the questions. Items approved by employers' representatives and linguists were then used to construct a possible test. In order to see whether pupils would find any items ambiguous and to try out a proposed pattern of test administration, the test was given to pupils in two schools in Sheffield. Following this, certain items were changed or replaced and, where necessary, re-piloted.

The date associated with items surviving these processes of scrutiny and piloting were banked in computer storage. The access to and use of that data is the subject of a Manual allied to this (Parry, 1980).

It was felt important in the course of the research to examine the technical vocabulary used in job-related reading materials. Such terms might be the 'in-house' or industry-specific jargon that appears in various documents; the technically specific words or phrases relating to some object or process; or a number of familiar, everyday words used in such a way as to give them specific technical meanings. All of these will have an effect upon the complexity of reading materials and hence the difficulty of the reading task associated with it.

For the purposes of the future work of the Project, it was proposed to identify and store the technical vocabulary contained in the reading passages of items being used in the Project. The identification of the vocabulary was undertaken by members of the Department of English and of Communication Studies, to whom many thanks are due. It was felt that members of these departments were sufficiently sophisticated in their knowledge of language not to include unfamiliar words merely because they were unfamiliar (but not necessarily technical), but also that they were sufficiently naive about the context of technical usage to recognise each occurrence. That is, persons who use the vocabulary everyday are less likely to realise what is technical and what is not technical; unsophisticated users of language may be overinclusive in their selection. Members of the two departments fall between these two stools and are therefore the appropriate judges.

In all, many hundreds of terms weee identified and were stored along with associated classification data. This Manual describes the processes by which searches can be made of that store for vocabulary items of specific or general types.

It is recommended that this Manual is used in conjunction with its associated Manual (SOFRP IBS), as much of the classification data is contained therein and not reproduced here. For clarity, figures and tables in that Manual will be called IBS figure ... or IBS Table ... and in this Manual, VBS ..., etc.

Data Stored

Each vocabulary item consists of four pieces of numerical data and a string of alphameric characters up to 4o characters long. The numerical values are:

i) the code of the industry from which the original passage was derived;

ii) the code of the job in which it was used;

iii) the code of the content-type (an American system adopted by Sticht (1975))

iv) the number of the item in the IBS Item Bank.

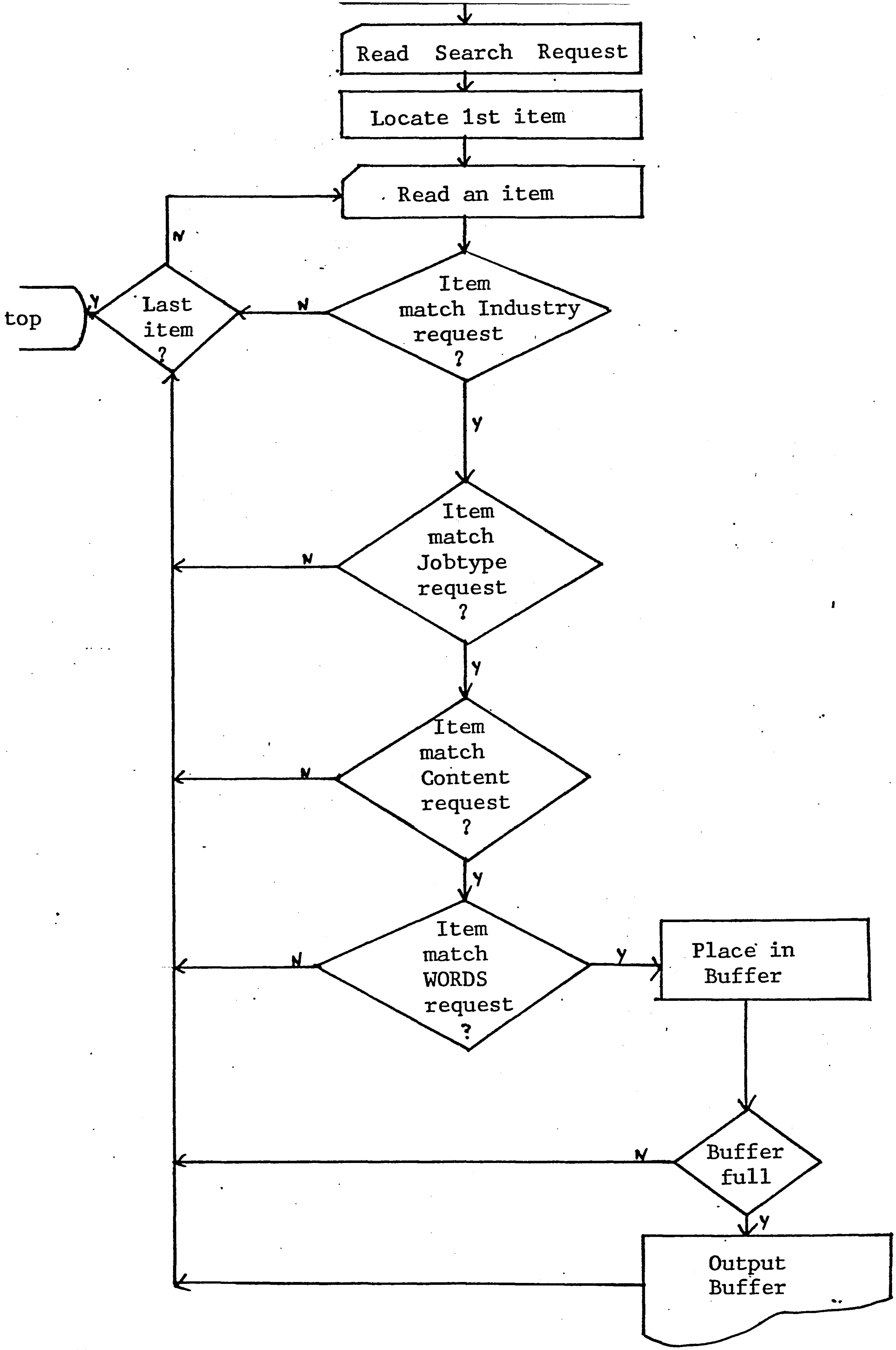
The alphameric string is the actual word or phrase banked.

The codes given above are those used in the IBS and the Codebook may be assessed via the command /EXEC CDBKIB at a terminal.

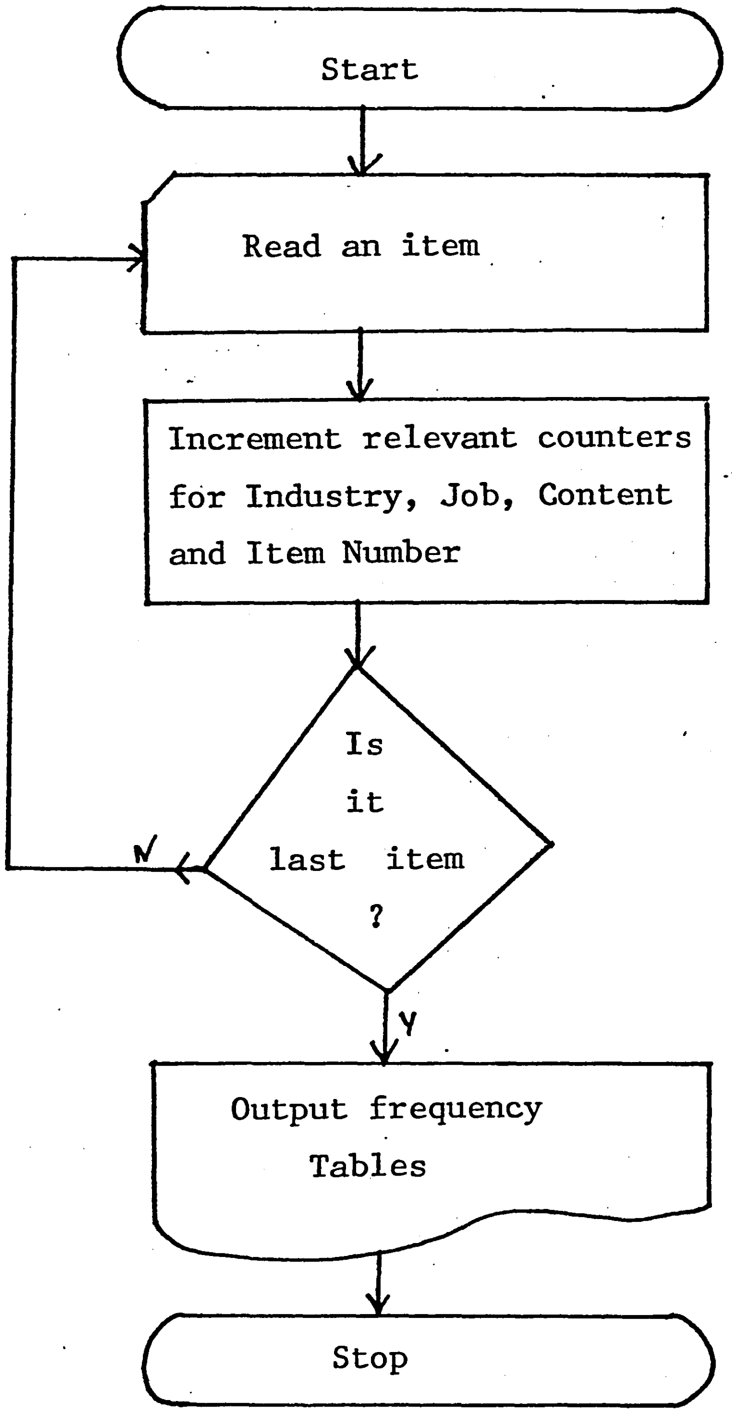
The data are sorted according to their numerical values, lowest first.

Control of the System

The VBS control system is very simple. The User provides a set of search requests which are then used by a searching program to access the VBS store. Conforming vocabulary items are output with their associated data, in sets of ten items. Similarly, use of the frequencies routine means that the command involves a system where the entire VBS is scanned and counts made of the number of vocabulary items with each data code. The lists of frequencies of each code are then output.
VBS Figure 1 gives a flowchart of the search routine.
VBS Figure 2 gives a flowchart of the frequencies routine.



VBS Figure 1: Flowchart of Search Routine



VBS Figure 2: Flowchart of FREQs routine

Provision of Search or Frequencies Requests

The provision of requests is handled interactively at a terminal and the User should type /EXEC DATAVO

This invokes a simple program which invites the User to enter a data-type name and then a numerical value, or word or phrase.

Acceptable names are : INDUSTRY
JOBTYPE
CONTENT
FREQS
WORDS

INDUSTRY is related to the IBS meanings of industrial types and the <u>same</u> codes are used

Similarly, JOBTYPE and CONTENT have the same codes as variables JOB and CONT in the IBS.

Numerical values are specified in association with these three names and are used in the search through the VBS store. In essence what happens is thus: given values of, say, INDUSTRY 2, JOBTYPE 1 and CONTENT 3, the search program accesses first a management file which quickly locates the start of vocabulary items with those characteristics and counts how many there are. If these are none, a suitable message is output and the job ends. Otherwise, each item is placed in an output buffer and when ten have been found or no more are available, the buffer is output.

If one of the names is omitted or is given as zero, then the search is undertaken for <u>each and every</u> code associated with that name. For example, for INDUSTRY 2, JOBTYPE 1 and CONTENT 0, the search is undertaken for INDUSTRY 2, JOBTYPE 1, CONTENT 1; then INDUSTRY 2, JOBTYPE 1, CONTENT 2; and so on. Similarly, for the omission of INDUSTRY, each code from 1 to 38 would be used in a search.

The name WORDS adds to this search. The word or phrase specified after the WORDS name is used in searching the alphameric string making up part of the vocabulary item's data. Any occurrence of the word or phrase within the string (not necessarily starting at the beginning, but it must be continuous) will advance the item for selection. (To be actually selected, its numerical data must also conform to the values, if specified, for INDUSTRY, etc). For instance, for INDUSTRY 2, JOBTYPE 1, CONTENT 3 and WORDS FOREMAN, the following item would be selected:
INDUSTRY 2 JOBTYPE 1 CONTENT 3 BY THE MANAGER OR FOREMAN
whilst the following would not
INDUSTRY 3 JOBTYPE 1 CONTENT 3 BY THE FOREMAN
INDUSTRY 2 JOBTYPE 1 CONTENT 3 THE MAN WHO USED HIS FOREARM

FREQS may only be used on its own and it will cancel any search associated with it. The use of the name causes the entire VBS store to be scanned and the number of items conforming to each code for each industry, job, content and IBS item number are counted. This is output as a list by industry, by jobtype, by content and by IBS item number. This enables the User to keep an up-to-date list of what sort of vocabulary items are available to him and avoids fruitless searches.

VBS Figure 3 gives an example of a typical search request

VBS Figure 4 gives an example of output from a search

VBS Figure 5 gives an example of output from FREQS

```
MAIN    = ØØØ6A6
ØØ6878 BYTES USED
EXECUTION BEGINS      5.9S

SUBMISSION PROGRAM FOR VOCABULARY BANK

PLEASE ENTER A DATA TYPE NAME
?
INDUSTRY

AND THE VALUE OR WORD FOR THIS DATA TYPE
?
2

DO YOU WISH TO SPECIFY FURTHER TYPES ?
?
YES

PLEASE ENTER A DATA TYPE NAME
?
JOBTYPE

AND THE VALUE OR WORD FOR THIS DATA TYPE
?
1

DO YOU WISH TO SPECIFY FURTHER TYPES ?
?
YES

PLEASE ENTER A DATA TYPE NAME
?
CONTENT

AND THE VALUE OR WORD FOR THIS DATA TYPE
?
3

DO YOU WISH TO SPECIFY FURTHER TYPES ?
?
YES

PLEASE ENTER A DATA TYPE NAME
?
WORDS

AND THE VALUE OR WORD FOR THIS DATA TYPE
?
FOREMAN

DO YOU WISH TO SPECIFY FURTHER TYPES ?
?
NO

DO YOU WISH THIS JOB TO BE SUBMITTED FOR BATCH PROCESSING ?
?
YES
JOB NUMBER R432 SUBMITTED OK

JOB COMPLETED.
```

VBS Figure 3: Search request

INDUSTRY CODE	JOBTYPE CODE	CONTENT CODE	ITEM NUMBER	WORD OR PHRASE
4	5	2	137	SAFE HANDLING
4	5	2	137	DIES
4	5	2	137	TOOLS
4	5	2	137	PRESS
4	5	2	137	POWER PRESS REGULATIONS
4	5	2	137	TOOLING
4	5	2	137	PRESS OPERATION
4	5	2	137	COMPONENTS
4	5	2	137	DESIRED QUALITY STANDARDS
4	5	2	137	FINAL CHECK

INDUSTRY CODE	JOBTYPE CODE	CONTENT CODE	ITEM NUMBER	WORD OR PHRASE
4	5	2	137	COMMENCE PRODUCTION
4	5	2	137	INSPECTS
4	5	2	137	APPROPRIATE ACTION
4	5	2	137	ROUTINE MAINTENANCE PROCEDURES
4	5	2	137	DAILY LUBRICATION
4	5	2	137	GOOD HOUSEKEEPING
4	5	2	137	PRESS CLEANLINESS
4	5	2	137	STACKING OF WORK
4	5	2	137	WORKING AREA
4	5	2	137	LIAISES

INDUSTRY CODE	JOBTYPE CODE	CONTENT CODE	ITEM NUMBER	WORD OR PHRASE
4	5	2	137	JOB DESCRIPTION
4	5	2	137	RESPONSIBILITIES
4	5	2	137	HOT OR COLD
4	5	2	137	PROTECTIVE CLOTHING
4	5	2	137	CORRECT MANNER
4	5	2	137	OPERATOR
4	5	2	137	PERIODICALLY
4	5	2	137	MATERIAL
4	5	2	137	REQUIRED STANDARD
4	5	2	137	FOREMAN

VBS Figure 4 : Output from Vocabulary Banking System

VOCABULARY ITEMS AVAILABLE, BY INDUSTRY

S.I.C.CODE	NUMBER OF ITEMS
1	0
2	132
3	52
4	31
5	22
6	94
7	0
8	0
9	0
10	0
11	55
12	0
13	0
14	1
15	10
16	49
17	0
18	38
19	30
20	23
21	21
22	0
23	0
24	0
25	0
26	0
27	0
28	0
29	0
30	0
31	0
32	0
33	0
34	0
35	0
36	105
37	0
38	30

VOCABULARY ITEMS AVAILABLE,BY JOBTYPE

CODE OR VALUE	NUMBER OF ITEMS
1	300
2	0
3	112
4	61
5	81
6	137
7	2
8	1

VBS Figure 5: Output from FREQS

Entering new data

New vocabulary items may be coded on VBS Form 1, with the Industry code in columns 1 and 2, with a leading blank or zero if the code is less than 10; Jobtype and Content are coded in columns 3 and 4 respectively; and Item Number is coded in cloumns 5 to 7 inclusive. The vocabulary string itself is coded in columns 8 to 47 inclusive. Strings with less than 40 characters should be followed with a colon (:) in the next column, as this serves as an end-of-string marker.

VBS Figure 6 shows an example of some new coding:

VBS. CODING SHEET FOR NEW VOCABULARY ITEMS

Ind	J	C	Item		Vocabulary item (Max. 40 characters)
Ø3	2	3	1Ø4		ENGINEERING TOOLS:
Ø9	3	4	Ø99		DIE-PRESSING AND CUTLERY MANUFACTURE:
12	6	1	211		CASH REGISTER:

VBS Figure 6: Coding of new Vocabulary Items

The User should submit a job in batch (see IBS Manual, p 5) with the following cards:

/ID etc

/INCLUDE CRVOBN

then the VBS Coding sheets

and

/END

These should be given in at Job Reception, Computer Services.

d	J	C	Item		Vocabulary item (Max. 40 characters)

APPENDIX VI

PROVISIONAL MANUAL FOR THE FUNCTIONAL

READING TEST, FORM A

Sheffield City Polytechnic

Language Development Centre

FUNCTIONAL READING TEST A

PROVISIONAL MANUAL of INSTRUCTIONS

W. Latham & O.Parry

August, 1980

Introduction

In the Green Paper, 'Education in Schools - A Consultative Document' (Cmnd 6869, 1977) it was argued that 'schools must prepare their pupils for the transition to adult life and work'. If such preparation is accepted as one of the duties of the school, it seems reasonable to expect secondary schools to prepare their pupils, as far as may be possible, for the functional reading (ie reading associated with a task or role) which will be required of them when they first leave school. For instance, the pupil leaving school at 16+ should be prepared for functional reading related to his/her role as a job seeker or worker (and, one might add, as consumer and citizen).

Unfortunately, it is not reasonable to expect the secondary school to prepare pupils for functional reading related to any of the roles mentioned above, as the reading tasks involved have yet to be identified and classified; and, obviously, no tests exist to assess the pupil's progress towards success in functional reading or diagnose his/her difficulties.

In September 1977, Sheffield City Polytechnic, in association with Sheffield Metropolitan District Council, initiated a research programme aimed at providing Sheffield schools with both information, and assessment and diagnostic tests, related to functional reading in the areas of job seeking and employment.

This test arises out of that research project.

Criterion - referenced Measurement

Criterion-referenced measurement is different from the norm-referenced measurement that most testers are used to, in that the latter measures the performance of an individual with reference to all others taking the test. The individual is at, above or below the norm for those taking the test. In criterion-referenced measurement, however, the reference point is some external measure (the criterion) and it is this absolute standard with which we are concerned. The criterion in this work is the performance of individuals on job-related reading tasks, and they are assessed in relation to actual job performance. To make this clearer, consider a driving test: no matter what anyone else can do, <u>you</u> still have to satisfy the examiner that you can drive. Similarly, the school-leaver must satisfy the tester, via the test, that <u>he</u> can cope with job-related reading tasks, not that he is as good, or better, than anyone else.

Test Format

The Functional Reading Test, Form A, consists of 31 questions based on passages. It is arranged in two sections. The first has seven questions of different types, where the testee has to either complete the whole set of answers in a particular way (Q7) or to select up to four correct answers from up to ten possibilities. These questions were derived from reading tasks where a young person has to know, for instance, <u>all</u> the safety rules for a job, not just one. The second section has 24 questions, three per passage, and are straight forward "5-possibilities, one correct answer" multiple-choice questions.

There are 13 subtests, as well as the total score, which may be used to indicate areas of particular weakness or strength.

Uses of the Functional Reading Test, Form A

The test passages are derived from materials actually used by school-leavers in their first few weeks at work (see below). It is felt that after that time, the young person becomes sufficiently socialized into his task, and is building upon what he has learned, so that any test given prior to starting employment would be irrelevant to any later period than the beginning weeks.

The test is primarily designed, therefore, for use in the first term of the last year at secondary schools, either by subject-specific departments (eg English) or by those responsible for careers guidance. It might be given in the last term of the penultimate year, but any earlier would not really be giving the young person a chance to develop his functional skills to a level related to his level of skill at leaving-time. It is suggested that the first term of the ultimate year be used, as this allows time for intervention and, where possible, remediation.

In careers guidance, the job subtests (see below) allow a tester to see areas of weakness related to type of job. He is able, therefore, to act appropriately, either in terms of guidance or in arranging help.

The test is not designed for use with pupils with general, severe reading difficulties. It is felt that such pupils would regard a test of this complexity as upsetting, being beyond their abilities. At the other end of the range, however, the very brightest may have shortcomings in functional reading skills not demonstrated by ordinary school tasks. Conversely, the middle to low ability pupil may score very highly.

This test has not been validated beyond the initial employment period (see below). There is no evidence, therefore, that it predicts long-term adequate job-performance. The use of the instrument for personnel selection, therefore, is to be regarded with some reservations.

Collection of Job-related Reading Materials

The Project proceeded with the identification and classification of job-related reading tasks for this group of young people. A representative sample of firms in Sheffield was visited, selected by a random process (Latham and Parry (1979)) to give a broad range of industries and types of job. Details of numbers of leavers employed, types of job, training procedures, etc were collected, along with specimens of job-related reading materials. Industry Training Boards, Group Training Associations and Colleges of Further Education were similarly visited, where relevant. Materials were also collected from the local Careers Service.

A large volume of material was thus collected, of many varying types. Reading materials included job cards, stock lists, safety regulations, descriptions of machinery, leaflets on security, personal appearance, credit card completion, and many other types. This plethora of disparate material becomes more comprehensible, however, if one groups similar types of job together. Six job-types have been used, corresponding, in the main, to those categories commonly employed by the Careers Services. These are given in Table 1.

Table 1: Definitions of job-types

1	Apprenticeship:	in which articles of apprenticeship are signed and agreed, with national regulations, block and/or day release training;
2	Professional:	recognised training in one of the professions, eg articled clerk in accountancy;
3	Clerical:	general office and clerical duties which may or may not include any training;
4	Distribution Operatives/Others:	those employed in the Distribution industry; operatives: 2+ months of on-the-job training; others: no formal training;
5	Operatives/Others:	as above, all other industries;
6	Induction:	applies to reading material common to all job-types, given as part of induction into the company.

Content - categories

All materials used in the construction of items were further classified according to a system adopted from the work of Sticht and his associates (Sticht et al,1972). This classification is reproduced in Table 2.

Table 2: Definition of content-type categories

1 Tables of content and indexes:

Content designating the location of information with a publication.

2 Standards and specifications:

Content setting forth specific rules or tolerances to which task procedures or the completed product must conform.

3 Identification and physical description:

Content attempting to symbolically represent an object via an identifying code (stock ##, nomenclature) and/or by itemizing its distinguishing physical attributes.

4 Procedural directions:

Content which presents a step-by-step description of how to carry out a specific job activity. Essential elements are equipment/materials/ingredients to be used, and how they are to be used, with presentation organized in a sequential step-wise fashion.

5 Procedural check points:

Content which presents a key word or highly summarized version of what should be done in carrying out a task rather than how it should be done. This content differs from the content classified under

/continued overleaf

5 Procedural check points: continued

Procedural Directions in that it assumes the user knows how to carry out the steps once reminded that the step exists and/or reminded of the decision factors which determine whether the step is required.

6 Functional description:

Content which presents an operating (cause and effect, dependency relationships) description of some existing physical system or subsystem, or an existing administrative system or subsystem.

(By kind permission: Human Resources Research Organisation)

A further classification system, based on linguistic tasks was also used. This is given in Table 3.

Table 3 Definition of linguistic tasks

1	Referential:	where the item poses a task requiring the locating of correct information;
2	Regulative:	where the item requires an answer in terms of a behaviour, either an act of commission or omission;
3	Attitudinal:	where the question requires the description of a state of mind, or approach to a task;
4	Definitive:	where the item requires a description of the meaning of a word or words.

Piloting of Items

All items surviving the content validation procedures (see below) were piloted amongst the target population. Over one hundred middle-ability Fifth Form pupils in a Sheffield comprehensive school took various sets of items in short tests. No group was smaller than seven nor larger than thirty. This initial pilot was to discover any obvious ambiguities in text or questions, to note frequently omitted items and those where a wrong answer was chosen more often than the correct answer. (See below for item analysis details). Test administration procedures were also examined at this time.

After scrutiny, certain items were rejected or amended, these latter being repiloted.

Construction of Tests

Using a computerised item bank, piloted items were used to construct a test. The pilot study had suggested that the maximum number of items that could be answered in a double school period was about thirty. To construct a test wherein every content-type and every linguistic task was represented for each job-type would have meant the number of items necessarily exceeding thirty. Instead, a test (Form A) was constructed, with items for each job type, content-type and linguistic task. No job-type, however, had items for every content-type or linguistic task. Similarly, no content-type had items for every job.

Figure 1 indicates the format of Form A.

Figure 1 Number of items by job and content type: Form A

	Apprentice	Clerical	Distribution	Op/other	Induction	Totals
Standards & specifications	1			1	3	5
Identification & Physical description	4	5		1		10
Procedural Directions	3		3			6
Procedural Checkpoints	3		1			4
Functional Description	3				3	6
Totals	14	5	4	2	6	31

Test Reliability

A criterion-referenced test is one which attempts to assess an individual's ability or performance in relation to some external criterion (eg whether or not a person can drive a car is assessed at a driving test). The reliability of such a test can therefore be seen as the reliability of classification as 'master' or 'non-master' on repeated testing. The test-retest procedure is indicated as the method of determining reliability for this type of test.

The calculation of measures of reliability remains a controversial area within criterion-referenced measurement. Various non-standard statistical techniques have been suggested - and decried - (see eg Livingstone (1972) Harris (1972)). It was decided to use standard, product-moment correlation techniques and the decision-theoretic approach adopted by other workers in this area (Swaminathan et al 1974); Close (1977); Berk and DeGangi (1979); Subkoviak (1978)), using Cohen's K (Cohen 1960) as an unbiased estimate of the degree of agreement in classification as master/non-master. The coefficient was calculated for each possible score, and its maximum value taken as the score at which decisions were most reliable. This score will be used as part of the predictive validity study. (See below).

The results for Forms A are given in Table 4, for repeated administration to 47 middle ability students in two schools.

Table 4: Reliability Coefficient

	Pearson Correlation	Cohen's K	Optimal Cutting Score	Total number of items
Form A	0.87	0.81	17	31

The Functional Reading Test, Form A, also correlates with the Edinburgh Reading Test, Form 4, with $r = 0.81$ ($N = 375$).

Internal consistency measures suffer from the restrictions given above, but the correlation table below is reported for the sake of completeness.

Table 5

Product-moment correlations of subtest totals with whole test total score:

Subtest	r
JOB 1	0.93
JOB 3	0.65
JOB 4	0.63
JOB 5	0.56
JOB 6	0.71
CONT 2	0.73
CONT 3	0.86
CONT 4	0.79
CONT 5	0.72
CONT 6	0.67
LING 1	0.92
LING 2	0.92
LING 4	0.50

$N = 148$

$p < 0.001$ in all cases

Item Analysis

Following on from this brief discussion of problems in criterion-referenced tests, it is clear that many of the other statistics used in classical item analysis are also of limited value for the same reason: problems arising from restricted variability as between scores.

The commonest measures, difficulty and discrimination, are, in fact, of some use in criterion-referenced measurement, but not in the same sense as in classical test theory. Item difficulty is interesting but not of overwhelming importance. A test constructor may require a mixture of easy and hard items, but more often in criterion-referenced measurement he requires a test of a certain area. If the test has content validity, the difficulty of the items is secondary. Discrimination indices for an item are not valuable in terms of seeing how the item contributes to spreading out scores on the test, but to see whether an item positively or negatively discriminates. The latter is as much an indicator or a poor item in criterion-referenced measurement as in norm-referenced measurement (eg Smith, 1974).

Difficulties in the test range from 0.17 to 0.97 (expressed as the proportion correctly answering the item).

There were no negative discriminators.

Validation

Two types of validity are relevant to this test: content validity and predictive validity. The former is a measure of whether the test contains material related to, and representative of, the domain which the test is supposed to represent. The latter is a measure of the extent to which a high score does well on the actual task, at work, and vice versa.

Every item was subject to scrutiny by panels of personnel and training officers from industry and commerce to assess its content validity. Different panels were invited, according to the type of job predominant in that industry, and asked to assess each item in terms of the representative nature of the reading passage and the relevance of the question. If the reading material failed to be representative of a type commonly encountered, the item was rejected. Similarly, should the question-part of the item be rare, irrelevant or not applicable to school-leavers in their initial period, it was either modified or rejected.

All surviving items were then examined by two linguists from the staff of the Polytechnic. Their task was to note any unseen ambiguities or unnecessary complexities and to check the content classification of each item.

The ultimate value of a test of functional skill is its ability to predict adequate performances. In this case, the test of occupational functional reading ability should predict performance on job-related reading tasks in the first six weeks of employment or training for a job.

As a measure of this predictive validity, about 500 fifth-form school-leavers, mainly of average ability, were given Form A of the test, and the Edinburgh Reading Test, Form 4. This latter is a norm-referenced test for comparison purposes; extrapolated quotients were used for those pupils already 16 years old.

These pupils left school and found or sought employment. The Sheffield Careers Service kindly allowed access to its files, so that each subject might be followed up when in employment. Each employer was asked to allow the young person's immediate supervisor to fill in a rating form, assessing relevant aspects of job-performance on tasks involving and not involving reading. These include perseverance, need for oral guidance, speed, etc.

Test scores and ratings were then related via multiple and linear regression analyses, to establish the test's predictive validity.

Due to local labour market factor and a very low return rate of the rating forms, the predictive validity of the test could not be demonstrated.

Test Interpretation

As the test has content validity, its main use is to demonstrate the range of tasks that a testee can do, and those he cannot. The content subtests and the linguistic tasks are important indicators as to the areas of difficulty in a testee's performance.

Two uses of the test are specifically recommended at the present time. Firstly a teacher may use the test to investigate the specific weakness of a given pupil or group of pupils in either the Fifth Form or those nearing the end of their Fourth Year. The use of the test is then descriptive. It remains for the tester to decide on a course of action, but it can clearly be suggested that the lower the score, the more help is needed, although no single score has been established as a cut-off point.

Secondly a teacher involved in careers guidance may use the test and concentrate on the job subtest relevant to the particular pupil who he is counselling . He will therefore be able to see whether the pupil possesses the skills and knowledge such that he is able to perform the reading tasks associated with his desired career. On this point, it must be noted that the Induction subtest (Job subtest 6) should be considered for _all_ jobs, especially operatives and others.

Administration Instructions

Section I : Max. time = 25 minutes

Section II : Max. time = 25 minutes

Administrator should make a short personal introduction if he is a stranger to the testees.

Administrator should, in all cases, explain, in so far as is possible, the reasons for giving the test (eg. 'to see how well you can deal with passages to read taken from different sorts of job')

Administrator should indicate need for private work rather than copying or talking and indicate that there will be plenty of time for each of the two sections.

SECTION I (Action-items)

1. Instruction

'I am going to give out some booklets. Please do not open them until I tell you to do so, but fill in the details on the front cover.'

Action

Give out booklets.

2. Instruction

'Open your booklets at the first page and follow what is written as I read it through. This page tells you how to answer the first set of questions. It begins 'please read this'.'

Action

Hold up the booklet and indicate the page. Read the text of the page (up to 'Examples'). Ask 'Any questions?'

3. Instruction

'Now turn over to the first example. (Indicate pages to the group). You will see that the question to be answered is on the left-hand page; and that the passage to be read is on the right-hand page.

'Now try to answer this question.' 'When you have finished, please put your pencil down so I can see you are ready. Now try to answer this question.'

(Allow 4 minutes for all to finish)

'Stop now please and check the answers with me.'

'In this example, there are four (4) duties which are the same as those of the Warehouse and General Assistant.

'These are:

'B' which is number 11 in the list of duties
'E' " " " 6 " " " " "
'F' " " part of number 12 in the list of duties
'H' " number 2 in the list of duties

'To get this question right, you must have put a cross in the boxes next to all four answers. Are there any questions?'

'When you do this set of questions remember that you may have to mark more than one box to get the question right. Sometimes, it may be one box, sometimes two, sometimes three or sometiems four boxes. Read each question very carefully so that you do not miss any answers. I'll repeat that.'

Repeat last paragraph.

4. Instruction

'When I tell you, turn over and carry on working through the questions until you come to a page that says 'STOP'.

Then go back and check through your answers. When you have finished that, put your pencil down and sit quietly so that I can see you have finished.

Now turn over and carry on.'

5. After 25 minutes or when everyone has clearly finished 'Stop now please, and put your pencils down.'

SECTION II - Multiple-choice

6. Instruction

'Turn over to the next page (page 23) and follow me as I read through, please.'

Action

Indicate page.

'This page tells you how to answer the next set of questions. Do not turn over yet.'

Action

Read page.

7. Instruction

'Now turn over to the example (Action: indicate page).

You will see that the questions are again on the left-hand page (indicate), three questions this time, and the passage to read is on the right-hand page.

'This time there is one and only one right answer for each question, and you should answer with a tick in the box you think is right.

'Now try to answer these 3 questions. When you have finished, put your pencil down and show me you are ready. Now try to answer these 3 questions. '

(Allow 3 minutes to finish)

'Stop now, please and check the answers with me.'

'In the first question 'C' is right
In the second question 'C' is right and
In the third question 'D' is right'

'Are there any questions?'

8 Instructi on

'When I tell you to do so, turn over and carry on answering the questions until you reach the end of your booklet. Then, go back and check your answers - for the whole booklet if you wish. When you have finished, please put your pencil down and sit quietly, so that I can see that you have finished.

9. After 25 minutes or when everyone has clearly finished.

'Stop now please, and put your pencils down.'

10. Collect booklets.

Marking Procedure (Manual)

The marking key is given below. In questions 1 to 6, the numbers refer to the boxes on the question. To get the questions correct, all key boxes must be correctly marked and no others. Each item correctly answered counts one mark to the total. No fractions should be awarded.

In every case, where the intention of the testee is clear, and it is a correct response, the mark should be awarded, even if he has indicated his response in the wrong way (eg a cross instead of a tick). An unambiguous answer is all that is required. This also applies where a testee has changed his mind. Providing his response in the final analysis is clear, if correct, he should be awarded the mark.

In question 7, all boxes must be correctly completed to obtain the mark.

MARKING KEY (MANUAL)

Question	Box(es)	Question	Box(es)	Question	Box(es)
1	2-3-5-8	2	2-3-5-8	3	1-3
4	1-4-6-8	5	4	6	2-3-4-8
7	see overleaf				
8	E	9	A	10	C
11	D	12	E	13	B
14	A	15	E	16	C
17	D	18	C	19	B
20	C	21	B	22	C
23	C	24	D	25	E
26	C	27	D	28	B
29	D	30	B	31	D

For ease of scoring, the pupil's answers may be coded onto the Class Sheet, the master for which is given overleaf. In columns 1 to 31, the marker should insert a 1 for correct and a 0 for incorrect. When all the tests have been marked, the marker may use the acetate sheet provided to find out the total and subtest totals for each pupil. The sheet is placed over the Class Sheet and moved over each pupil's row of 1's and 0's. By counting up the number of '1's enclosed in heavy blocking, the marker can enter the total correct in the appropriate columns, corresponding to the line on the acetate he is using eg line Cb on the acetate total is entered in columns 42 and 43 for that pupil.

Marking Procedure (Machine)

The test is open to marking and scoring of results by computer using the SOFRP Item Banking System. Users should refer to the authors to obtain coding sheets and instructions.

3

Ashton-sur-Vale
Cheshire
CH6 9AB

4

Jupiter Rentals Ltd
9 Fawcett Road
Welwyn Garden City
Herts

1

J P Brass & Co Ltd
Rutland Avenue
Merseytown
Lancs

7

Mr J Reckitt
Middle Chambers
Law Row
OXFORD
Oxon

8

N P Tolman Esq
57 Englebrook Street
Sheffield
S1 3PZ

2

Anne Charpal
Avenue des Champs-Elysees
PARIS - 84103
France

5

Mrs J Morgan
High Tower
Chesterman
Essex
CL15 9PT

6

Mr F Price
3 Providence Mews
LONDON
N6 3YP

Class: Teacher:

Name	Question Number																														T		C2		C3		C4		C5		C6		J1		J3		J4		J5		J6		L1		L2		L4	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59

References

Berk R A & De Gangi G A — Technical Considerations in the Evaluation of Pediatric Motor Scales <u>American Journal of Occupational Therapy</u> 1979, 33, No 4 pp 240-244

Close J S — Personal communication

Cohen J — A coefficient of agreement for nominal scales <u>Educational & Psychological Measurement</u> 1960, 20, pp 37-46

Department of Education and Science — <u>Education in Schools - A Consultative Document</u> HMSO Cmnd 6869 1977

Harris C W — An interpretation of Livingstone's reliability coefficient for criterion-referenced tests. <u>Journal of Educational Measurement</u> 1972, 9, pp 27-29

Latham W & Parry O — Functional Reading and the Schools in <u>Growth in Reading</u> ed Thackray D V London, Ward Lock Educational 1979

Livingstone S A — Criterion-referenced application of classical test theory <u>Journal of Educational Measurement</u> 1972, 9, pp 13-25

Smith C W — Criterion-referenced Assessment in <u>Contemporary Issues in Educational Testing</u> ed Crombag H F & De Gruyter D N The Hague, Mouton 1974 pp 138-147

Sticht T G, Caylor J S, Kern R P and Fox L C — Project REALISTIC: determination of adult functional literacy skill levels <u>Reading Research Quarterly</u> 1972, 7, pp 424-465

Subkoviak M J — Empirical Investigation of Procedures for Estimating Reliability for Mastery Tests <u>Journal of Educational Measurement</u> 1978, 15, No 2 pp 111-116

Swaminathan H, Hambleton R K & Algina J — Reliability of Criterion-Referenced Tests: a decision-theoretic formulation <u>Journal of Educational Measurement</u> 1974, 11, pp 263-7

APPENDIX VII

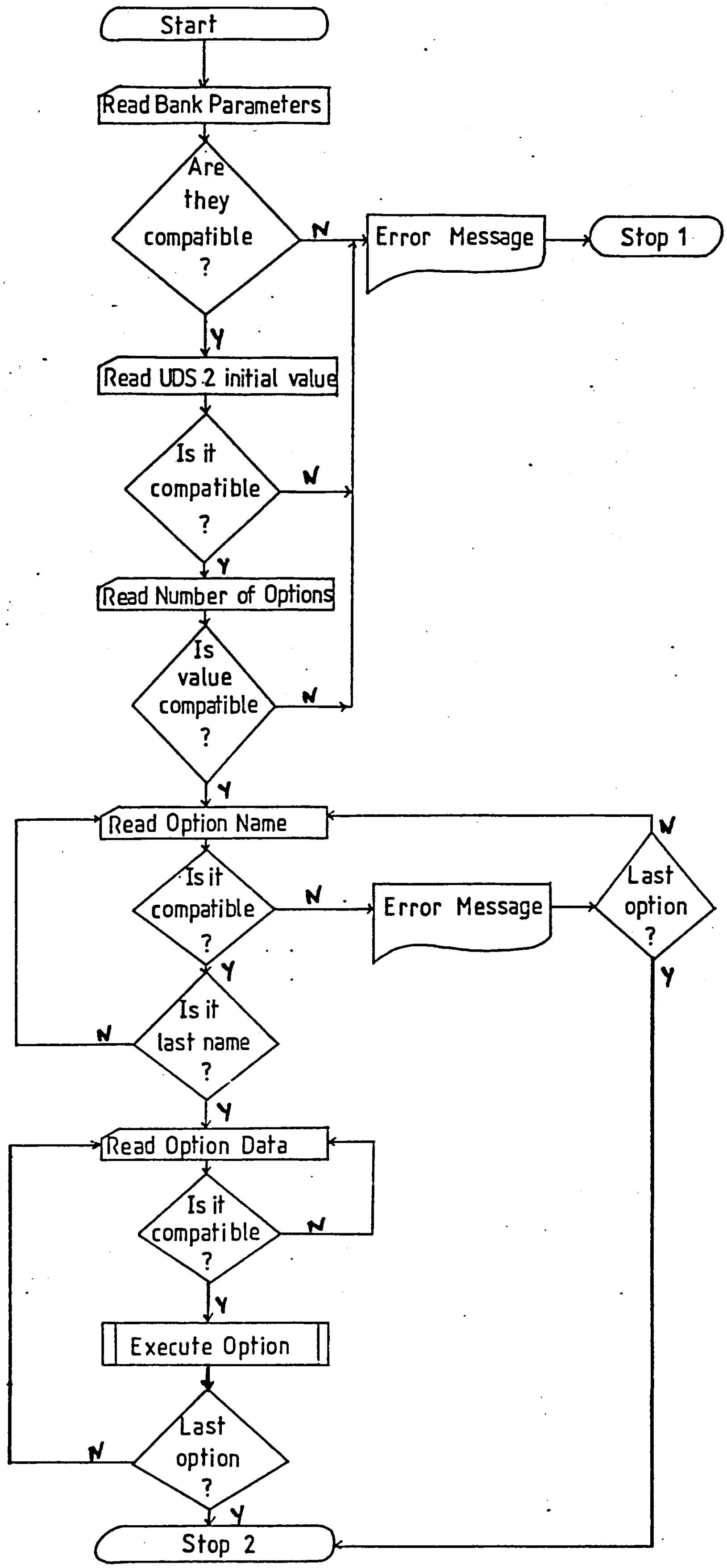
SPECIMEN COPY OF THE FUNCTIONAL

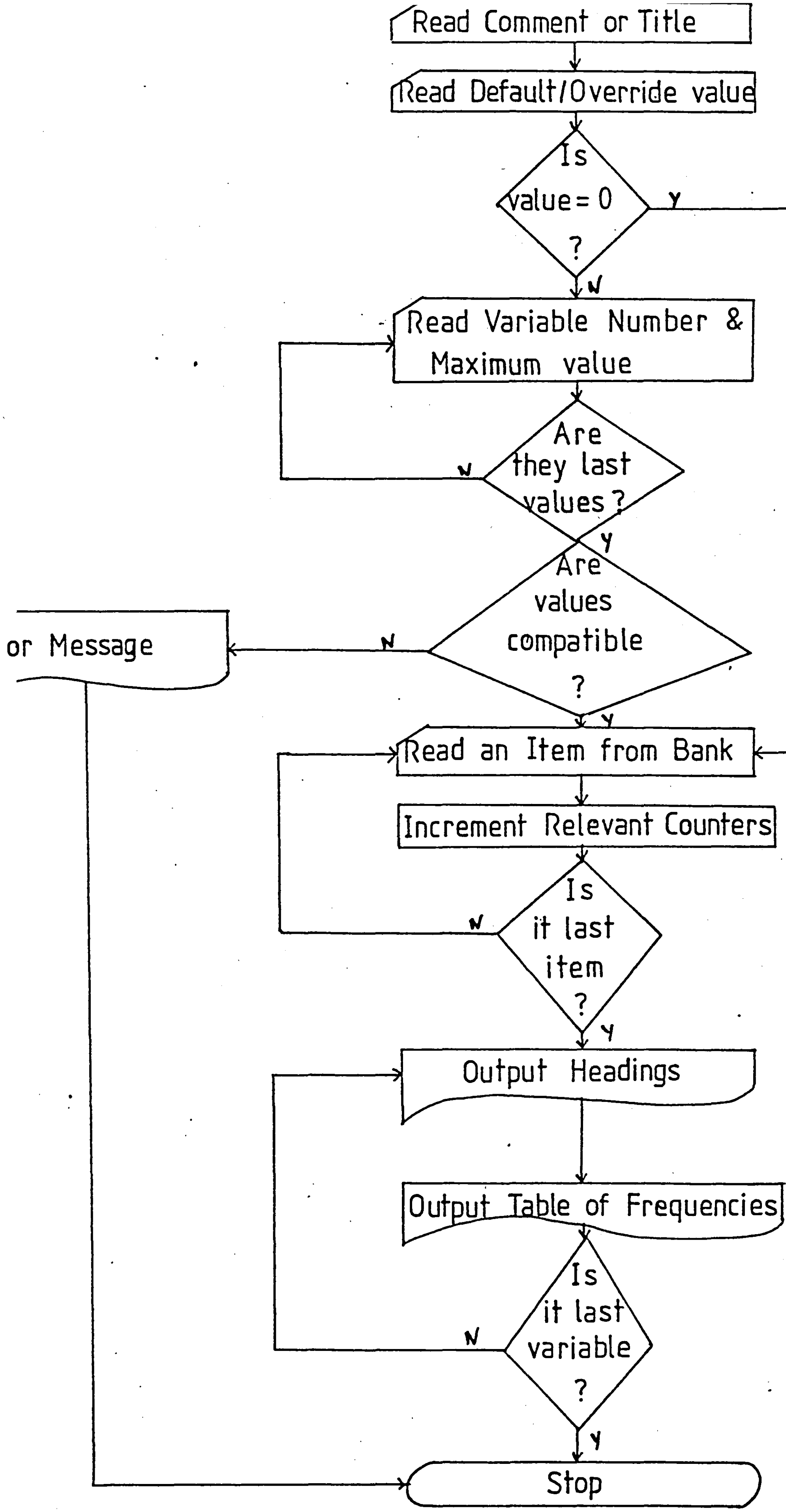
READING TEST, FORM A

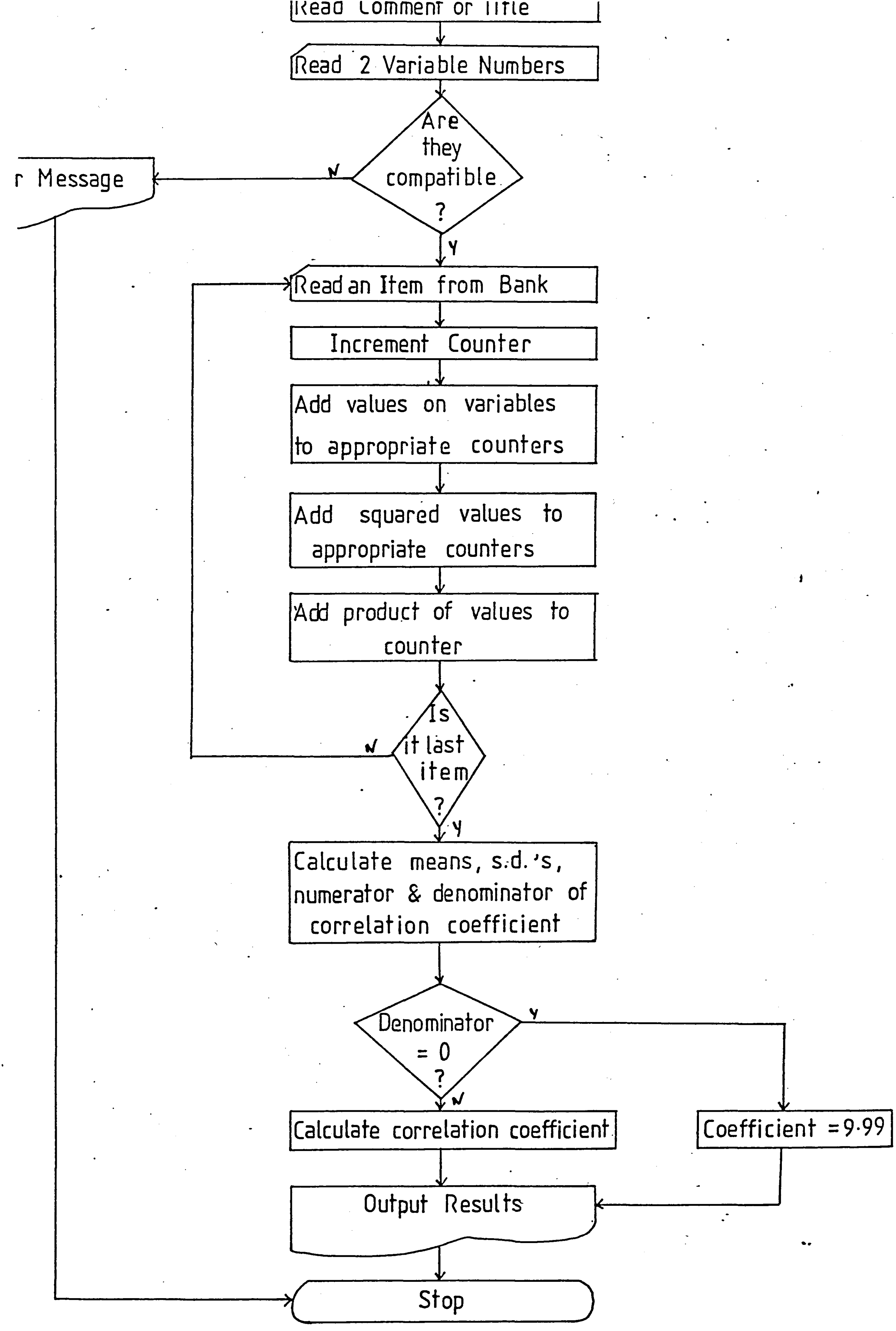
(AS USED AT ALL STAGES FROM RELIABILITY ESTIMATION ONWARDS)

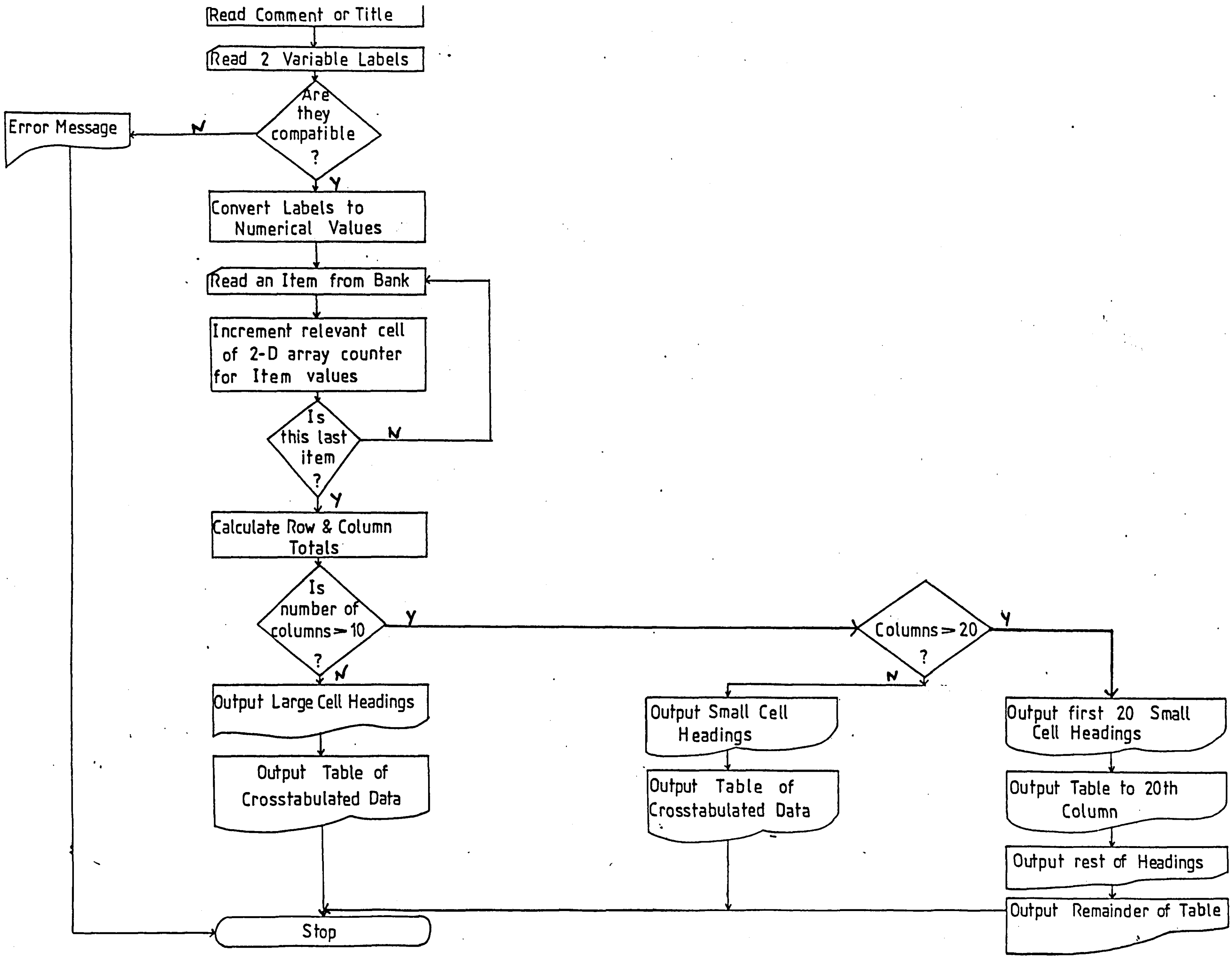
APPENDIX IX

FLOWCHARTS OF THE COMPUTING SYSTEMS

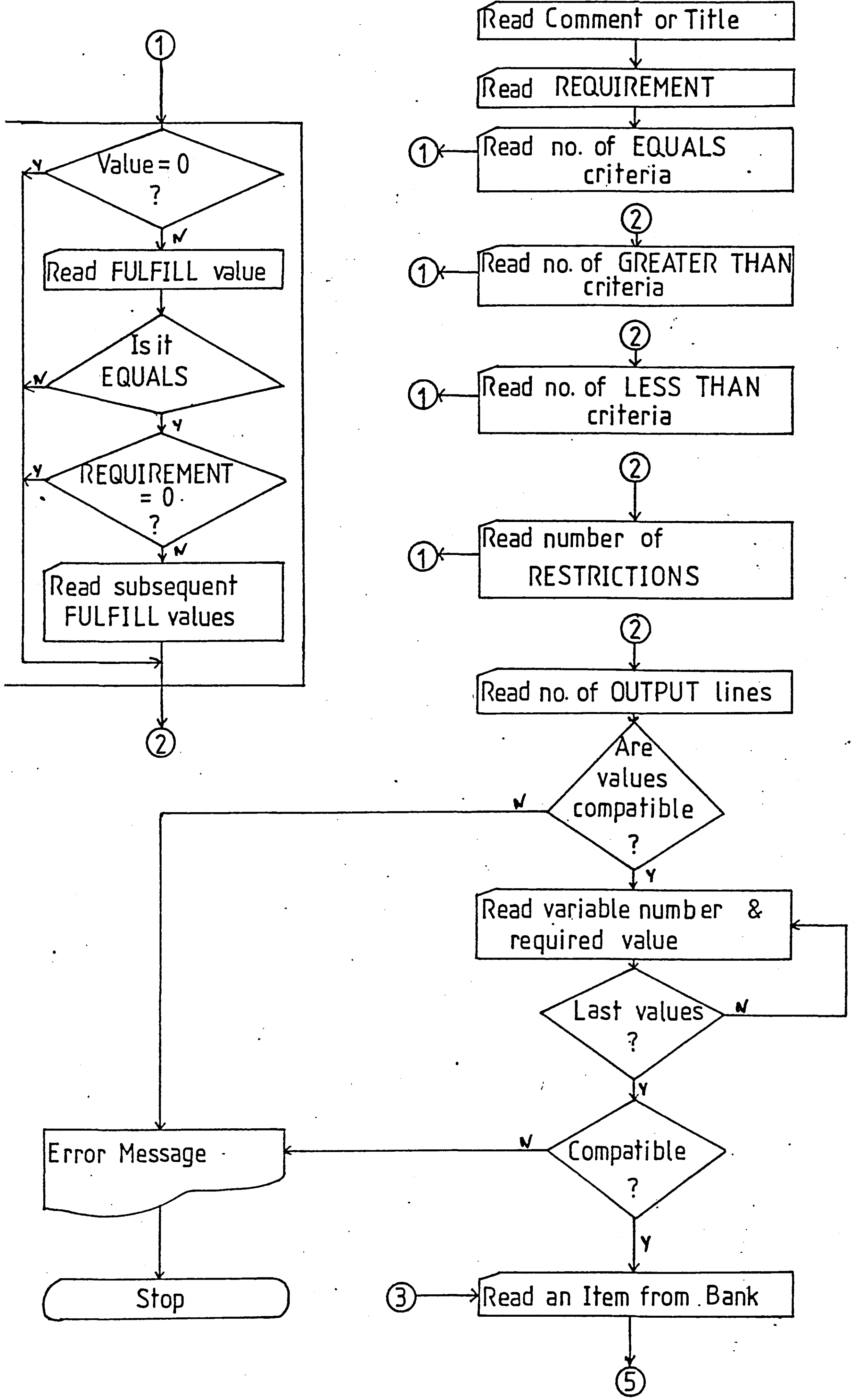

Start
Read Bank Parameters
Are they compatible ?
N
Y
Error Message
Stop 1
Read UDS 2 initial value
Is it compatible ?
N
Y
Read Number of Options
Is value compatible ?
N
Y
Read Option Name
Is it compatible ?
N
Error Message
Last option ?
N
Y
Y
Is it last name ?
N
Y
Read Option Data
Is it compatible ?
N
Y
Execute Option
Last option ?
N
Y
Stop 2


Read Comment or Title
Read Default/Override value
Is value = 0 ?
Y
N
Read Variable Number & Maximum value
Are they last values ?
N
Y
Are values compatible ?
N
Y
or Message
Read an Item from Bank
Increment Relevant Counters
Is it last item ?
N
Y
Output Headings
Output Table of Frequencies
Is it last variable ?
N
Y
Stop

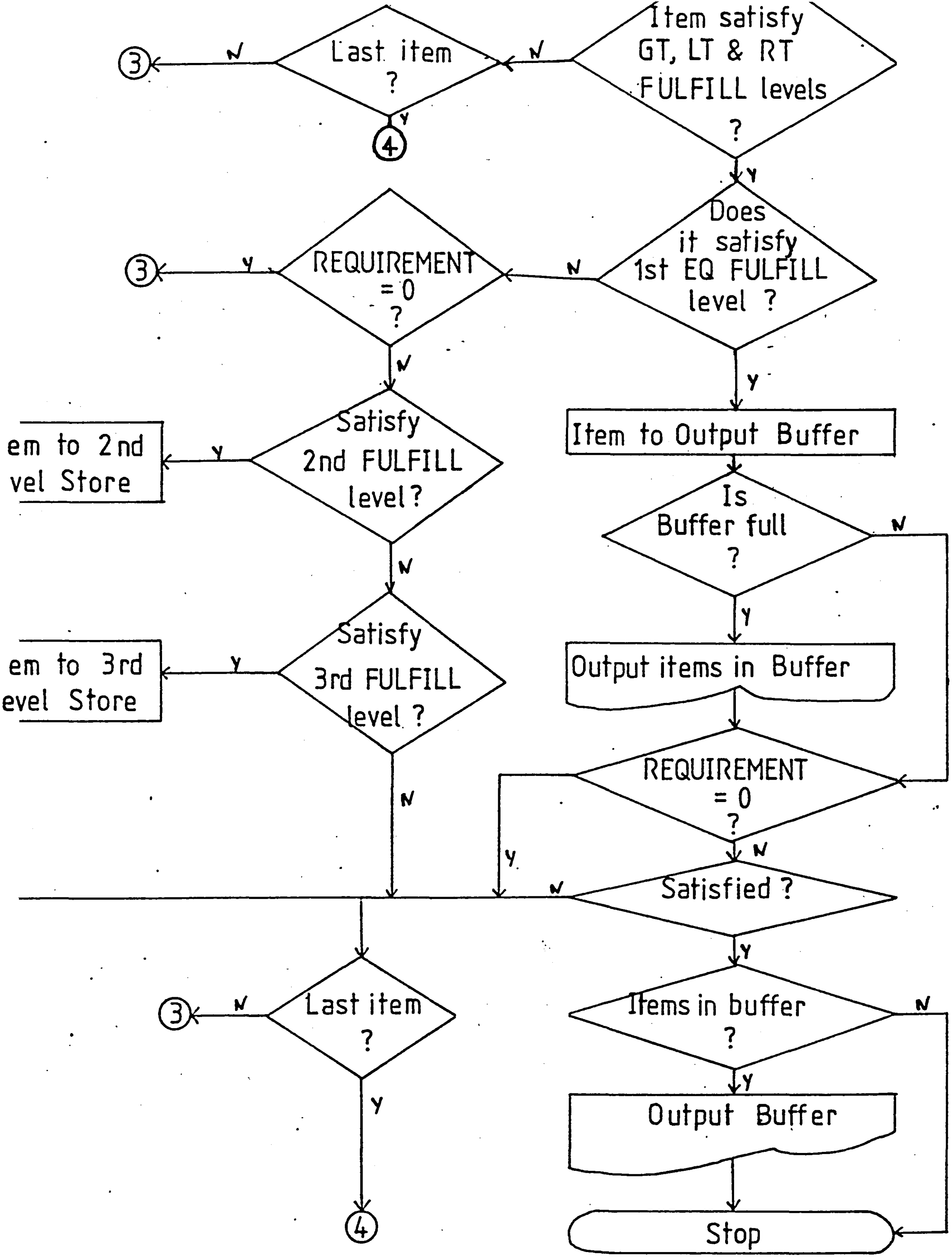

Read Comment or Title
Read 2 Variable Numbers
Are they compatible ?
N
Y
r Message
Read an Item from Bank
Increment Counter
Add values on variables to appropriate counters
Add squared values to appropriate counters
Add product of values to counter
Is it last item ?
N
Y
Calculate means, s.d.'s, numerator & denominator of correlation coefficient
Denominator = 0 ?
Y
N
Calculate correlation coefficient
Coefficient =9·99
Output Results
Stop



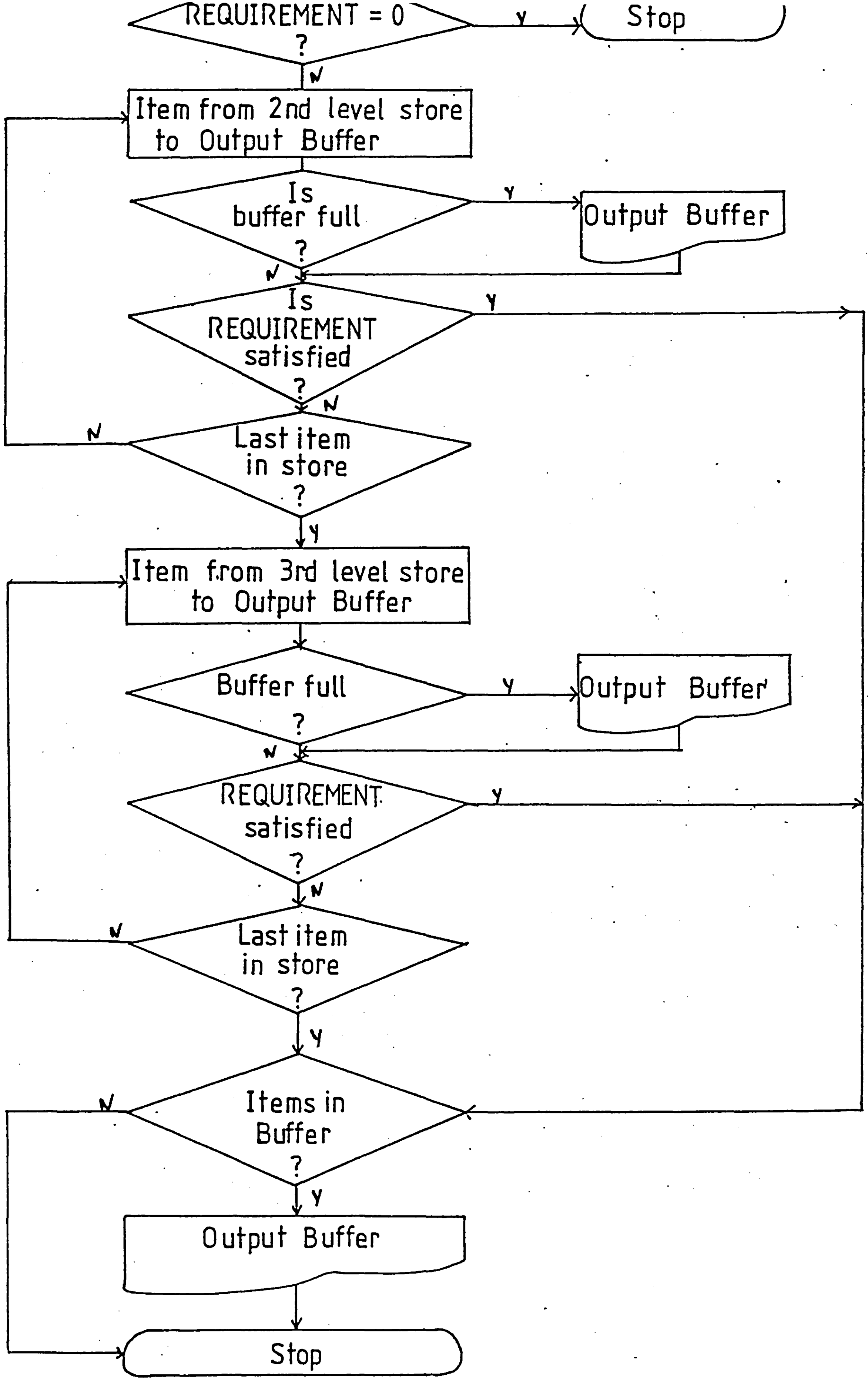
Read Comment or Title
Read 2 Variable Labels
Are they compatible ?
N
Error Message
Y
Convert Labels to Numerical Values
Read an Item from Bank
Increment relevant cell of 2-D array counter for Item values
Is this last item ?
N
Y
Calculate Row & Column Totals
Is number of columns > 10 ?
Y
N
Output Large Cell Headings
Output Table of Crosstabulated Data
Columns > 20 ?
Y
N
Output Small Cell Headings
Output Table of Crosstabulated Data
Output first 20 Small Cell Headings
Output Table to 20th Column
Output rest of Headings
Output Remainder of Table
Stop

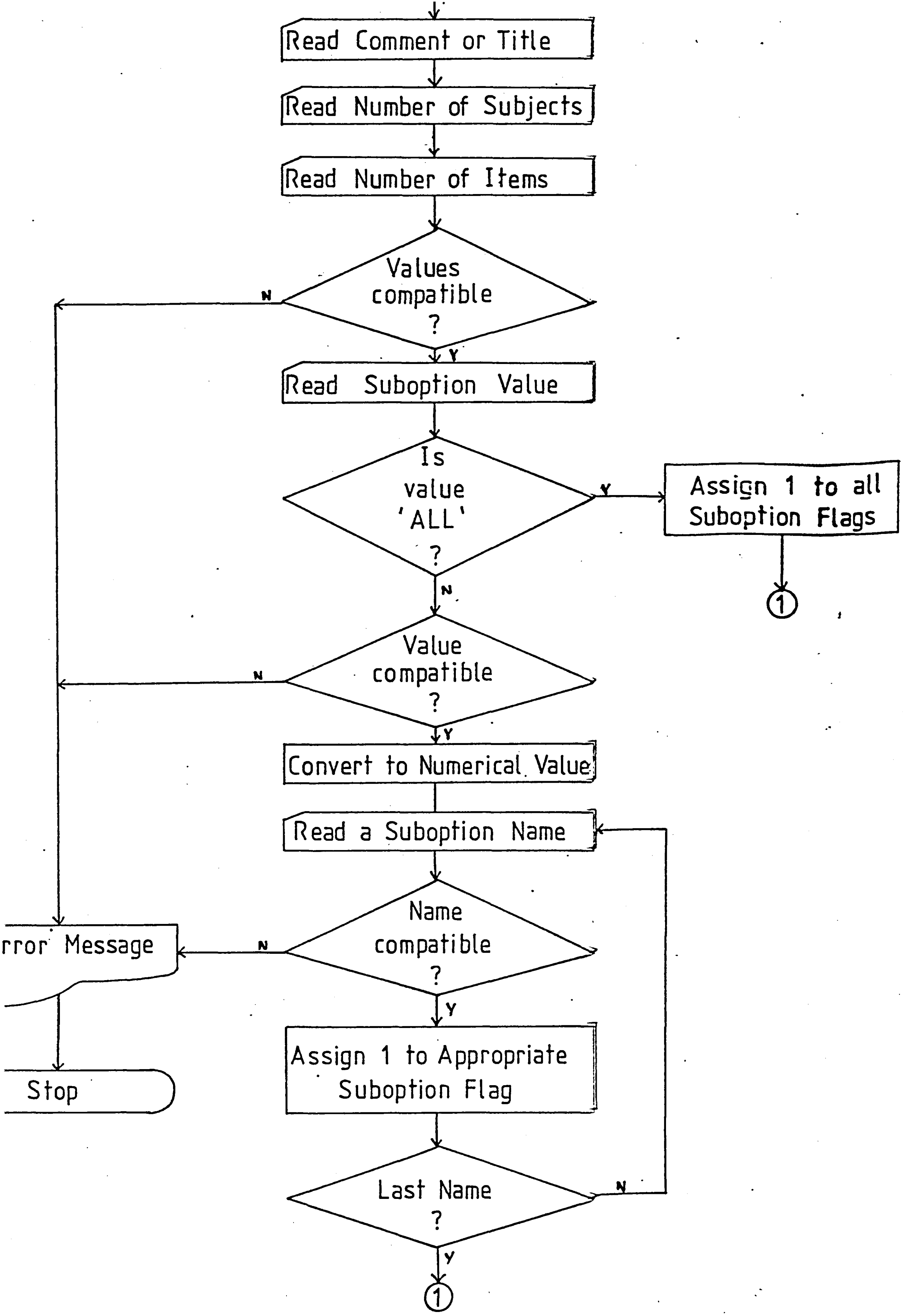


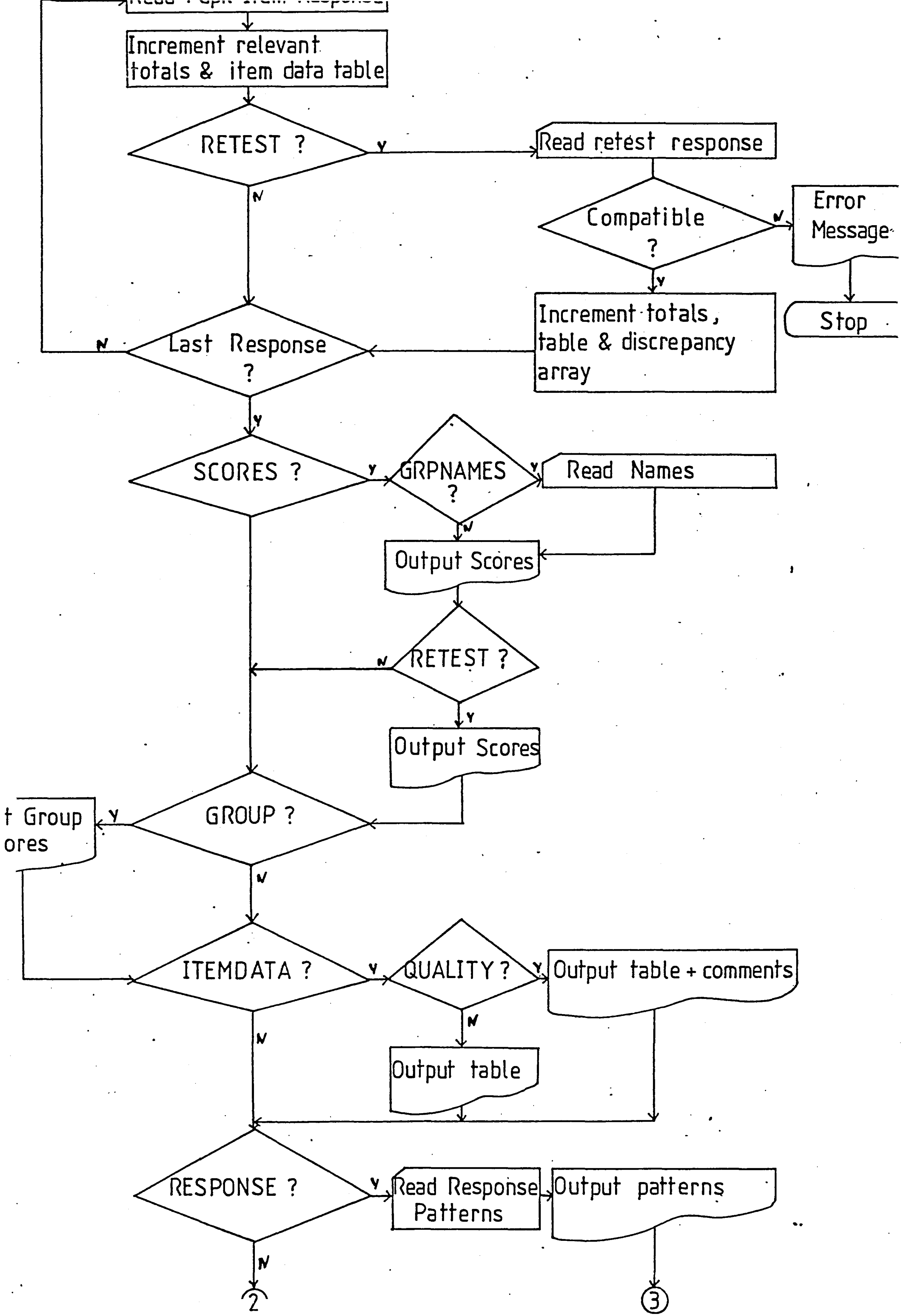
1
Value = 0 ?
Y
N
Read FULFILL value
Is it EQUALS
N
Y
REQUIREMENT = 0 ?
Y
N
Read subsequent FULFILL values
2
Read Comment or Title
Read REQUIREMENT
Read no. of EQUALS criteria
1
2
Read no. of GREATER THAN criteria
1
2
Read no. of LESS THAN criteria
1
2
Read number of RESTRICTIONS
1
2
Read no. of OUTPUT lines
Are values compatible ?
N
Y
Read variable number & required value
Last values ?
N
Y
Compatible ?
N
Y
Error Message
Stop
3
Read an Item from Bank
5



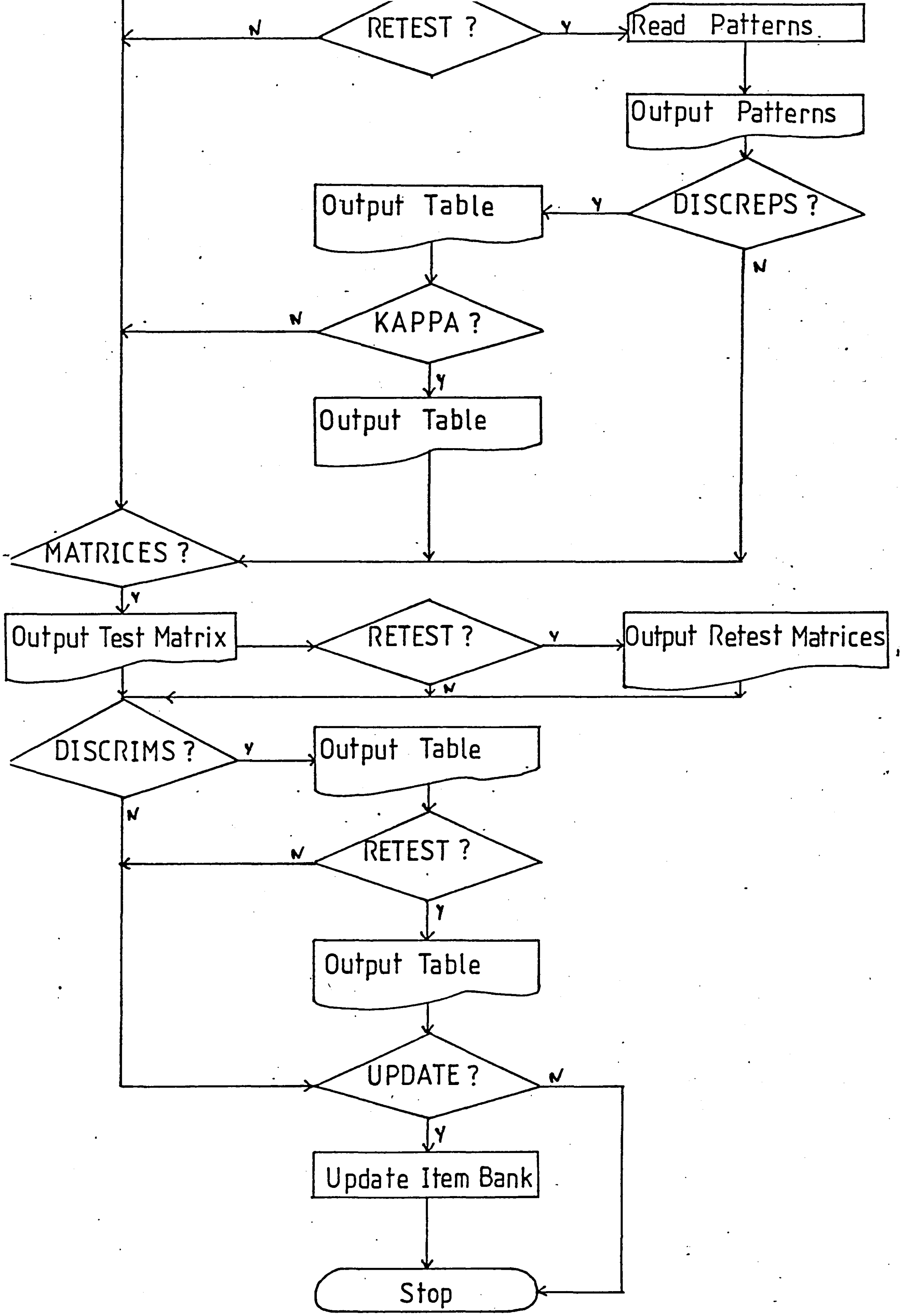
Last item ?
Item satisfy GT, LT & RT FULFILL levels ?
Does it satisfy 1st EQ FULFILL level ?
REQUIREMENT = 0 ?
Satisfy 2nd FULFILL level?
em to 2nd
vel Store
Satisfy 3rd FULFILL level ?
em to 3rd
evel Store
Item to Output Buffer
Is Buffer full ?
Output items in Buffer
REQUIREMENT = 0 ?
Satisfied ?
Last item ?
Items in buffer ?
Output Buffer
Stop

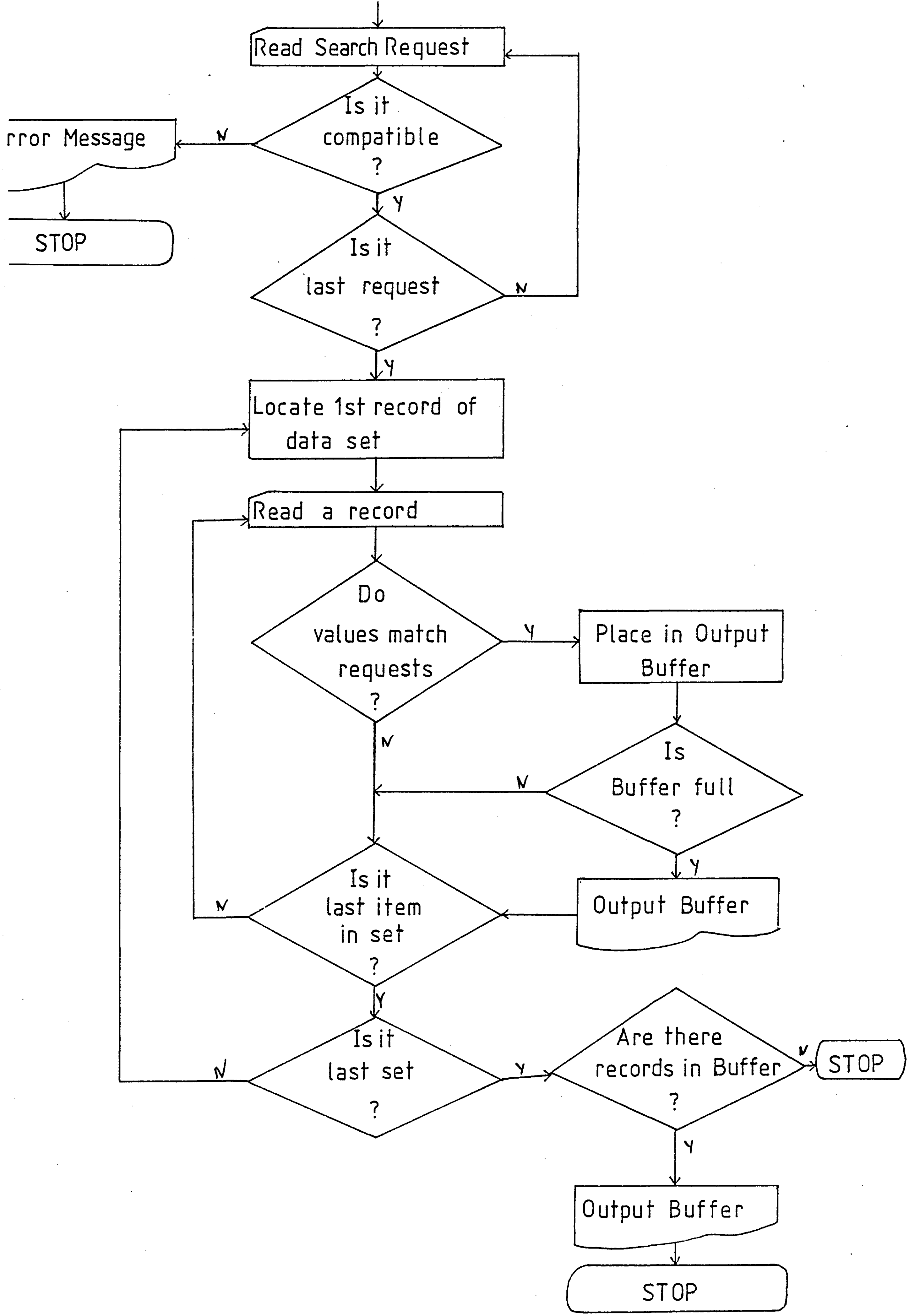

REQUIREMENT = 0 ?
Y
Stop
N
Item from 2nd level store to Output Buffer
Is buffer full ?
Y
Output Buffer
N
Is REQUIREMENT satisfied ?
Y
N
Last item in store ?
N
Y
Item from 3rd level store to Output Buffer
Buffer full ?
Y
Output Buffer
N
REQUIREMENT satisfied ?
Y
N
Last item in store ?
N
Y
Items in Buffer ?
N
Y
Output Buffer
Stop


Read Comment or Title
Read Number of Subjects
Read Number of Items
Values compatible ?
N
Y
Read Suboption Value
Is value 'ALL' ?
Y
Assign 1 to all Suboption Flags
1
N
Value compatible ?
N
Y
Convert to Numerical Value
Read a Suboption Name
Name compatible ?
N
rror Message
Stop
Y
Assign 1 to Appropriate Suboption Flag
Last Name ?
N
Y
1



Increment relevant totals & item data table
RETEST ?
Read retest response
Compatible ?
Error Message
Stop
Increment totals, table & discrepancy array
Last Response ?
SCORES ?
GRPNAMES ?
Read Names
Output Scores
RETEST ?
Output Scores
GROUP ?
t Group ores
ITEMDATA ?
QUALITY ?
Output table + comments
Output table
RESPONSE ?
Read Response Patterns
Output patterns
2
3


RETEST ?
N
Y
Read Patterns
Output Patterns
DISCREPS ?
Y
N
Output Table
KAPPA ?
N
Y
Output Table
MATRICES ?
Y
Output Test Matrix
RETEST ?
Y
N
Output Retest Matrices
DISCRIMS ?
Y
N
Output Table
RETEST ?
N
Y
Output Table
UPDATE ?
N
Y
Update Item Bank
Stop



Read Search Request
Is it compatible ?
N
rror Message
STOP
Y
Is it last request ?
N
Y
Locate 1st record of data set
Read a record
Do values match requests ?
Y
Place in Output Buffer
N
Is Buffer full ?
N
Y
Output Buffer
Is it last item in set ?
N
Y
Is it last set ?
N
Y
Are there records in Buffer ?
N
STOP
Y
Output Buffer
STOP

Functional reading and the school

William Latham and Owen Parry

Introduction

In the Green Paper *Education in Schools – a Consultative Document* (Cmnd 6869 1977) it was argued that 'schools must prepare their pupils for the transition to adult life and work'. If such a preparation is accepted as one of the duties of the school, it seems reasonable to expect secondary schools to prepare their pupils, as far as may be possible, for the functional reading (i.e. reading associated with a task or role) which will be required of them when they first leave school. For instance, the pupil leaving school at 16+ should be prepared for functional reading related to his/her role as a job seeker or worker (and, one might add, as consumer and citizen).

Unfortunately, it is not reasonable to expect the secondary school to prepare pupils for functional reading related to any of the roles mentioned above, as the reading tasks involved have yet to be identified and classified; and, obviously, no tests exist to assess the pupil's progress towards success in functional reading or diagnose his/her difficulties.

In September 1977 Sheffield City Polytechnic, in association with Sheffield Metropolitan District Council, initiated a research programme aimed at providing Sheffield schools with both information and assessment and diagnostic tests related to functional reading in the areas of job seeking and employment.

The research programme

The research programme has the following three objectives:

1 the identification and classification of reading tasks associated with job seeking and employment in Sheffield, which a 16+ year old school leaver faces immediately he leaves school;
2 the construction and validation of a criterion-referenced group test based on the reading tasks identified in achieving objective (1) above;
3 the construction of diagnostic tests to be used with individuals who fail to reach the criterion scores on the group test.

A fourth objective, subject to discussion with the schools, will be the production of materials to be used in teaching related to functional reading. The progress of the research towards objectives (1) and (2) above is described below.

The research
In order to investigate the reading requirements of job seeking and employment, we must first ask what areas of employment take on school leavers at 16+; what sort of jobs the leavers seek and enter; and what are the reading requirements involved. To answer the first of these questions, the help of the Sheffield Careers Service was sought. They kindly provided statistics relating to the 'occupational destinations' of school leavers since the raising of the school leaving age. We have found it useful to use the categorization of types of employment and types of job to which these statistics are related. Types of employment are categorized by the Standard Industrial Classification (SIC) (Central Statistical Office 1968) (see Table 1); and types of job into the categories shown in Table 2.

Table 1 gives a breakdown of numbers entering the various areas of employment. The metal working and engineering industries combined are gaining a larger proportion of leavers over time, whilst certain manufacturing industries take fewer leavers every year. The Distributive trades continue to take on the largest number in any one category. There is some change within groups, however: 'Engineering' and 'Engineering Small Tools' have taken on fewer and fewer leavers in the past few years, whilst 'Other Metal Manufacturer' has increased greatly.

Table 1 Analysis of school leavers' first jobs; 1976/7

Industry	*Total*	*Percentage*
Agriculture	32	0.58
Mining	41	0.74
Food, Drink and Tobacco	193	3.52
Chemical and Allied Industry	9	0.16
Metal Manufacture	487	8.84
Engineering	139	2.52
Electrical Goods	54	0.98
Vehicles	8	0.15
Engineering Small Tools	206	3.74
Hand Tools	230	4.17
Cutlery	434	7.88
Other Metal Goods	385	6.99
Textiles	14	0.25
Clothing	109	1.98
Bricks, Pottery and Glass	27	0.49
Timber and Furniture	73	1.32
Paper, Printing and Publishing	56	1.01
Other Manufacturing Industries	28	0.51
Construction	419	7.60
Gas, Electricity and Water	47	0.85
Transport and Communications	78	1.42
Distribution	1209	21.94
Insurance Banking and Finance	194	3.52
Professional and Scientific Services	273	4.95
Miscellaneous Services	193	3.52
Hairdressing	96	1.74
Motor Repairs	229	4.16
Public Administration	247	4.48
Total	5510	100.01

Table 2 Definitions of job types

Apprenticeship	in which articles of apprenticeship are signed and agreed, with national regulations, block and/or day release training
Professional:	recognized training in one of the professions, e.g. articled clerk in accountancy
Clerical:	general office and clerical duties which may or may not include any training
Operative Training:	specific, on-the-job training for a minimum period of two months
Others:	no formal training longer than two months, or none at all; non-clerical

Sampling

It was proposed to take a proportional stratified sample of each industrial category. Before doing this, however, it was decided to try and decrease the number of categories to be considered, as many areas employed few school leavers at 16+. All categories employing less than 1 per cent of the total number of school leavers (1976/7) were excluded for the following reasons:

1 This will exclude categories which have previously only accounted for 4.71 per cent of all school leavers in 1976/7 (260 persons, see Table 1).
2 The number of categories is reduced to nineteen, which will save research time and effort reasonably employed elsewhere.
3 No excluded category represents more than 1.5 per cent of any job type and therefore specific job-related reading materials from these categories would be of limited value.

It was felt that random sampling within each category might well result in the 'swamping' of the sample with small firms. This was considered undesirable as it was thought that the reading requirements of jobs in larger firms may be greater than in smaller firms, where there is likely to be greater oral communication. Further, small firms, *per se*, are less likely to take on school leavers in any significant numbers. Therefore, firms employing less than a certain number of employees (all employees, not just school leavers) were to be eliminated from consideration in this sample. The data kindly made available by the Employment Service Agency did not include firms with less than thirty employees; this was therefore taken as an appropriate cut-off point, after due consideration.

Firms were then stratified according to size, into 'Small', 'Medium' and 'Large' categories; 'Small' firms usually had 30 to 100 employees, 'Medium' had 101 to 500 employees and 'Large' had 501+ employees. Occasionally, however, eight or nine firms appeared in the data for a single employment category as having 1000+ employees. Rather than group all of these very large firms with smaller ones, it was decided that, for groups where there were several 1000+ firms, they should be the 'Large' size category, whilst 'Medium' would be 101 to 1000.

A further problem was that of heterogeneity of industries within a category; for instance, the very mixed set of employments grouped together under the heading 'Miscellaneous Services'. With a limited number of firms to be sampled from each category, it might well be that not all industries in a highly heterogeneous category would be sampled, or that less important industries, in terms of the number of school leavers employed, would be included in the sample at the expense of the more important. With this in mind, some of the more heterogeneous categories were examined in more detail and certain areas of employment excluded, usually on the grounds of the small number of school leavers employed.

The category 'Public Administration' needed special consideration. It is, again, a fairly heterogeneous category. It was felt that access to branches of the Civil Service (except in relation to job seeking, i.e. Job Centres) would be difficult and that the time and effort involved in adequately covering this organization was not justified by the small number of school leavers employed by it. The Armed Forces remove their recruits for training elsewhere and so may be discounted. Also, in Local Government Service the Police and Fire Services recruit very few school leavers at 16+ years of age. Therefore, in its study of 'Public Administration' the project confined itself to Sheffield Metropolitan District Council, its collaborating institution; the nature and scope of that study to be determined in conjunction with the Council.

With the above considerations in mind, a proportional, random sample, stratified by SIC category and by size, was taken. The role of Industry Training Boards (ITBs) and Group Training Associations (GTAS) is a key factor in the sampling method. ITBs all exist to provide standards within their industry, but many also provide basic training, as do the GTAs. Such training may be 'off the job' at a special training centre, 'on the job' at the workplace, or a combination of these two. With little exception, block or day release to a College of Further Education makes up some part of this training. There are often limits placed on the types of employee allowed to go on such courses, however; typically, all apprentices attend, as do many operative trainees. It is useful to take those categories covered by a Training Board or Association and consider them separately.

The Sampling Model for Employment (Figure 1) shows how this aids the obtaining of reading materials and the sampling of employees. From the sample of firms in categories with ITB/GTAS, firms will be divided up into those that use the ITB/GTA training schemes (or related schemes) and those that do their own training. Of the former group, information can be gained direct from the Board or Association whilst job types for which they provide no schemes will require separate investigation. Further Education courses are also a source of information concerning reading requirements. In short, there are five routes to the gaining of reading materials from employment for test design: via ITB/GTAS; via firms using these schemes; via other firms in these categories; via firms in other categories; and via Further Education college courses. Job Centres and Trade Unions add further dimensions to test design.

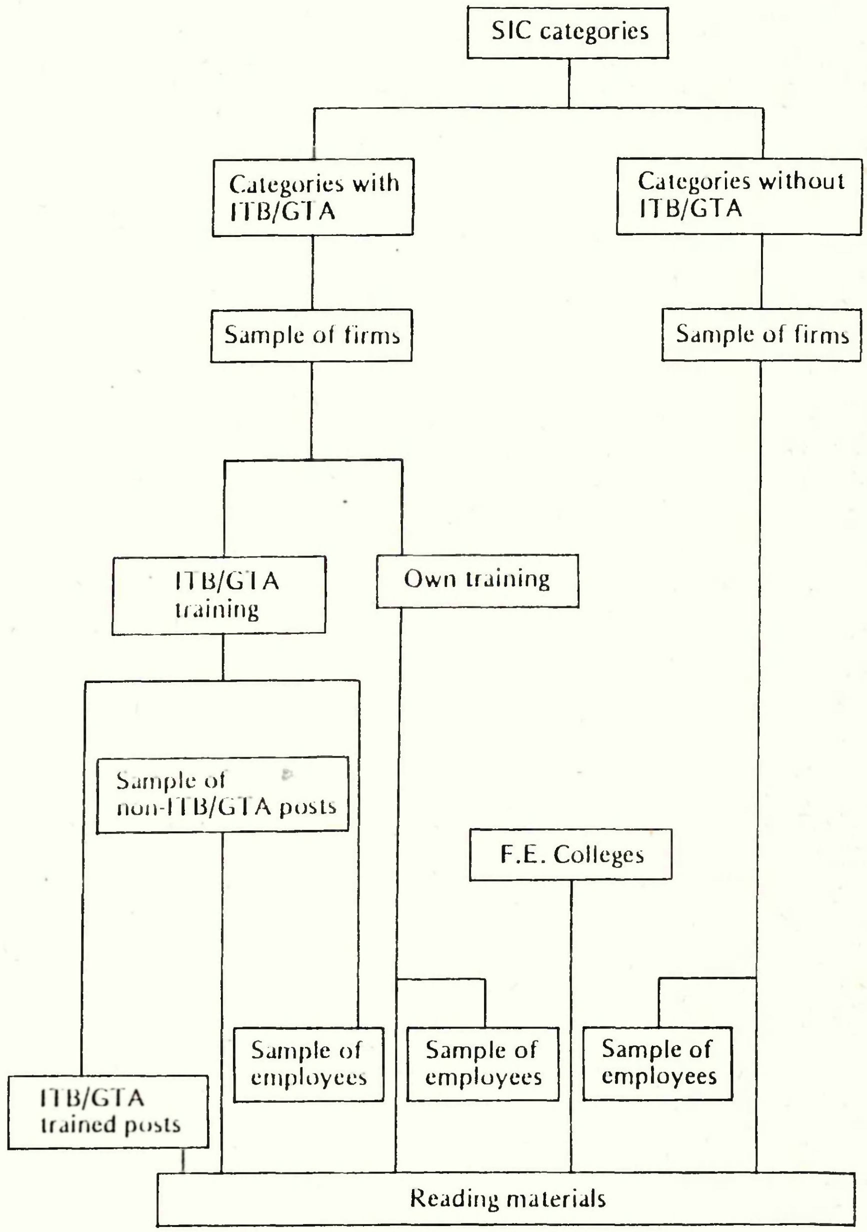
Nineteen categories were finally selected; fifteen are covered by a Training Board or Association, four are not. The proportion of the sample allotted to each category was determined by the proportion (of the total) of school leavers entering employment in that category. The data in Table 1 have been used to arrive at the number of firms to be samples from each category. Whilst it is recognized that the data in Table 1 include all school leavers, including those leaving at 17 and 18+, it was felt that the sampling method was sufficiently accurate for the purposes of the study. A sample of the 116 firms, plus departments of Sheffield Metropolitan District Council, was taken.

In summary, we have attemped to use a sampling model which will give widest possible coverage of job-related reading tasks encountered by school leavers in Sheffield.

Contacts

Before each SIC group was contacted, representatives of the relevant ITB were seen and asked to give their opinion on the representativeness of their area of the sample. In three cases this meant that resampling to replace a firm had to take place. Replacement was for such reasons as recent takeovers, bankruptcies etc. not mentioned in the original data from the Employment Service Agency. Where applicable, details of the ITB/GTA scheme were noted so that relevant personnel could be interviewed at a later stage.

Figure 1 Sampling model for employment



Information was also obtained about the sort of data which might be encountered in that industry, such as the type of training systems used etc.

The Sheffield Careers Service was again most helpful in allowing names of contacts for each firm to be extracted from their files. One hundred and three names were obtained in this way, or provided by the ITB/GTAS. To date, of the firms contacted, twenty-five have declined or been unable to help, but sixty-five interviews have been successfully completed. Firms were contacted one SIC category at a time, in order of importance; but beginning by joining all the metal working and engineering industries together (41 per cent of all school leavers). Initially, four areas of cooperation were requested:

1 an initial discussion between the firm's representative and the Research Assistant to the project;
2 the inspection of reading materials actually used on the job by school leavers;
3 discussion with some of the firm's recent school-leaver employees about the situations on the job in which reading is required;
4 observation of some of the leavers at work in situations which require reading.

It was found, however, that the latter two areas were inappropriate at that time of year (March to July), as most school leavers are taken on in the August–September period and, hence, the material being used was not the initial material in which we were most interested. It was decided, therefore, that a second stage of interviewing of a sub-sample of school leavers, later in the year, would fulfil the requirements of the discussion and observation stages. This second round of interviewing has the added advantage of coming at a time when areas of overlap in reading requirements will have been discovered, thus cutting down on the work involved.

In parallel with the collection of job-related reading materials and information concerning their use, detailed work has been carried out on the factors involved in item construction. Presented with the wealth of material from the sample, it is necessary to sort out the materials which are representative of the jobs undertaken by school leavers and to construct items which reflect reading tasks performed by them. The factors involved are discussed below.

Item construction

It is proposed to classify materials according to a system developed by Sticht and his associates in their work with the United States Army (Sticht *et al* 1972). Using this classification system, we can then relate reading materials to job type and employment area and construct items which adequately represent the range of reading tasks involved. The classification is given in Table 3.

Table 3 Content-type categories

1 Tables of content and indexes

2 Standards and specifications

3 Identification and physical description

4 Procedural directions

5 Procedural checkpoints

6 Functional description

The definition of each of these content-type categories follows:

1 *Tables of content and indexes*
Content designating the location of information within a publication.
An example would be use of a telephone directory.

2 *Standards and specifications*
Content setting forth specific rules or tolerances to which task procedures or the completed product must conform.
An example would be conditions in an insurance policy.

3 *Identification and physical description*
Content attempting to represent symbolically an object via an identifying code (stock, nomenclature) and/or by itemizing its distinguishing physical attributes.
An example would be an inventory or price list.

4 *Procedural directions*

Content which presents a step-by-step description of how to carry out a specific job activity. Essential elements are equipment/materials/ingredients to be used, and how they are to be used, with presentation organized in a sequential step-wise fashion.

A good example is a vehicle maintenance manual.

5 *Procedural checkpoints*

Content which presents a key word or highly summarized version of what should be done in carrying out a task rather than how it should be done. This content differs from the content classified under Procedural Directions in that it assumes the user knows how to carry out the steps once reminded that the step exists and/or reminded of the decision factors which determine whether the step is required.

Here, a job card related to an industrial process is a good example.

6 *Functional description*

Content which presents an operating (cause and effect, dependency relationships) description of some existing physical system or sub-system, or an existing administrative system or sub-system.

A description of an industrial process, say in a training manual, would be an example of this.

Certain factors are of importance when considering which types of test item – multiple-choice, sentence completion, cloze procedure etc. – are appropriate for the assessment of ability in this area. Firstly, the pupil cannot be expected to bring prior technical knowledge to the test (technical not only in content but also in the sense of special jargon or vocabulary). The main focus of the items, therefore, must be on the different content-type categories given above and the pupil's ability to cope with them despite unfamiliar terms. Secondly, items must try to represent real reading requirements of jobs, of the sort actually undertaken by school leavers. As an example, cloze procedure would be an inappropriate form of item, as no employee is ever required to fill in blank spaces in a written passage, while an item that requires ticking off items on a price list does represent an appropriate item, this being a fairly common task at work.

Various forms of multiple-choice items (where the questions refer to some action the reader might have to take, for example asking for the next step in a sequence of instructions) and 'action-type' items (such as a filing exercise of the ticking off on a list already mentioned); these seem to be the most appropriate types of item, though not all the material is yet in from the sample to completely confirm this.

Item banking

Also in parallel with work already mentioned is the work on item banking. If the items constructed exceed in number the items actually used in the final test or tests (as seems likely), it would seem useful to keep all the items for future use, for example the construction of parallel forms. Some item-banking projects are far advanced (for example the NFER Project on Item Banking), but it is not proposed to enter into such complexities in this study.

Items will be filed by assigning each a number, which will relate to a set of data about the item stored by the computer. This data will consist of a 'history' of the item (where the material was derived, which employment area, which type of job), classificatory data (content-type, item-type, key etc.), and 'usage' data (number of testees, per cent correct answer, date of last use, number of users etc.). Such data will assist test construction, as specific types of item can be swiftly selected, and simple statistical calculations performed as necessary.

Work to be carried out in the immediate future

An outline of the next stages of the research is given below.

Job Centres and Trade Unions

After the completion of the interviews in the employment areas, job seeking will be considered, via interviews with Job Centre personnel. Relevant Trade Unions will be contacted about documentation they may provide for school leavers starting work.

Content validation

It is proposed to use panels of employers to obtain content validation of items as representatives of tasks in their areas; and of linguists to examine the structure of the tasks for any undue complexities in the questions and underlying structures not previously considered in the construction of the items.

Piloting and further validation
Items will be piloted amongst the target population in the autumn term, leading to initial measures of reliability. A sweep test of the target population will be carried out in the following term. A predictive validity study is planned, to consider the relationships between test performance and adequate job performance.

References

CENTRAL STATISTICAL OFFICE (1968) *The Standard Industrial Classification* HMSO

Cmnd 6869 (1977) *Education in Schools – a Consultative Document* HMSO

STICHT, T. G., CARLOY, J. S., KERN, R. P. and FOX, L. C. (1972) Project REALISTIC: determination of adult functional literacy skill levels *Reading Research Quarterly* 7, 424–65

A CRITERION-REFERENCED TEST OF OCCUPATIONAL FUNCTIONAL READING ABILITY

Paper presented to the 4th Annual Conference of the Reading Association of Ireland, St Patrick's College, Dublin, 13-15 September 1979

by

William Latham, Co-ordinator

and

Owen Parry, Research Assistant

Language Development Centre
Sheffield City Polytechnic
Sheffield
S10 2LD

A criterion-referenced test of occupational functional reading ability

W Latham and O Parry

I Introduction

Functional reading may most simply be defined as reading related to a task or role. The study of functional reading ability, therefore, is concerned with the investigation of the different tasks in which reading is required and the assessment of both the level of ability required and the extent to which it is being achieved by the individual.

It was suggested recently, in the Green Paper: 'Education in Schools - A Consultative Document', that 'schools must prepare their pupils for the transition to adult life and work'. If this suggestion is accepted, it seems reasonable for schools to attempt to prepare their pupils for the functional reading they will encounter on leaving school, not only as an employee or trainee, but also as a citizen and consumer. Unfortunately, the schools cannot be expected to attempt such preparation as studies of functional reading ability, of the type mentioned above, have not yet been undertaken.

Such a study is the purpose of the Sheffield Occupational Functional Reading Project. In collaboration with Sheffield Metropolitan District Council the Project is studying the functional reading required of 16 year old school leavers on entering employment or training for a job, and attempting to devise assessment and diagnostic instruments of occupational functional reading ability.

II Collection of Job-related Reading Materials

As suggested above, the Project first proceeded with the identification and classification of job-related reading tasks for this group of young people. A representative sample of firms in Sheffield was visited, selected by a process previously discussed (Latham and Parry (1979)) to give a broad range of industries and types of job. Details of numbers of leavers employed, types of job. training procedures, etc. were collected, along with specimens of job-related reading materials. Industry Training Boards, Group Training Associations and Colleges of Further Education were similarly visited, where relevant. Materials were also collected from the local Careers Service.

III Classification by Type of Job

A large volume of material was thus collected, of many varying types. Reading materials included job cards, stock lists, safety regulations, descriptions of machinery, leaflets on security, personal appearance, credit card completion, and many other types. This plethora of disparate material becomes more comprehensible, however, if one groups similar types of job together. Six job-types have been used, corresponding, in the main, to those categories commonly employed by the Careers Services. These are given in Table 1.

Within each category, it has been possible to identify broad areas of overlap and many common elements. The Project is concentrating on a possible critical period in employment - the first six weeks - in which a school-leaver undergoes an adjustment to working life. Of course, technical language differs to some degree from one industry to the next, but the procedures involved in data collection resulted in a sampling of such language across the board. Here, then, we are concerned with the type of structure and content of job-related reading materials, taking vocabulary into account, and with the linguistic tasks posed by the materials encountered in the first six weeks. These factors will be considered in more detail below.

IV Test Item Construction

It is an aim of the Project to construct a criterion-referenced test of occupational functional reading ability, based on those reading materials, for use at the start of the last year at school, with those leaving at 16+. It was necessary, therefore, to construct test questions or items for this purpose.

Certain a priori criteria existed: the test must be objective (ie. have clearly identifiable correct answers for each question); it must fit into a reasonably short time span (eg. a double school-period); there must be ease of administration and marking. Further, each item must, in so far as possible, reflect a job-related reading task.

It is immediately clear that acceptable types of question are fairly restricted. There can be no essay questions, or questions requiring short written answers. Sentence completion would be a dubious type, as would cloze procedure.

Table 1: Definitions of job-types

1.	Apprenticeship:	in which articles of apprenticeship are signed and agreed, with national regulations, block and/or day release training;
2.	Professional:	recognised training in one of the professions, eg. articled clerk in accountancy;
3.	Clerical:	general office and clerical duties which may or may not include any training;
4.	Distribution Operatives/Others:	those employed in the Distribution industry; operatives: 2+ months of on-the-job training; others: no formal training;
5.	Operatives/Others:	as above, all other industries;
6.	Induction:	applies to reading material common to all job-types, given as part of induction into the company.

If we consider the sorts of reading material we have - lists, forms, text, etc. - it is clear that, in the main, the young person is required to find one or more pieces of data, to transfer information, to order data, etc. There is little extended writing, no sudden blanks in the middle of sentences.

Multiple choice questions suggest themselves, but, although certainly representative of a large number of job-related reading tasks, they are in themselves insufficient to represent all these reading tasks.

Three main types of item have therefore been constructed. The first is the familiar five-answer one-correct multiple choice. The second is an orally-administered type, where a number of correct answers are required for the item to be correct. The third type, known as 'action-type', requires up to four selections from up to ten possible answers. It can be seen that the first type represents tasks where a simple answer is needed. The second is for tasks where written administration may be more difficult than the actual reading material - safety signs, for example - and the third type represents those tasks where several pieces of data are required for successful completion.

V Content-validation

All materials used in the construction of items were classified according to a system adopted from the work of Sticht and his associates (Sticht et al, 1972). This classification is reproduced in Table 2.

Every item was then subject to scrutiny by panels of personnel and training officers from industry and commerce to assess its content validity. Different panels were invited, according to the type of job predominant in that industry, and asked to assess each item in terms of the representative nature of the reading passage and the relevance of the question. If the reading material failed to be representative of a type commonly encountered, the item was rejected. Similarly, should the question-part of the item be rare, irrelevant or not applicable to school-leavers in their initial period, it was either modified or rejected.

All surviving items were then examined by two linguists from the staff of the Polytechnic. Their task was to note any unseen ambiguities or unnecessary complexities and to check the content classification of each item. In fact, they performed an additional task, and suggested a classification

Table 2 Definition of content-type categories

1. Tables of content and indexes:

   Content designating the location of information with a publication.

2. Standards and specifications:

   Content setting forth specific rules or tolerances which task procedures or the completed product must conform.

3. Identification and physical description:

   Content attempting to symbolically represent an object via an identifying code (stock ##, nomenclature) and/or by itemizing its distinguishing physical attributes.

4. Procedural directions:

   Content which presents a step-by-step description of how to carry out a specific job activity. Essential elements are equipment/materials/ingredients to be used, and how they are to be used, with presentation organized in a sequential step-wise fashion.

5. Procedural check points:

   Content which presents a key word or highly summarized version of what should be done in carrying out a task rather than how it should be done. This content differs from the content classified under Procedural Directions in that it assumes the user knows how to carry out the steps once reminded that the step exists and/or reminded of the decision factors which determine whether the step is required.

6. Functional description:

   Content which presents an operating (cause and effect, dependency relationships) description of some existing physical system or subsystem, or an existing administrative system or subsystem.

(By kind permission: Human Resources Research Organisation) system based on types of linguistic task. This is given in Table 3.

VI Computerized Item-Banking

Byrne, in his comprehensive review of question-banking systems, (Byrne, 1975) suggests that "Question banks are built in order to facilitate the speedy construction of a test from a given specification" (p.6). Although the number of items surviving the content validation procedures was fairly small (about 170), both the number of data per item (31) and the complexity of the Project's requirements of such items were judged sufficient to construct a question (or "item") bank; in this case, a computerized item-banking system.

Item banks can be complex or they may be very simple. There are, however, certain irreducable minima below which a computer system becomes a waste of effort. The facilities corresponding to these minima are:

i) the selection of items on the basis of given parameters;

ii) various statistical routines;

iii) the scoring and marking of tests;

iv) the updating of the content of the item bank with new data;

v) an output of present bank contents.

These five facilities are included in the present system in varying degrees of sophistication. The item bank has been designed to be "user oriented", rather than "item-banker oriented". Virtually every user requirement has been considered and most that are feasible - and a few that are quite esoteric - have been included. Over one hundred selection criteria are available with up to 120 possible values on each; there are six different item and test analyses available; scores on tests are reported as totals and eighteen subscores; individual and group performance is reported; the provision of unreliable or inconsistent commands is not allowed and error routines pinpoint and explain each error to the user; and many other facilities.

The value of this item-banking system has been made clear to us by the ease with which it allows tests to be constructed and their results analysed. The specification of require-

Table 3: Definition of linguistic tasks

1. Referential: where the item poses a task requiring the locating of correct information;

2. Regulative: where the item requires an answer in terms of a behaviour, either an act of commission or omission;

3. Attitudinal: where the question requires the description of a state of mind, or approach to a task;

4. Definitive: where the item requires a description of the meaning of a word or words.

ments leads, within a day, to a review of selected items and the assembly of the relevant master-copies for printing. The results for a test of 50 pupils can be available within three days. Further, the ability to investigate trends and relationships within the bank may be of great use to us in the production of diagnostic tests.

Enquiries have also been received from outside bodies as to the possibility of commissioning tests tailor-made to their requirements.

The data stored per item are given in Table 4.

VII Piloting of Items

All items surviving the content validation procedures were piloted amongst the target population. Over one hundred middle-ability Fifth Form pupils in a Sheffield comprehensive school took various sets of items in short tests. No group was smaller than seven nor larger than thirty. This initial pilot was to discover any obvious ambiguities in text or questions, to note frequently omitted items and those where a wrong answer was chosen more often than the correct answer. (See X below for item analysis details). Test administration procedures were also examined at this time.

After scrutiny, certain items were rejected or amended, these latter being repiloted.

VIII Construction of Tests

Using the item bank, piloted items were used to construct a test. The pilot study had suggested that the maximum number of items that could be answered in a double school period was about thirty. To construct a test wherein every content-type and every linguistic task was represented for each job-type would have meant the number of items necessarily exceeding thirty. Instead, a test (Form A) was constructed, with items for each job type, content-type and linguistic task. No job-type, however, had items for every content-type or linguistic task. Similarly, no content-type had items for every job.

Figure 1 indicates the format of Form A.

Hesitating, however, to reduce all the collected reading materials to only thirty-one questions, a long form of the test, involving the items in Form A, together with additional items (Form B), was constructed, to give wider coverage of the job, content and linguistic areas. Thus, should a clear-cut decision not be available using Form A alone, Form B might be administered and the results added together.

Figure 2 indicates the format of Form B.

IX Test Reliability

A criterion-referenced test is one which attempts to assess an individual's ability or performance in relation to some external criterion (eg. whether or not a person can drive a car is assessed at a driving test). The reliability of such a test can therefore be seen as the reliability of classification as 'master' or 'non-master' on repeated testing. The test-retest procedure is indicated as the method of determining reliability for this type of test.

The calculation of measures of reliability remains a controversial area within criterion-referenced measurement. Various non-standard statistical techniques have been suggested - and decried - (see eg. Livingstone (1972; Harris (1972)). It was decided to use standard, product-moment correlation techniques and the decision-theoretic approach adopted by other workers in this area (Swaminathan et al 1974); Close (1977); Berk and DeGangi (1979); Subkoviak (1978)), using Cohen's K (Cohen 1960) as an unbiased estimate of the degree of agreement in classification as master/nonmaster. The coefficient was calculated for each possible score, and its maximum value taken as the score at which decisions were most reliable. This score will be used as part of the predictive validity study (see XI below).

The results for Forms A and B are given in Table 5, for repeated administration to 44 middle ability students in two schools.

It can be seen that Form A is acceptably reliable, whilst Form B is not. Further analysis of the results is being undertaken.

Table 5: Reliability Coefficient

	Pearson Correlation	Cohen's K	Optimal Cutting Score	Total no.of Items
Form A	0.83	0.90	17	31
Form B	~~0.18~~ 0.54	~~0.23~~ 0.51	~~22~~ 27	30

Table 4: Data per item in Item Bank

1. Unique, Identifying Number
2. Industrial Category (B.S.I.C.)*
3. Content Type (Sticht, 1974, Classification)
4. Job Type (Careers Service Classification)
5. Physical Format of Material (Prose, Diagrams, etc)
6. Firm, ITB, College Originally Supplying Item
7. Size of Firm
8. Item Question Type (Action, Mult.Choice, etc)
9. Answering Key
10. Order of Item on the Page
11. Code of Last Test in Which Item was Used
12. Number of the Item in Last Test
13. Number of Testees who have taken the Item
14. Number of Answer Boxes (Max. = 10)
15. Number of Testees Omitting the Item
16. Number Marking First Box
17. Number Marking Second Box
18. Number Marking Third Box
19. Number Marking Fourth Box
20. Number Marking Fifth Box
21. Number Marking Sixth Box
22. Number Marking Seventh Box
23. Number Marking Eighth Box
24. Number Marking Ninth Box
25. Number Marking Tenth Box
26. Number Correctly Responding
27. Linguistic Task Type
28. Item Number of Item replacing this one
29. Index of Discrepancies in Test-Retest

* British Standard Industrial Classification, Central Statistical Office HMSO (1968)

Figure 1 Number of items by job and content type: Form A

	Apprentice	Clerical	Distribution	Op/other	Induction	Totals
Standards & specifications	1			1	3	5
Identification & Physical description	4	5		1		10
Procedural Directions	3		3			6
Procedural Checkpoints	3		1			4
Functional Description	3				3	6
Totals	14	5	4	2	6	31

Figure 2 Number of items by job and content type: Form B

	Apprentice	Clerical	Distribution	Op/other	Induction	Totals
Standards & specifications		2	4	6		12
Identification & Physical description	3	1	6	3	1	14
Procedural Directions					3	3
Procedural Checkpoints						
Functional Description		1				1
Totals	3	4	10	9	4	30

X Item Analysis

Following on from this brief discussion of problems in criterion-referenced tests, it is clear that many of the other statistics used in classical item analysis are also of limited value for the same reason: problems arising from restricted variability as between scores.

The commonest measures, difficulty and discrimination, are, in fact, of some use in criterion-referenced measurement, but not in the same sense as in classical test theory. Item difficulty is interesting but not of overwhelming importance. A test constructor may require a mixture of easy and hard items, but more often in criterion-referenced measurement he requires a test of a certain area. If the test has content validity, the difficulty of the items is secondary. Discrimination indices for an item are not valuable in terms of seeing how the item contributes to spreading out scores on the test, but to see whether an item positively or negatively discriminates. The latter is as much an indicator of a poor item in criterion-referenced measurement as in norm-referenced measurement (eg. Smith, 1974).

XI Predictive Validation

The ultimate value of a test of functional skill is its ability to predict adequate performance. In this case, the test of occupational functional reading ability should predict performance on job-related reading tasks in the first six weeks of employment or training for a job.

As an initial measure of this predictive validity, about 150 fifth-form school-leavers, mainly of average ability, were given Form A of the test, and the Edinburgh Reading Test, Form 4. This latter is a norm-referenced test for comparison purposes; extrapolated quotients were used for those pupils already 16 years old.

These pupils have now left school and have found or are seeking employment. The Sheffield Careers Service has kindly allowed access to its files, so that each subject may be followed up when in employment. Each employer is being asked to allow the young person's immediate supervisor to fill in a rating form, assessing relevant aspects of job-performance on tasks involving and not involving reading. These include perseverence, need for oral guidance, speed,etc.

Test scores and ratings will then be related via multiple and linear regression analyses, to establish the test's predictive validity.

XII Summary

The Sheffield Occupational Functional Reading Project has as its remit the development of a criterion-referenced test of occupational functional reading ability of pupils, leaving at 16+ years of age, in their last year at Sheffield secondary schools. Related diagnostic tests are also to be constructed.

Progress towards these aims has taken the form of identification and classification of job-related reading tasks and collection of relevant reading materials. These have been used to construct test items, the data about which have been banked by computer. Content validity was established by the use of panels of experts. Test items were piloted and a short and long form of a prototype test constructed. This test has been assessed for reliability and then used to test a large group of school-leavers, prior to their leaving school. These testees are now being followed up at work and their employers asked to rate them on job-performance scales. The predictive validity of the test will be established by relating test score to performance at work.

References

Berk R A & De Gangi G A	Technical Considerations in the Evaluation of Pediatric Motor Scales *American Journal of Occupational Therapy* 1979, 33, No.4 pp.240-244
Byrne C J	*Computerized Question Banking Systems* Institute of Educational Technology, The Open University, Milton Keynes 1975
Close J S	Personal communication
Cohen J	A coefficient of agreement for nominal scales *Educational & Psychological Measurement*, 1960, 20, 37-46
Department of Education and Science	*Education in Schools - A Consultative Document* HMSO Cmnd.6869 1977
Harris C W	An interpretation of Livingstone's reliability co-efficient for criterion-referenced tests. *Journal of Educational Measurement*, 1972, 9, 27-29
Latham W & Parry O	Functional Reading and the Schools in *Growth in Reading* ed. Thackray D V London, Ward Lock Educational, 1979
Livingstone S A	Criterion-referenced applications of classical test theory *Journal of Educational Measurement* 1972, 9, 13-25
Smith C W	Criterion-referenced Assessment in *Contemporary Issues in Educational Testing* ed. Crombag, H F & De Gruyter, D N The Hague, Mouton, 1974, pp 138-147
Sticht T G, Caylor J S, Kern R P and Fox L C	Project REALISTIC: determination of adult functional literacy skill levels *Reading Research Quarterly*, 1972, 7, 424-465
Subkoviak M J	Empirical Investigation of Procedures for Estimating Reliability for Mastery Tests *Journal of Educational Measurement* 1978, 15, No.2, pp 111-116

Swaminathan H, Hambleton R K & Algina J	Reliability of Criterion-Referenced Tests: a decision-theoretic formulation _Journal of Educational Measurement_, 1974, _11_, 263-7

Journal of Research in Reading *Vol 3 No 2 1980*

Functional reading and the schools: a progress report

W. Latham and O. Parry, *Sheffield Polytechnic.*

ABSTRACT

A progress report of the work of the Sheffield Occupational Functional Reading Project is made. A résumé of previously reported work is given, including the surveying of the job-related reading tasks facing 16+ year old school-leavers in their initial period of employment or training. The development of test items is explained and examples given, with a brief report of the computerised item-banking system used. Content validation procedures are outlined and the piloting of test items reported. The construction of two criterion-referenced tests of occupational functional reading ability is given in some detail and estimates of reliability for these tests is reported. Progress of work to establish the predictive validity of these tests is included, with indications of how this work leads into diagnosis of individual problems.

RÉSUMÉ

La lecture fonctionnelle et les écoles: état périodique.

Le but du *Sheffield Occupational Functional Reading Project* (projet de la lecture prefessionnelle fonctionnelle) est d'abord l'identification et la classification des tâches de lecture que rencontrent des jeunes quand, à l'âge de seize ans, l'enseignement scolaire termine et ils commencent à travailler ou entrent en apprentissage. Cette démarche est suivie par la construction d'un test de groupe, indiquant les critères adoptés, censé mesurer l'aptitude des individus à ces tâches, et par le développement des épreuves diagnotiques relatives.

L'échantillon comprenait des firmes de Sheffield, choisies au hasard, représentatives au point de vue d'industrie et de grandeur, d'autant de sortes de travail que possible. On les a visitées et noté les tâches de lecture portant sur le travail, et on a ramené des exemples des matériaux écrits et de ces firmes-ci et d'autres établissements d'entraînement.

Ces matériaux ont été classés suivant l'emploi, l'industrie, la grandeur de la firme etc, et les données inclues dans un programme d'ordinateur établi pour le Projet. On a adopté aussi une méthode de classification pour le contenu du matériel écrit, développée aux Etats-Unis.

Au moyen de ces matériaux, on a construit des articles pour le test de groupe. Tous les articles ont été soumis à un examen rigoureux par des employeurs et des linguistes pour en affirmer la validité de contenu et la nature représentative des matériaux.

Tous les articles qui résistaient à cet examen ont été mis au point dans un école à Sheffield, parmi des élèves de quinze ou seize ans de dernière année d'étude. A ce stade, des détails d'ordre administratif ont été dressés aussi. Par suite de cet exercice, plusieurs articles ont du subir des modifications, ou la suppression, et être remis au point.

A partir de ces articles, deux tests ont été générés. Formule A est un test indépendant qui consiste de trente-et-un articles. Formule B est une série de questions supplémentaires qu'on peut ajouter à A pour mieux recouvrir des aspects divers de l'emploi, du contenu et des éléments linguistiques.

Un processus de test-retest, appliqué à deux écoles, a dégagé des évaluations de la regularité des épreuves. On a fait l'analyse des résultats par des méthodes d'étalonnage de corrélation produit-moment et de la statistique décision-théorique, K. Formule A s'est montrée suffisamment fidèle, mais pas Formule B, pour laquelle d'autres recherches sont nécessaires.

Utilisant Formule A et Formule 4 de l'*Edinburgh Reading Test* (en tant que comparaison) on a testé à peu près 150 élèves de 15 ou 16 ans, provenant de quatre écoles. Maintenant, on poursuit ces élèves en demandant à leurs employeurs de faire un rapport pour évaluer leur performance. La relation entre cette mesure et les scores du test sera la validité prophétique des deux tests dans ce domaine-ci.

On utilisera des analyses détaillées des performances pour établir des tests diagnostiques relatifs.

INTRODUCTION

Functional reading—reading related to a task or role—is encountered by all of us, every day. Whether it be in our job or in our leisure, as a consumer or as a citizen, there is constant recourse to the written word. Yet little study has been made of any aspect of functional reading in this country. Williams (1976) has made some valuable inroads into the world of the consumer, and Moyle has looked at basic literacy for adults (1978). Working life has been the subject of some neglect, however, particularly at the level of the school-leaver. It is of little surprise then that James Callaghan, in his Ruskin College speech, 1976, should suggest a wide gulf between education and work, when so little is known in such a major area as the reading required of young people at work.

The Green Paper, *Education in Schools—A Consultative Document* (HMSO 1977) suggests: 'It is not the task of schools to prepare pupils for specific jobs but experience has long shown that studies and activities which are practical and obviously relevant to working life can be valuable as a means of learning, including the learning of basic skills' (p. 11). Little can be done by schools, however, in the area of occupational functional reading whilst reading tasks faced by the school leaver have yet to be identified.

The purpose of the *Sheffield Occupational Functional Reading Project* (SOFRP) is firstly the identification and classification of reading tasks encountered by a 16 year old school-leaver upon entering work or training for a job. Following on from this is the construction of a criterion-referenced group test to assess the abilities of individuals in relation to these reading tasks and the development of related diagnostic tests.

The work in the first nine months of the Project has been previously published (Latham and Parry, 1979) but a brief résumé is given below.

In order to identify the reading texts which might be faced by 16 year old school leavers in the first six weeks of their employment, a large number of firms in Sheffield were visited. The firms were selected, by a random process, from groups of firms representing the main employment areas entered by 16+ school leavers. Each firm gave details of the types of jobs into which leavers were recruited, details of training (where applicable) and induction procedures. Job related reading materials were inspected and specimens collected. Samples of reading were also obtained from a Job Centre and from FE Colleges that new employees attended as a condition of their employment.

The reading materials obtained were used to construct test items (an item consisting of a passage to be read and *one* question related to it). A passage might be prose, a form to be filled in, etc. Three types of item were constructed: the majority were 5-answer, multiple-choice questions. Examples of three such items, all relating to the same passage, are given in Figure 1. The second and third types required more than one answer to be marked as correct, the former being administered orally. In these types, up to four answers out of up to ten possibilities might be required. A small number of items, including a filing exercise, required that one answer involving several steps be given. Examples of these latter types are given in Figures 2 and 3 respectively.

HYGIENE

It is encumbent upon all staff whose work takes them into kitchens, serveries, storerooms, dining rooms, or ancillary premises to observe the tenets of personal hygiene:-

a) To keep clean the hands, especially the finger-nails and the exposed parts of the arms.
b) To wash the hands frequently and always after using the 'conveniences'.
c) To keep covered any cut, wound, abrasion or sore by an approved waterproof dressing.
d) To report to the supervisor any stomach disorder, skin complaint, ear discharge or other illness or infection. If in doubt, to play safe and report.
e) To keep clean all protective clothing, including suitable head-coverings, the wearing of which is obligatory upon all persons working with food in this organisation.
f) 'Thou Shalt Not Smoke' except in those areas designated for that purpose.
g) Not to wear visible personal adornments when working with or serving foodstuffs—wedding rings are excepted.

1. What should you do if you cut yourself?

- ☐ A. Wash your hands
- ☐ B. Cover the cut with an appropriate dressing
- ☐ C. Cover the cut with your handkerchief
- ☐ D. Leave it to stop bleeding of its own accord
- ☐ E. None of these

2. What should you do if you have a skin rash?

- ☐ A. Go straight to the doctor.
- ☐ B. Carry on working normally
- ☐ C. Stay away from work
- ☐ D. Report it to the supervisor
- ☐ E. Cover up the rash

3. Which of these statements is incorrect?

- ☐ A. You cannot wear earrings at work
- ☐ B. You must wear a head-covering
- ☐ C. You must keep your nails clean
- ☐ D. You can smoke almost anywhere
- ☐ E. You must report ear discharges

Figure 1: *Example of three multiple choice items.*

Put a cross (X) in the box next to each statement below which is *wrong*.

The first is done for you.

- ☒ You can use defective tools
- ☐ You should keep your bench tidy
- ☐ Anyone can use moving machinery
- ☐ Chisels with mushroom heads are good tools
- ☐ Handles are needed for scrapers
- ☐ Vices should only be used in emergencies

Defective Tools

Do not use defective tools as they can cause severe injury—particularly to the hand and eye. Make sure that handles are fitted on files, hammers, scrapers and screwdrivers, and that they are not split or insecure.

Ill-fitting spanners, hammers with chipped heads and chisels or drifts with 'mushroomed' heads must not be used.

Bench Work

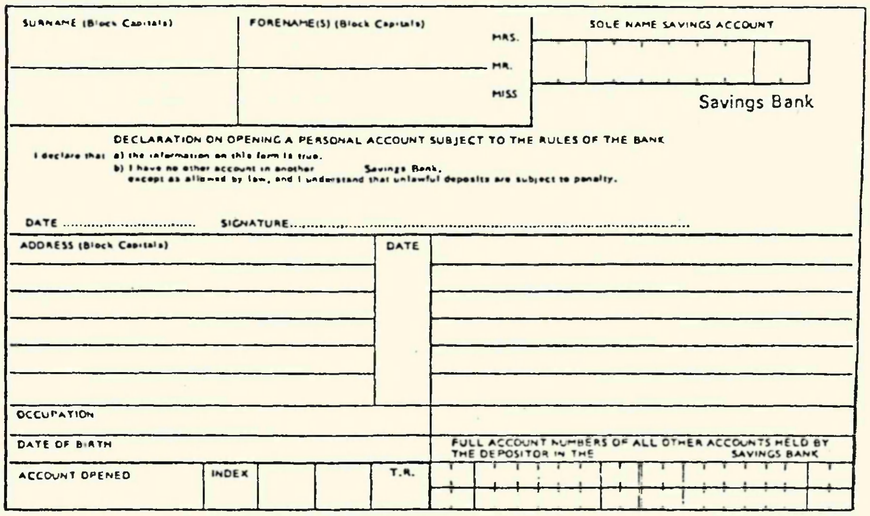
Keep your bench tidy. Tools laid aside should be put in a safe place where they cannot be dislodged or fall. When you have finished using them put them away tidily and in a serviceable condition.

Where necessary, clamp your work in a vice or use some other effective method. A vice should grip the workpiece firmly so that it will not fall out or slip when you are working on it. Don't use worn vice jaws.

Work on Moving Machinery

Only specially authorised people (Machinery Attendants) may work near or on unguarded moving machinery and then only under circumstances specified by Regulations made under factory law.

Figure 2: *Example of an 'action-type' item.*


SURNAME (Block Capitals)
FORENAME(S) (Block Capitals)
MRS.
MR.
MISS
SOLE NAME SAVINGS ACCOUNT
Savings Bank
DECLARATION ON OPENING A PERSONAL ACCOUNT SUBJECT TO THE RULES OF THE BANK
I declare that a) the information on this form is true.
b) I have no other account in another Savings Bank, except as allowed by law, and I understand that unlawful deposits are subject to penalty.
DATE .... SIGNATURE ....
ADDRESS (Block Capitals)
DATE
OCCUPATION
DATE OF BIRTH
FULL ACCOUNT NUMBERS OF ALL OTHER ACCOUNTS HELD BY THE DEPOSITOR IN THE SAVINGS BANK
ACCOUNT OPENED
INDEX
T.R.

Fill in the Savings Bank form using the following details:

Mr John Sidney Barlow
63 Middle View Road
Uptown-on-Sea
Berks

Account number 02 967310 MB

Today's date is 22nd June

Mr Barlow is an accountant

Please write *your* name clearly in the space for the signature.

Figure 3: *A further 'action-type' item.*

These test items were classified according to the type of job in which they were encountered, the type of industry and a content-classification system adopted from the work of Sticht and associates for the US Army (Sticht *et al*, 1972). This latter system includes such categories as 'procedural directions', where the content is such that it details each step in a given procedure. The small sticker often seen on supermarket tills, instructing on the use of a £50 cheque-card, is a good example. (Figure 4).

The information available on each item, its job-type, content-type, etc., was stored by computer in the initial stages of the development of the Project's computerised question-banking system.

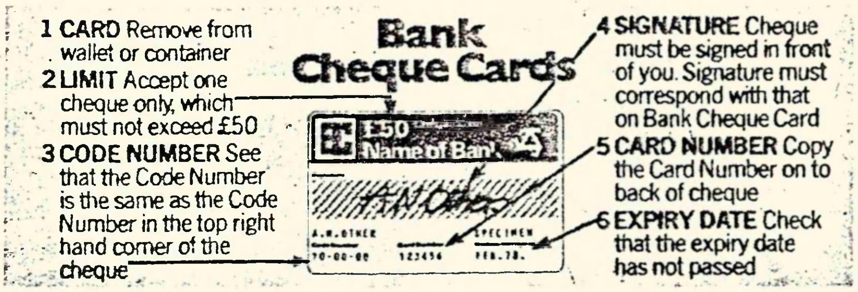


Figure 4: *Example of content-type 'Procedural Directions'.*

CONTENT VALIDATION

The work following on from the last report started with the assessment of the content validity of the test items, i.e. the extent to which they truly represent typical occupational reading tasks. The selection of material around which to construct items and the questions themselves are largely judgemental processes and it is only by the use of external experts that one can attempt to ensure the validity of the items.

Employer Panels

The scrutiny of the test items required a return to source. Panels of personnel and training officers were invited to scrutinise all the available items for the type of job into which they most often recruited school-leavers (e.g. apprentices, operatives) and all the panel members viewed the items related to company induction procedures. Panel members were drawn from the fields of steel and tool manufacture, engineering, distribution, catering and other major employers in the Sheffield area. The purpose of each panel was to consider each item and to answer two questions: 'Is it based upon a piece of reading material representative of material encountered in my industry?' and 'Is the question posed or task involved a relevant, common and important one?'

From a pool of about 250 items, many were rejected outright and others were revised in the light of the panels' comments.

Linguist Panel

Two linguists, members of the staff of Sheffield Polytechnic, were asked to undertake a second scrutiny of all the items, to identify any ambiguities, check the content classifications and to give their opinion on the question of whether undue complexity had been introduced at any stage. In fact, they exceeded these requests by suggesting a further classification system in addition to the content system, based upon linguistic tasks.

Four categories were suggested:

i) *Referential:* where the item poses a task requiring the locating of correct information.

ii) *Regulative:* where the item requires an answer in terms of a behaviour, either an act of commission or omission.

iii) *Attitudinal:* where the question requires the description of a state of mind, or approach to a task.

iv) *Definitive:* where the item requires a description of the meaning of a word or words.

Other changes were also effected in the wording of questions and the correct content-classification of some passages.

PILOTING OF ITEMS

Despite all the processes described above, it would be unwise to merely construct a test and use it. Each item must be carefully checked again using a piloting technique, giving it to a number of appropriate subjects, drawn from the target population. The piloting was undertaken in a Sheffield comprehensive school, using groups of pupils taken from the middle-ability range (predominantly taking CSE examinations rather than GCE). Every item was taken by at least seven pupils on the first occasion and results checked for distractors with high answer rates and items consistently omitted, whilst the groups were also used to test out various administration procedures. On review, several items were deleted, distractors changed and whole items re-written. Changed items were then repiloted.

It was considered that the usual forms of item analysis, such as difficulty and discrimination indices, were inappropriate *per se* in the development of a criterion-referenced test. Here, the ultimate emphasis is upon assessing each individual child's ability in relation to the test materials, rather than the more common norm-referenced (or standardised) test where a pupil is assessed in relation to his peers. Such analysis as is appropriate is considered in a later section.

FORMAT OF THE LATEST VERSIONS

Physical Format

The test has been designed to last for one double-school-period, about 70 minutes, to include all administration and working periods. From piloting, it was found the maximum number of items which can be answered by pupils in this period is about 30. Items are contained in two sections: the former consists of several 'action-type' items with the appropriate administration sheet; the latter of 'multiple-choice' items. Timing is arranged such that all testees have an opportunity to answer all items.

Short and Long Forms

It is possible to select thirty items which do cover all aspects of the content-types, job-types and linguistic tasks, but not as completely as the panels of employer experts have suggested would be representative of their industry. A short form (Form A) and a long form (Form A plus Form B) of the test have been constructed, the former to cover all aspects as stated above, the long form to be second test to cover more completely each area. Both contain about 30 items and are timed to last one double-school-period. It is hoped that the short-form alone will be shown to be sufficiently valid in the predictive validity study.

RELIABILITY OF THE TEST

Each of the two forms, Form A (the short form) and Form B (additive second half of the long form) were given separately to groups of the target population. Form A was administered on two occasions to forty-five pupils in the original pilot school, and Form B was given on two occasions to forty-four pupils in a second pilot school.

A great deal has been written about the reliability in criterion-referenced measurement, with, as yet, little agreement as to acceptable measures, (e.g. Livingstone (1972), Harris (1972)). What is certain, however, is that criterion-referenced tests are designed to assist in decision-making and that a reliable test here is one which is consistent in its recommendation. The test and retest procedure is clearly the appropriate method, whilst there remains disagreement on the appropriate form of subsequent analysis. The picture is complicated by the fact that if everyone taking the test scores exactly the same, it is not necessarily a bad test. It may equally be that those subjects are masters of the area in which the test assesses. Normal correlation methods, however, would produce some very strange results in a calculation of reliability on such figures. Other methods are rare and often under-supported by theory. One statistic that does lend itself, however, is Cohen's K (Cohen, 1960; Swaminathan, Hambleton and Algina, 1974), an estimate of the agreement in decision-making on a test and retest.

We report, therefore, the whole-test correlation coefficient for repeated administrations and the optimal value of K. These are, for Form A: 0.83 and 0.90 for a cutting score of 16 out of 31; Form B: 0.54 and 0.51 for a cutting score of 27 out of 30.

It is clear that the Form B has lower reliability and its items are at present undergoing further analysis to discover the nature of the relationships within the test.

SWEEP TESTING AND PREDICTIVE VALIDATION

The criterion-referenced test under development here is for use in identifying which pupils will have difficulty with job-related reading tasks upon leaving school. It is therefore necessary to obtain an estimation of how well a pupil's score on the test can predict his or her performance at work on tasks involving reading, i.e. an assessment of the test's predictive validity.

It was initially proposed to test up to 500 16 year old pupils in their last year at school in the Spring Term 1979. The combination of industrial action, oil shortages and extreme weather conditions made this, unfortunately, impossible. A two-stage sample was therefore devised: the city was divided into six major areas on the basis of the number of schools and the prevalent type of industry. Eight schools were selected on a random basis from within them. They were asked to provide two groups of about 25 pupils each, according to specified ability ranges (randomised between the schools). Those able and willing to assist in the opening weeks of the Summer Term had their pupils tested immediately; the remainder assured their help for Spring 1980, the second phase of testing.

163 pupils were administered the *Functional Reading Test, Form A (FRT A)* and the *Edinburgh Reading Test, Form 4 (ERT 4)* for comparison. Not all groups overlapped for the two tests and a pair of scores is available on about three-quarters of the pupils.

The product-moment correlation between the *FRT A* raw score and the *ERT 4* quotient was 0.81 (N = 123), but earlier comments on the use of correlational techniques with criterion-referenced test scores must also be borne in mind here.

The Sheffield Careers Service has continued its invaluable support for the Project by allowing access to its files on school-leavers, so that each pupil may be located when in work.

Rating scales of performance on tasks involving reading and on those not involving reading are under development at present. These scales cover the areas of perseverence, speed, requests for help and overall performance.

The immediate supervisor of each of the testees in employment will be asked to rate the performance of the testee on each of these scales. These ratings will then be related to the test scores to provide an estimate of the ability of the two tests tasks at work.

PREDICTION INTO DIAGNOSIS

Provided that the work mentioned above results in an acceptable level of predictive validity, it is proposed to produce some specific diagnostic tests.

This would proceed with a detailed analysis of the group test items which have caused difficulty to individuals. The analysis will initially be concerned with possible problems associated with:

i) the format of the reading materials,
ii) vocabulary or aspects of the relevant semantic field,
iii) the sentence or phrase structure of the material,
iv) the general nature of the linguistic task posed by the material,
v) the overall nature of the content of the material.

The panel of linguists who provided help earlier in the project will be asked to comment on this pattern of analysis and the approach to using it.

The second phase of sweep testing will provide extra data both for a second predictive validation, but also, wider ranges of scores to assist in the diagnostic test development.

NOTE

An earlier version of the paper was presented at the 16th Annual Course and Conference of the United Kingdom Reading Association, Leeds, July 1979. The project is being undertaken in collaboration with the Sheffield Local Education Authority.

REFERENCES

COHEN, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20,* 37-46.

HARRIS, C.W. (1972) An interpretation of Livingstone's reliability coefficient for criterion-referenced tests. *Journal of Educational Measurement, 9* (1), 27-29.

LATHAM, D.W. and PARRY, O.D. (1979) Functional reading and the schools. In D. Thackray (ed.) *Growth in Reading*. London: Ward Lock, 197-208.

LIVINGSTONE, S.A. (1972) Criterion-referenced applications of classical test theory. *Journal of Educational Measurement, 9* (1), 13-25.

MOYLE, D. (1978) *Perspectives on Adult Literacy*. Ormskirk: UKRA (Monograph series).

STICHT, T.G., CAYLOR, J.S., KERN, R.P. and FOX, L.C. (1972) Project REALISTIC: determination of adult functional literacy skill levels. *Reading Research Quarterly, 7,* 424-465.

SWAMINATHAN, H., HAMBLETON, R.K. and ALGINA, J. (1974) Reliability of criterion-referenced tests: a decision-theoretic formulation. *Journal of Educational Measurement, 11* (4) 263-267.

WILLIAMS, A. (1976) *Reading and the Consumer*. London: Hodder Stoughton.

APPENDIX XI

COURSES OF STUDY UNDERTAKEN BY THE CANDIDATE

Courses of Study undertaken by the Candidate

1. Courses in

   The Teaching of Reading

   Linguistics and Reading

   Assessment in Reading

   Research Methods & Statistics

   given as part of the Diploma in the Teaching of Reading (University of Sheffield) were attended between September and December 1977.

2. Courses in

   FORTRAN programming

   Further FORTRAN

   Use of Statistical Packages

   as part of the Courses for Staff given by the Department of Computer Services, SCP, were attended in September 1978, April and May 1979 respectively.