Trainability testing

A practical approach to selection

Sylvia Downs

A TSA Publication
Trainability testing

A practical approach to selection

Sylvia Downs

*Industrial Training Research Unit,
Cambridge*

London: Her Majesty’s Stationery Office
The Training Services Agency has taken over from the Department of Employment responsibility for producing specialist publications designed to encourage and speed up the introduction of new ideas, methods and techniques into the training process. These publications include the Training Information Paper series, intended to encourage the application of research findings, or of innovative instructional methods, to the task of solving practical training problems. The aim is to present—as concisely as possible—the results of a particular research project, or a group of related projects, in a way that will enable them to be assimilated by anyone actively involved in the training process.

Authors themselves are entirely responsible for the contents of these booklets; they may contain speculations and conclusions not necessarily supported by the Training Services Agency.

In all cases, however, the Agency believes these booklets will provide valuable topics for discussion, and their usefulness will be considerably enhanced if they are included in seminars or during courses for training officers and managers.

Enquiries relating to the research on which these papers are based, or other related matters, should be addressed to:

Richard Smith
Head of Training Information
Training Services Agency
162-168 Regent Street
London W1R 6DE
<table>
<thead>
<tr>
<th>Chapter I: Trainability testing</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Summary of trainability test procedure</td>
<td>2</td>
</tr>
<tr>
<td>The research background</td>
<td>3</td>
</tr>
<tr>
<td>— Selecting older workers</td>
<td>3</td>
</tr>
<tr>
<td>— The clothing industry</td>
<td>4</td>
</tr>
<tr>
<td>— Other industries</td>
<td>5</td>
</tr>
<tr>
<td>Applicability of trainability tests</td>
<td>5</td>
</tr>
<tr>
<td>Advantages and disadvantages of trainability tests</td>
<td>8</td>
</tr>
</tbody>
</table>

| Chapter II: Constructing, validating and administering a trainability test | 10 |
| Designing a trainability test      | 10 |
| — Summary of the procedure         | 10 |
| — Analysing the job                | 11 |
| — Selecting the work-piece or task | 11 |
| — Writing the error check-list     | 12 |
| — Ratings                         | 13 |
| — Instructor’s script              | 13 |
| Running a pilot study              | 15 |
| — Training of instructors          | 15 |
| — Comparing test results with other measures of success | 18 |
| — Making adjustments in the test   | 20 |
| The trainability test in use       | 22 |

| Chapter III: Summing up           | 24 |
| Appendix 1: References           | 26 |

**Appendix 2:** Job examples: their essential components and tasks set for trainability tests  
**Appendix 3:** Statistical validation of test results
Acknowledgements

This booklet on Trainability Testing would not have been possible without the interest and help given by many individuals.

I would particularly like to thank my colleagues at the Industrial Training Research Unit, especially Carol Smith, Michael Smith and Anne Warren, who have analysed and conducted much of the research; and Eunice Belbin, Director of the Unit, for her inexhaustible help and guidance.

Mr Sagar, the Group Training Manager of Pasolds Ltd, and Mr Carter, the General Manager of Fork Truck Training, unsparingly gave their co-operation and encouragement.

Mr Logan Colbeck, of the Knitting, Lace and Net Industry Training Board, kindly allowed us to quote from his own researches and also provided us with valuable notes about his experience in designing new tests and training instructors.

Many of the Industrial Training Boards co-operated closely, especially the Clothing and Allied Products, Construction, Road Transport and Shipbuilding Boards; and the Training Services Agency gave generous support and help.

Finally, thank you to Janet Toye and Joy Larcom who assembled and distilled many research reports into the content of this booklet.

Sylvia Downs
Introduction

Employers have for many years appreciated in principle the advantages of using good selection methods. They have recognised the benefits of choosing the right person for the job, thereby saving the time and resources which might otherwise have gone into training or employing someone unsuitable.

But there has often been dissatisfaction with the methods actually available. Sophisticated selection procedures, sometimes elaborate, time-consuming and costly, have often had disappointing results. In some cases the employer feels he would be no worse off returning to old-fashioned rule-of-thumb methods.

The scientific approach to the problem has centred on the assessment of intelligence plus special aptitudes. Skills are interpreted as being the potential product of particular clusters or combinations of aptitudes; such as spatial ability, manual dexterity, mechanical ability, and general reasoning ability. The procedure is to work back from the skill one is interested in to the basic aptitudes and level of intelligence that are needed to perform it. This approach, however, is cumbersome and also very difficult to carry out.

It is not easy, even in theory, to define the various aptitudes required for a job. Where results have been properly evaluated there has often proved to be little or no correspondence between success in tests designed to measure the aptitudes in question and success at doing the job itself. The apparent irrelevance of the tests, moreover, may be off-putting to applicants for a job and distasteful to the work force and organisation.

Tests based on a job sample clearly overcome some of the limitations of this classical approach. Selecting the job sample is a more straightforward business than theorising about the aptitudes involved, and the test is something which can be devised on the spot with regard to the particular circumstances of the case.

Compared with aptitude tests, there is much less risk of creating a test which has no bearing on the applicant's ability to perform the whole job, and it makes more sense to most applicants to carry out a task which is intrinsic to the job.

Nevertheless, if a person is asked to carry out a job sample without prior information about how it is to be performed, the test may not give him a full opportunity to show his potential. A person is not normally expected to perform adequately without prior instruction.
Moreover, in the majority of cases, instruction, in other training, will be provided if the person is accepted. There is an important advantage to be gained, therefore, from including in one’s assessment of an applicant’s suitability, an assessment of his ability to learn the job in question. Let us call this ability to learn, his ‘trainability’.

The trainability test has been devised to take into account those factors otherwise ignored by an ordinary job sample test. Its objective is to assess whether the applicant has the potential to reach a satisfactory standard of performance after training. This is done by asking the applicant to perform a specific task in which he has been given prior instruction.

A note on terminology

The Industrial Training Research Unit (ITRU) has usually referred to the procedure described in this booklet as ‘Trainability Assessments’. This usage developed mainly out of the desire to impress upon applicants that what they were being asked to do was not a test in the conventional sense; the term ‘assessment’ seems more reassuring. There has, however, been increasing dissatisfaction with this terminology as a blanket description of what is involved: the process of assessment in reality is only one part of the whole exercise and needs to be referred to specifically as such; the word ‘test’ is in fact a more accurate description of the whole exercise. For this reason the term ‘trainability test’ has been used for the purpose of this paper. It seems reasonable to suppose that people who take the test will not be put off by the use of the word, provided the nature of the exercise is properly explained to them.

Summary of trainability test procedure

The test is administered by a specially trained instructor.

1. Using a standardised form of instruction and demonstration, the instructor teaches the applicant the test task, chosen because it incorporates essential elements of the work for which selection is being made. During the teaching period the applicant is free to ask questions.

2. The applicant is asked to perform the task unaided.

3. The instructor evaluates the applicant’s performance by noting all errors on a standardised error check-list and by making a rating of the applicant’s likely performance in training, usually on a five-point scale.

4. The applicant’s test result is compared with expected results for successful applicants; if his result falls below the cut-off point decided upon in the light of previously validated results, he is rejected; if he comes above it he is accepted for training.

The illustrated pages in the centre of this book show a sample trainability test, in this case making a three-sided bag as a test of
ability to an overlocking sewing machine. Included in it are the form of words the instructor should use to introduce the applicant to the task, the instructions he uses when teaching the task, and the form on which he records errors and makes his overall rating of the applicant.

The research background

Selecting older workers

Early work with the trainability test was carried out in the late 1960's. The particular problem to be met was that of selecting older workers for retraining. In the past, few training schemes had accepted recruits aged over 35 years, but as a result of redundancies created through technological change increasing numbers of older people would need re-training. Would it be possible to identify the people who would be unlikely to be able to learn a specific skill?

Standard selection techniques had a number of disadvantages for this group. Employment history had limited usefulness when the applicant's previous experience was entirely different from the skill he wanted to learn. For instance, a former agricultural worker might wish to go on a welding course. In these circumstances, it was especially difficult to predict the ability to learn a new skill from a straightforward interview. Nor were there any standard tests of aptitude or intelligence which had been devised for older people. In practice, the decision was often made in the course of a probationary training period. In the Government Training Centres (now called Skillcentres), for instance, a three-week probationary period was allowed. But the rejection of an applicant after initial acceptance is likely to be particularly distressing for the individual concerned, and it is also wasteful of time and resources. It is far better to make an adverse decision at the interview stage if this is possible.

The Industrial Training Research Unit surveyed possible selection techniques which might be employed in this situation and found a useful lead in a study of manual sorters in the London Postal School. Here it had been found that performance on a learning task could be, for mature adults, a useful predictor of success in training. A test of similar nature was devised for two manual skills, carpentry and welding, and a pilot study was carried out in nine Government Training Centres where courses in these skills were being run (Ref. 1).

The carpentry test involved making a simple joint; the welding test was a welding exercise making straight runs along chalk lines on mild steel. Each new recruit to the course in question was given a demonstration of the task he was to carry out and during that time he was allowed to ask questions. He was then asked to carry out the task and the instructor made a note of his errors and made a rating on a six-point scale of his likely performance in training. At the end of the six-month course each recruit's performance in training was compared with the trainability test error and rating scores. Trainees included both younger men (under 35) and older men (35 or more).
The result was that almost all trainees who did well on the trainability test did well on the course. But of those who did badly on the test there was a clear distinction according to age. Nearly half the young men who did badly on the test in fact succeeded on the course, but of the older men who did badly on the test (a total of 23) only three were successful on the course.

It was concluded that by using this type of trainability test most of the older recruits who would fail in the course of training could be identified in advance.

The clothing industry

Another early trial of the method was conducted in four factories making children's clothing (Ref. 2).

The company's existing method of selection was by interview, together with a test of manual dexterity and a performance test. The manual dexterity test involved picking up short metal rods and placing them, as quickly as possible, in holes in a board. In the performance test the applicant was asked to select and arrange wooden pieces to make outline shapes. The management lacked confidence in this method and thought it was particularly inappropriate for older trainees whom, among others, they wished to recruit. A trainability test was designed as an alternative means of selection. The essential components of the skill to be assessed were cloth-handling and hand-eye-foot co-ordination.

The task of machining three seams between two pieces of cloth to make a small open bag was chosen as a means of testing this skill (see centre pages). Instructors from the company's four factories were trained to administer the test and it was given to recruits of all ages who, for the duration of the experiment, continued to be selected in each factory by the usual methods.

The trainability test was evaluated by comparing each trainee's score on the test with the training officer's rating of the trainee. This was made either when she left the training school because it was decided she could not be trained, or when she reached the standard of performance at which she could produce 70 per cent of the output expected from an experienced machinist. The statistical correlation between the trainability test score and the training officer's final rating was 0.6 for 82 trainees. This was a statistically significant relationship which would occur by chance less often than one in 100 times. The manual dexterity test and the performance test, however, did not correlate either with the training officer's final rating of the trainee or with the trainability test score. The trainability test was thus shown to be a relatively efficient means of selection: it gave a clear indication of those who would or would not prove successful and also picked out the marginal trainees.

The original tests were designed for overlock machines. Subsequent research in which trainability tests were constructed for lock stitch machines gave similar results. More recently the Knitting, Lace and
Net Indu-Training Board became interested in the trainability test after unsatisfactory results with manual dexterity, performance and perception tests in a number of companies, and, in 1973, 50 companies were involved by them in a project to test more widely the usefulness of the trainability test (Ref. 4). A series of instructor training courses was run and instructors were subsequently visited to ensure competence in test administration and to amend the tests, where necessary, to suit the particular machines in use at the factory. There were some difficulties about carrying out validation because a complete set of performance data suitable for analysis could not be obtained from all factories. Nevertheless there did appear to be a direct relationship between performance on the shop floor and good test scores; and in the interval the Knitting, Lace and Net Industry Training Board has introduced trainability tests to at least a hundred more firms.

Other industries
Trainability tests have been devised and implemented with equal success in a variety of other industries. A trainability test for the selection of fork lift truck drivers was carried out at three training establishments and validated by comparison with the competence test recommended by the Fork Truck Training Joint Committee of Industrial Training Boards.

A large electronics firm consulted the Industrial Training Research Unit about a trainability test for electronic assemblers because the selection test they were using (a memory and speed test) seemed to bear little relationship to the work done in the factory, and because there were complaints from supervisors on the shop floor about the competence of new staff (Ref. 6). The trainability test scores were found useful for predicting end of training attitude, quality of performance and on-the-job speed of performance.

In the case of dentistry, 'A' level results have been found unpredictive of the operative skills involved and alternative selection instruments are required.

A pilot study carried out at the University of Bristol Dental School has led to the conclusion that a trainability test given to assess the manual skill required for dentistry should be tried out on a national scale (Ref. 7).

Other jobs for which trainability tests have been designed and validated include electrical wiring, brick-laying and bottling.

Applicability of trainability tests
The situation of older people applying for a job new to them highlighted the irrelevance of many standard selection methods. Older people, long since free of school and the examination system, may be nonplussed by written tests. After years in manual work they will be ill at ease with pen and paper. They are often suspicious of intelligence and aptitude tests which seem irrelevant or even incomprehensible, a guarantee in advance that they will fail.
What is wanted is something that will encourage the applicant to do his best, motivation being a key factor in adult learning. The trainability test, both because of its obvious connection with the work to be done, and because it includes what can be a reassuring period of instruction, is likely to do this, both for older and for younger people. Whatever the age of the applicant, the trainability test is likely to tell the employer more about the applicant's potential for a practical job than a verbal interview.

What is almost equally important is that the informative nature of the test contributes to the applicant's own assessment of his suitability for the job: the test gives him a good idea of the work he will be doing and also of his likely competence. Moreover he meets a member of the staff he will be trained by if he gets the job. The consequences of this were neatly demonstrated in the course of some of the work done with sewing machinists. As the trainability test was being carried out on an experimental basis, applicants rated poorly on the test were still invited to join the firms for training. Yet a very high proportion (76.9 per cent of those rated E on a scale A-E) failed to start work. Of those rated A only 9.2 per cent failed to start work (Ref. 4).

A similar tendency was observed in some recent work with graduate applicants for jobs in a building society. The task used for the trainability test was to conduct an interview with a member of the public applying for a mortgage. (The member of the public was in fact the staff instructor.) The test turned out to be very helpful to the applicants themselves. In some cases their response was to the effect, 'if that's what you have to do, I'm not interested', whereas other applicants were much more enthusiastic about the job after the test than before it.

Trainability tests have been used explicitly for the purpose of self-assessment by the Training Services Agency on their Wider Opportunities Courses for the unemployed. Here the client learns the task in order to assess himself in the skill and decide whether he would like to be trained in it. In this case the task is called a 'skill sample', but it retains its structured form.

This use of trainability tests is being further developed in a pilot project with the British Steel Corporation. Redundant men are encouraged to try a number of 'Suitability Assessments' as a means of interesting them in retraining. A 'Suitability Assessment' differs from a trainability test in that after the initial test there is an unstructured period of practice and learning followed by a second test piece. The object of this additional period of practice is to reduce anxiety and give the client longer to 'taste' the skill. The effect of suitability assessments is being monitored in order to evaluate their bearing on commitment to training and eventual success in training and work.

The informative nature of trainability tests is not, of course, so relevant to applicants who have done the work in question before,
thought to opportunity to meet training staff and see something of the firm at first hand may help them to decide whether they want the job.

The test can prove a useful means of assessing an applicant's claim of previous experience and the extent of it, though if the task is very simple it may have a detrimental effect on the experienced applicant's attitude towards performing it. In these circumstances a slightly more difficult task can be substituted for that given to other applicants.

Although no study has as yet been carried out to assess the particular suitability of trainability tests for immigrants, some of the groups which have been involved in the research included immigrants: their performance on the tests was apparently unaffected by the fact that they were not born in this country. In the first experiments with sewing machinists for instance, the trainability test scores of immigrants were not as a group different from those of other applicants, and they were predictive of their performance in training (Ref. 3). In the study of electronic assemblers there was no correlation between an applicant's length of stay in Britain and her performance on the trainability test. Where there are language difficulties which could affect the applicant's understanding of the instruction he is given, it is necessary to have interpreters present. But in some cases the test instructions have been translated and successfully applied in languages other than English.

The majority of jobs for which trainability tests have been designed so far fall into the category of manual and semi-skilled. The test allows the applicant to express himself physically rather than verbally in a situation where his physical skill is the one in question: the test's justification lies in its ability to test this skill more accurately than other methods. Recent research developments, however, suggest it would be unnecessarily restrictive to assume that the test is applicable only to semi-skilled manual jobs. The manual skill for dentistry, for instance, is considerable, and there seems no reason, on the face of it, why trainability tests should not be developed for other jobs demanding a high degree of manual skill.

Nor, on the basis of an even more recent development, does it seem that the trainability test need be confined to manual skills. The exploratory work already mentioned with graduate applicants for jobs in a building society, suggests that the criteria might be widened. Here the task of conducting a mortgage interview was highly relevant to the nature of the work for which applicants were being chosen, and it was also something for which, although the job is ostensibly a white-collar one, it would be very difficult to judge a person's suitability by a pen and paper test. What this test did have in common with other trainability tests was that the task was practical in nature and that it could be taught in a standardised way.

If such criteria turn out to be relevant in this case, there would appear to be good grounds for exploring the usefulness of
trainability tests for other practical tasks in non-manual jo.

Yet a further example of the test’s adaptability was demonstrated by the Local Government Training Board who used a trainability test as a training aid for heavy goods vehicle drivers (Ref. 8). The instructor used the error check-list on each of a series of test drives which the trainee took in the course of training. After each test the instructor and trainee compared the trainee’s most recent score with his previous one to see how much learning had taken place. The training method for that trainee was then adjusted to correct the errors still being made. A pilot study showed that the addition of this aid to existing training methods accounted for an improvement in the pass rate.

Advantages and disadvantages of trainability tests

The first requirement of any selection method is that it selects well. The evidence so far shows that a good trainability test predicts success on a job involving manual skills more accurately than an interview. In most cases it also predicts success in a particular job better than general aptitude or intelligence tests. The informative nature of the test and the fact that it gives applicants the chance to judge their own competence may lead to some degree of self-assessment. Applicants who did not like what they learned about the job, even if they had acceptable test scores, are less likely to present themselves for work; both they and the employer are saved the fruitless time they might otherwise have spent in discovering that they did not want the job after all.

Many organisations use a probationary period of training as a means of selection. If a trainability test can be employed as a substitute for this and selection takes place as soon as an applicant presents himself, there are clearly advantages to the organisation and to the individual. The organisation is saved the time and money spent on trying to train a person who is eventually dismissed as unsuitable. From the individual’s point of view it is better to be rejected on initial application: taking up the new job may involve giving up a previous one, so rejection some way into training has severe practical as well as unpleasant emotional consequences.

The organisation benefits from having its training staff involved in the selection process. The training staff’s expertise is put to good use and the trainers are likely to work better with trainees they helped to select and in whom they have confidence. The relationship between the trainer and the trainee may be a significant factor in the progress which the trainee makes in learning and in the view which the trainee forms of his working environment. Training staff can immediately start to direct new recruits to the work for which they are best suited because the range of test scores can be designed so that they measure differences of expertise within the same skill area.
For example, if an applicant for a sewing machinist's job performs very well in the lockstitch trainability test, she can be trained for one of the more difficult tasks such as top stitching on a fashion or quality garment. If her results are poor she can be considered for training on a simple task such as hemming. As well as assessing the applicant's performance the instructor can spot potential training problems straight away. If the applicant is accepted the instructor has already had an opportunity to decide upon the most appropriate training methods for her.

A possible disadvantage of trainability tests is that they can be more time-consuming than other tests. They are best given individually and certainly to no more than two or three people at a time. They may also be expensive if materials are used up in the course of conducting them. These costs have to be balanced against the wastage of materials and time when unsuitable trainees have been recruited.

Because a trainability test predicts an applicant's training potential it is valid only for a situation where training is thorough and fairly systematic. Training staff are also crucial in their role as instructors for the trainability test. The test score will be meaningless if the instructor leaves out part of the standardised instructions, helps the applicant during the test piece or fails to observe errors. Moreover, the instructor's rating of the applicant's likely performance in training contains a subjective element not usually present in intelligence or aptitude tests. Therefore the effectiveness of the tests depends to a larger extent on the quality of the instructors. It is vital that instructors are themselves properly prepared for their task.

Another limiting factor is that the job-specific nature of the trainability test means that a new test has to be designed for each skill, though not necessarily each variation of a skill. A different test piece, for instance, was not necessary for the lock-stitch and overlock machinists, although of course, the instructions were modified to relate to the particular machine available for the test.

The skill specific nature of the tests was shown in a research project carried out in conjunction with the Shipbuilding Industry Training Board (Ref. 9). In this, apprentices were each given three trainability tests so that, for example, metal users were also given tests related to fitters and electricians. Basic training in all three skills was then given to the apprentices. Periodic phase tests given during training showed that each trainability test was related to its own skill area. This is a disadvantage in that the development of a new test takes time, skill and patience. The specific nature of each test, however, is its strength, and the possibility of designing a different test for each set out circumstances accounts for the relevance and versatility of this kind of selection method.
Chapter II: Constructing, validating and administering a trainability test

In the past few years the Industrial Training Research Unit has acquired considerable experience and expertise in the construction and administration of trainability tests, and will willingly advise on how practical assistance could be given to any firm or organisation wanting to develop tests for its own use.

This chapter describes in detail stages in the design and implementation of a trainability test which it is hoped will give firms a clear idea of what is involved. The information is also sufficiently detailed to act as a practical guide and could be useful for reference during the process of developing a test. Many practical points involved in the further development of trainability tests are covered in Ref. 12.

Designing a trainability test

A trainability test should normally be designed with the help of a member of the organisation's training staff, preferably an instructor who is involved in the training programme for which the test is to be developed. He or she will be familiar with the job in question and the conditions under which work is carried out. This knowledge is essential to the design of a satisfactory test. The instructor should be made responsible for consulting other training staff about the design of the test.

The procedure to be followed when designing a test is given in the summary below. Paragraphs (1)—(5) thereafter contain a full discussion of what is involved at each stage of this procedure.

Summary of the procedure

1. Analyse the job: identify the key operations and the essential skills which are required for its successful performance.
2. Select a work-piece or task which incorporates these skills and operations.
3. Write a check-list of errors which are liable to be made during the performance of the task.
4. Decide on the range of ratings which will be used and write a script to guide instructors on the use of the ratings.
5. Design and write the instructor's script. This tells instructors carrying out the test what to do and includes the wording which must be used when an instructor is testing an applicant.
1. **Analyse the job**

The purpose of analysing the job is to identify those operations which must be satisfactorily carried out for its successful performance and the essential skills which are involved in it.

A means of doing this is to compare the methods and attitudes of successful trainees and best workers on-the-job with those of the least successful. The design instructor can be asked to describe the work of a good trainee. He might, for instance, say of a good electrical trainee that ‘he made straight runs and neat bends’. Then he can be asked to describe the work of someone who had particular difficulty in learning the job. Here he might say that ‘he always removes too much insulation and leaves bare wires’. The instructor should be asked to identify as many critical points as possible—both positive and negative—in this fashion. Other instructors should also be consulted and asked to describe the best and poorest trainees they have had in recent months.

The points raised in these discussions throw light on critical areas of behaviour and also reveal those operations on which there is the greatest variation in performance. In the case of sewing machinists, for instance, the level of co-ordination between hand, eye and foot movements was found to be critical. A key operation was cloth handling. Good judgement of such perceptual features as height, tilt and position of the forks was essential for well-controlled operation of a fork-lift truck.

In addition to such specific perceptual and manual skills there are likely to be factors which differentiate good trainees from poor ones in almost any area. Discussions with instructors in fields as diverse as sewing, driving and dentistry have produced three factors common to good trainees: they are interested, they remember what they are told, and when they make mistakes they notice them and try to correct them in a sensible manner. A well-designed trainability test should give consideration to such features.

2. **Selecting the work-piece or task**

Selecting the work-piece or task must be done with considerable care as there are a number of different factors to take into account. It must incorporate those operations which have been defined as essential to the job and give scope for the observation of those areas of behaviour which are critical. It should also contain a number of opportunities for error. On the one hand, the greater the range of possible errors the more accurately the test is likely to distinguish between one applicant and another; while on the other, the task must not be so long or complex that it cannot be learned in one teaching session. Twenty to thirty minutes is usually the time allowed. No previous knowledge of the skill must be assumed.

There is usually little difficulty in choosing a task or test piece which seems to incorporate components already defined as essential.
to the job. At this point, however, it is wise to test the suitability of the task by listing all errors that could be made in performing it and comparing these with the essential skills involved. If for example, as in the case of electronic assemblers, soldering is an essential skill, then the test piece must involve soldering in such a way that faults which would impair job performance could occur during the test. Some modification in the task or work-piece first thought of may have to be made at this stage.

Appendix 2 summarises the essential skills which were defined for sewing machinists, fork lift truck operators and electronic assemblers, and the task which was chosen in each case to test those skills.

3. Writing the error check-list

Once the task or work-piece has been decided upon an error check-list should be written. This should contain, in a sequence matching the order in which mistakes could actually be made, those errors considered most relevant. The method used for recording errors must take account of the circumstances of the work. Some types of error can be recorded during the applicant’s performance, others may have to be noted immediately afterwards. This may affect the form in which the check list is written or the procedure the instructor must follow in using it. In the trainability test for sewing machinists, for instance, both the machinist’s hands have to be observed and errors can occur very quickly as the machines work fast. This makes it difficult to observe and record errors simultaneously. For this reason the applicant is asked to make three bags and the instructor observes errors while each of the three first seams are being made. While the other two seams are being made the instructor records the errors she observed on the first seam. The error check-list used in this case is shown in the centre pages section of this booklet.

In this instance all relevant errors can be observed from what the applicant does while performing the task. But in other tests, that of electronic assemblers for instance, not all errors can be observed in this way and the finished work-piece is also examined for errors. The most appropriate method must be decided upon according to the nature of the work in question.

Training staff involved in designing a trainability test sometimes want to write the check-list in positive rather than negative terms, that is by listing the things that the applicant does correctly. While this preference for encouraging the applicant is understandable, and for the purpose of long-term teaching a correct and important strategy, it is not appropriate for trainability testing.

Applicants tend to perform correctly far more often than incorrectly which, if the instructor does his recording properly, means that he would have to record a great many points. In practice however, he is more likely to notice every error than every instance of correct performance, consequently a record of errors is more likely to be accurate than a record of favourable points.
Research experience is that in every case where a list of correct instances of behaviour was used the accuracy of test scores was subsequently improved by reversing the statements and using an error check-list instead. An additional advantage of using an error check-list is that the type, range and frequency of errors provide clues which the instructor can use for planning training methods for the applicant.

4. Ratings

The instructor's rating of an applicant's likely performance in training is an holistic one and, to the extent that it takes into account such factors as the applicant's interest in the work and the effects of nervousness on his performance, it is bound to be subjective. However, personal bias is minimised by careful choice of an appropriate range and description of ratings and by giving guidance to instructors. Each rating takes the form of a judgement coupled with a prediction of the applicant's likely performance in training. An example of ratings and predictions is also included in the centre section of this booklet.

Discussion with instructors about variations in trainees' performance should be used as the basis for deciding on the range of ratings and the kind of prediction to be attached to each of them. If the organisation expects to reject a large proportion of applicants, more than one of the lower ratings are likely to be relevant. If the organisation expects to accept the majority of applicants only the last rating would probably recommend rejection.

Experience so far has shown a five-point scale to be the most useful range of ratings in most cases. In addition to ratings A to E in the overlock machine test, there are overall ratings which ask the instructor to comment on three types of behaviour found to be of particular importance in the performance of sewing machinists. Ratings of this kind should be included if there are positive aspects of performance which seem crucial and on which it would be relevant to require instructors to focus their attention. Note that the final factor in this case is the frequency with which the applicant notices and corrects the errors. This is likely to be generally applicable.

A factor, which on ITRU research experience to date, has not been found relevant is the time taken by an applicant in the test. Research with carpenters, sewing machinists and electronic assemblers found no correlation between the speed of test performance and success in training. Therefore, although the instructor may in practice take some account of speed in making his rating of an applicant, speed is not a factor which should be considered separately unless some clear indication is found to suggest that this element needs special consideration.

5. Instructor's script

The instructor's script includes two main elements: it tells the
instructor what to do and what materials he will need in order to carry out the test; and it provides the explanation the instructor will give the applicant about the purpose of the test and the text of instructions which he will use when demonstrating the task to the applicant. In order to write the script it is necessary to draw up a detailed and complete sequence of operations involved in teaching the task. The style and method of instruction should be similar to those which are used in training.

The script itself should be written in straightforward, unambiguous prose. It should not include technical jargon, which the applicant may not understand, and the aim should be to write short sentences whenever possible. It should indicate when the applicant should participate, and it may be necessary to interpolate the text of the instructions to the applicant with an occasional instruction to the instructor himself, as, for example: 'Do not tell the applicant but let her find out for herself'. These interpolations can be placed in brackets.

The drafts of the script and the assessment form should be tried out together in two stages.

First, one of the people designing the test should act as an applicant to an instructor who has not been much involved in its design. Once the instructor has had the chance to read the script, he should be asked to conduct the test following the script as he goes along. The designer/applicant should observe the instructor to see whether he understands from the script what he has to do, and at what point he has to speak to the applicant. This is also an opportunity to find out, through discussion, whether an instructor finds any difficulty in understanding and using the error check-list and the descriptions attached to the list of ratings.

Secondly, when all necessary alterations have been made the designer should then try out the draft script with one or more job applicants. The purpose at this stage is not to test the applicant's performance but to try to establish whether there is any ambiguity or difficulty in the instructions to applicants at the learning stage of the test. This can be done partly by observing whether the applicant is capable of following each instruction. If there is any failure on his part the designer should stop, and try to find out whether this is because the applicant has not understood something he has said. The designer may have difficulty in deciding how far failures are the result of inadequacy in the text and how far the result of poor comprehension on the part of the applicant. But if this occurs, the difficulty can be resolved by trying out the test on several applicants.

Any corrections which are required can be included in the final draft of the instructor's script which, together with the assessment form, is now ready for use.

The way in which the script will be used varies according to circumstances. The most common practice is for the instructor to memorise it. But sometimes the script is pre-recorded. In the case of the test
Example constructor's script for a trainability test.

INSTRUCTIONS FOR OVERLOCK MACHINE

Before the assessment starts, each applicant should be provided with:

1. A completed bag.
2. A pile of pieces for making the bags.
3. An overlock machine—threaded.

The following instructions are given while the applicant is sitting at the machine (Fig. 1).

For an inexperienced applicant

1. I understand that you would like to be trained as a skilled machinist.

2. I am going to teach you how to make a bag, (show completed bag) like this, using all the correct movements of your hands and feet and the right order of doing things. While I am teaching you how to make the bag you may ask as many questions as you like and I will do all I can to help you.

3. Then I will ask you to make three just like it, using the same methods as you have been shown. While you do this I cannot give you any help so make sure that you know what you have to do before you start.

For an experienced applicant

1. I know you have had some experience of machining but I am going to ask you to do a simple task like this (show completed bag).

2. I know you may be used to using the machine differently, but for this task would you do exactly what I ask you to do.

3. I will tell you how to make the first bag; then when you are ready I would like you to make three more by yourself.
**THE TEST**

**Introduction to the machine**

*(Make sure the applicant is sitting comfortably. If at any stage she does not seem to understand, the assessor must help her as much as possible.)*

*First of all we will see how the machine works. Underneath the table you can see two pedals, a large one on the left and a small one on the right. (These instructions will have to be amended for knee controls.)*

*Press the small pedal with your right foot and look at the machine to see what happens. (Do not tell the applicant but let her find out for herself.) The part which goes up and down is called the foot of the machine.*

*Press the large pedal on the left with both feet and see what happens. When machining you will always use both feet on the large pedal. When you are using the small pedal to lift the machine foot, you leave your other foot on the large pedal. Never press both pedals at once.*

**Cutting the thread**

*Take the thread, or chain as it is called, with your right hand and pull it across the front of the machine foot (Fig. 2). Now press the large pedal with both feet. What has happened? (Let the applicant reply.) Yes, this is the way we cut the thread with the machine. Will you please do it again.*

**Instructions for the first seam**

*Pick up two pieces of cloth and line up the pointed corners.*

*Press the small pedal to lift the foot and put the material under the foot for about ½ inch.*

*Now get the edge of the material level with the edge of the plate (Fig. 3). (Show the applicant what is meant by the plate.)*

*Hold the tail end of the pieces of cloth in your right hand, putting your thumb underneath and your fingers on top, and make sure the ends are together.*

*Put your left hand to the left of the foot and rest it lightly so that it can help to guide the cloth.*

*Now you are ready to sew. Put both feet on the big pedal and sew the seam in one burst of the machine. (Allow the trainee to do this.)*
Run off the end for about two inches so that you have a chain (Fig. 4). This will enable you to turn the cloth.

Instructions for the second seam

Turn the bag and position it under the foot, making sure that the edge of the material is level with the edge of the plate.

Hold the tail end of the pieces of cloth in your right hand, putting your thumb underneath and your fingers on top, and make sure the pieces are together.

Put your left hand to the left of the foot and rest it lightly on the cloth.

Put both your feet on the big pedal and sew the seam in one burst of the machine.

Run on for two inches for turning.

Instructions for the third seam

Turn the bag and put it under the foot, making sure that the edge of the material is level with the edge of the plate.

Hold the tail end of the pieces of cloth in your right hand, putting your thumb underneath and your fingers on top, and make sure that the pieces are together.

Put your left hand to the left of the foot and rest it lightly on the cloth.

Put both your feet on the big pedal and sew the seam in one burst of the machine, running off for two inches and cutting the thread.

Final instructions

(The applicant is allowed and encouraged to ask any questions at this stage. She is then asked to make three bags using the method she has been shown.)
OVERLOCK TRAINABILITY ASSESSMENT FORM

Factory
Name
Country of Birth
Assessor
Age
Date
Experienced/Inexperienced

<table>
<thead>
<tr>
<th>Seams</th>
<th>Bag 1</th>
<th>Bag 2</th>
<th>Bag 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seam 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seam 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seam 3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Aligns wrong seam first
Presents incorrect corner
Forgets to position cloth correctly
Forgets to align seam
Puts thumb on top
Does not use fingers of left hand correctly
Does not use fingers of right hand correctly
Seam not completed in one sew
Does not remember cutting method

Other errors (please describe)

Overall ratings
Positioning of hands: good (Always, generally, sometimes, rarely)
Positioning of feet: good (Always, generally, sometimes, rarely)
Notices errors and subsequently corrects:

Please circle appropriate letter.
A Extremely good. The assessor would expect her to become a very good machinist in a short time.
B Fairly good without being outstanding. The assessor would expect her to reach 100 performances in a reasonable time.
C Good enough for simple work. The assessor would expect her to become a steady worker on a simple machine, or task.
D Would have difficulty in training. The assessor would expect her to take longer training, and to perform a simple task.
E Would not be trainable. Even with a great deal of attention she would not make the grade, even on an easy operation.
for selectin...tistry students, for example, the verbal instruction is pre-recorded and played on tape while the lecturer demonstrates the task. The tape can be stopped while he deals with enquiries from the applicant. The advantage of this is that mistakes or omissions are avoided, but in some circumstances the playing back of a tape is not practicable.

Running a pilot study

However carefully a trainability test has been designed trials must be carried out before it can be put into use. Only when it has been established as a reliable predictor of performance in training within the individual organisation will the organisation be justified in using the test as a means of selection. The best way of trying out a test is by means of a pilot study which generally has three aims:

1. To provide a period of training for the instructors who will be using the test;
2. To give an opportunity for results from trainability tests to be compared with some measure of success in training in which the organisation has confidence;
3. To allow the test itself and the way it is administered to be monitored so that improvements can be made where these appear necessary from the test results.

1. Training of instructors

Because accurate assessment of performance in a trainability test depends on the ability to spot errors and the knowledge of which errors are likely to be crucial, it is important that all instructors who are chosen to administer the test should have good experience of training in the relevant job. However well qualified an instructor may be in general, he will not be suitable to administer the test unless he has experience of the job for which the test has been designed. Instructors should be selected for the task of administering the test according to this criterion. The ability to speak clearly and authoritatively is also a useful additional qualification.

Training sessions should be organised for the instructors to teach them about the principles behind trainability testing and how the test which has been designed for their job should be administered during the pilot study. The following paragraphs outline the kind of information which should be given if instructors are to acquire a proper understanding of what they are doing. It must, of course, be supplemented with information about the particular test which is being used for the pilot study.

Aspects of trainability testing

(i) The trainability test is a selection method based on the job sample. It has an important additional feature that the job sample does
not have: the task or work-piece selected as the job san is taught to the applicant before he is asked to perform it. The assessment which is made of his performance is thus intended to predict his ability to learn or to be trained in the job. This is why it is called a trainability test.

(ii) If the task selected for the test is taught in the same way to every applicant, and if each applicant’s performance is assessed in the same way, the test is expected to be a reliable guide to performance in training. The instructors who administer the test are provided with detailed instructions—an instructor’s script and assessment form—for teaching the test task and for assessing the applicant’s performance. These instructions are provided to ensure that the test is administered in a uniform manner.

(iii) Once the test is in use as a selection method, it has advantages for the applicant as well as for the organisation. Job applicants are less likely to be intimidated by this selection method than by other tests not so obviously related to the job. It also gives them a chance to find out what the job is like. Some of them may decide they do not want the job, which is a saving of effort to them and the organisation.

Test procedure

The test has three elements: the learning period, performance of the test task and assessment.

(i) During the learning period the instructor first explains to the job applicant what the trainability test is and what it is for. Then, using the instructor’s script—memorised, read, or on tape (as applicable)—he teaches the applicant the task which has been set for the test. During this time the applicant can ask as many questions as he likes and the instructor should answer them as fully as possible.

(ii) The applicant is asked to perform the task unaided. The instructor gives no help at all, either practical or verbal.

(iii) Assessment takes place both during and after the test performance. Assessment involves: (a) the recording of errors, and (b) a rating of the applicant’s performance according to guidelines provided on the assessment form. Errors should be recorded, as far as possible, during the test performance: ratings must be made immediately afterwards while the instructor’s recollection is quite fresh.

The instructor’s role

The instructor obviously has a crucial role in administering the test. It is his job to ensure that the test is given in the same way to each applicant; the reliability of the test depends on this. The test results can be critically impaired if the instructor fails to carry out his task correctly.

(i) It is essential that the whole of the instructor’s script be used in teaching the applicant the test task. If parts of the instructions are
omitted the applicant will clearly be at a disadvantage compared with other applicants, and thus it will be impossible to make a fair assessment of him. If it is necessary for the instructor to memorise the script—and yet he finds memorising difficult—he should keep the script constantly beside him for reference. Key points can be marked in red ink to help him find his place.

(ii) During the applicant’s performance of the test task some instructors are tempted into helping the applicant. This is an understandable mistake on the part of people who are accustomed to teaching, but the temptation must be resisted. If poorer applicants are helped they will appear to have performed better than they would have done unaided; so again there is no proper comparison with other applicants. During this part of the test the instructor must regard himself as an observer, not as a teacher.

(iii) It is very important to record all errors including, where applicable, an error which is repeated. Where errors are recorded on the basis of the work-piece itself as well as of behaviour during the test performance, it is easy to pay more attention to work-piece errors than to behaviour, just because they can be observed at leisure. But experience shows that behaviour and errors of technique can be more important for future performances than work-piece errors. Every effort must be made therefore to observe and record them correctly.

(iv) As previously mentioned, ratings should be made immediately after the test performance. Our memories are very short and a minute’s conversation on some subject other than the one in hand can mean that recollections are incomplete. This obviously affects judgement.

Instructors must also make sure not to ignore extreme scores on the rating, either on the argument that it does not do to award very high scores—‘no-one is perfect’—or because they simply dislike giving anyone the bottom score. If the range of scores is realistic, there will be a few applicants whose ability is correctly represented by scores at the extremes.

Reassuring the applicant

The instructor’s ability to make initial friendly contact with applicants is also very important. Applicants are likely to do their best if they are put at their ease. Some will be alarmed at the mention of a test, but may be reassured if the instructor explains to them carefully the nature of this one. Stress should be laid upon the following points:

(i) The applicant will be shown what to do and may have a chance to work at the task set, before being asked to perform it by himself. He can ask as many questions as he likes.

(ii) No-one gets it completely right the first time and no-one is expected to. The purpose of setting the task is to see how people
Instructor’s role during the pilot study

While giving a general explanation of the reasons for a pilot study and the way it will be conducted in this particular case, special emphasis should be put on the instructor’s role in observing faults in the test design or problems which arise in administering it. Instructors should be asked to take note of any difficulties they have in using the instructor’s script, the error check-list or the ratings. Suggestions they make for improvements will be welcomed.

Upon the instructor’s ability to keep records also depends the possibility of making a proper comparison between the trainability test scores and whatever measures of success in training are chosen for the purpose of comparison. Instructors should be given full and clear instructions about the manner in which the records are to be kept.

2. Comparing test results with other measures of success

Measuring the validity of a trainability test is something which must be done with care, but the process is not complex. Basically it involves making a statistical comparison of two sets of scores: between the trainability test error scores and some independent measure of success in performance and between the trainability test ratings and the same measures of performance. Crucial to the success of this operation is that proper records are kept of all trainees who take the trainability test and who enter training during the period of the pilot study. If the correct information is gathered, analysis of it is a fairly straightforward matter; but if information is missing, or has been gathered in a haphazard fashion the task of analysis becomes very difficult, if not impossible.

It is essential that no action is taken as a result of the trainability test scores during the pilot study. No applicant should be rejected on the basis of his trainability test scores.

Deciding on criteria for comparison

If trainability test scores are to be used as a means of selection it is important that they can be established as predictors of success and failure in training and on the job. Therefore, if practicable, they should be compared with the trainee’s performance at whatever point in training a final decision is usually made about each individual’s suitability. The relevance of this comparison clearly relies upon there being confidence in the correctness of decisions being made at this point.
There are times factors in the situation which make it difficult to judge impartially between one trainee and another at the end of training. Instructors may, in the course of training, give poorer trainees easier work than others and pay them more attention. This happened in large-scale trials in the clothing industry, with the result that poorer trainees appeared to have higher performances and rates of learning while in training than subsequently proved to be the case. On the other hand there was a direct relationship between performance on the shop floor and good test scores. It would, however, be a departure from usual practice for many organisations to keep poor trainees in employment until they reached the shop floor; usually they are dismissed before that point. Alternatively weaker trainees, who are nevertheless employable, may be engaged in simpler work on the shop floor.

The main conclusion to be observed in the light of these variations in conditions is simply that account should be taken of them when making judgements about a trainee’s competence.

It would be impracticable to ask instructors to teach all trainees in exactly the same way. Therefore if a single test of performance is made at the end of training it is important that this should be supplemented by measures of performance on the shop floor. The category of work an employee is doing may be in itself sufficient indication of his competence. If, alternatively, the decision about whether to reject or accept a trainee depends upon the instructor’s subjective assessment of competence at the end of training, he should also be asked to grade successful trainees—while bearing in mind the comparative difficulty of the work they have been doing.

Trainees who are rejected in the course of training ipso facto are the poorest trainees and should be treated as such for the purposes of comparison. Records should also be kept of applicants who fail to attend for work after taking the trainability test in order to see whether this has any connection with scores on the test. Instructors should be asked to make an assessment of any trainees who drop out of training and keep this in their records.

In the case of independent training establishments where trainees will subsequently go to a variety of companies, assessment must be done during or at the end of training. Follow-up on the job would be impracticable. During the trial period for the trainability test for fork truck operators, the trainability test scores were compared with trainees’ performance on a competence test at the end of training.

**Keeping records**

Each trainee will have two trainability test scores: his error score and his rating. Once these scores have been recorded they should be set aside and no longer be available to the instructors who are involved in training. Ideally a trainee should be trained by a different instructor from the one who gave him the test, and the test instructor should not tell the training instructor what the test scores were. If
this is not possible, removal of the scores is the best that can be done as a means of ensuring that the instructor is not influenced by them in his attitude towards the trainee.

Where trainees are usually dealt with only on an acceptance or rejection basis, with no records kept of their performance in training, it will be necessary to provide instructors with a special assessment form to be used during the pilot study. Performance in this case should be measured against the kind of criteria which are mentioned in the trainability test guide to the use of ratings. Reference might also be made to the kind of task the trainee can now perform competently. Instructors will obviously have to be consulted if an extra assessment form is to be drawn up for this purpose.

**Comparing the results**

In order to establish whether trainability test scores predict performance, statistical comparisons must be made between them and the other scores which have resulted from the measurement of their performance. The relationship between the two sets of scores is described in terms of the strength or weakness of the statistical correlation which exists between them. A convenient method of making the comparisons is by means of what is known as product moment correlation. The procedure for doing this and an explanation of how the results should be interpreted, including the use of statistical levels of significance, is contained in Appendix 3.

**3. Making adjustments in the test**

If there is a good correlation between both sets of trainability test scores and the performance scores, that is evidence that the trainability test is working well. Should, however, either or both correlations appear low, this suggests that there is something the matter either with the test itself or the way in which it is being administered. Information which has been gathered during the pilot study can be examined in the light of these results to suggest ways in which the test or its administration should be adjusted.

**Monitoring the test**

Throughout the pilot study instructors should be observed regularly to see whether they are administering the test correctly. Areas in which mistakes could occur are described in the paragraph headed 'The instructor's role', on pages 16 and 17. All such mistakes should be discouraged and instructors given help in carrying out their task correctly, but it is also important to take note of these mistakes. Instructors should also be asked at intervals for their comments on the test and on any difficulties they find in using it. Records should be kept of their remarks.
Analysing test results

Inadequacies in the test or its administration can usually be discovered through an analysis of the test results, and comparison of these results with criteria of performance and the information which has been gathered during the monitoring process.

If the correlation between performance criteria and the trainability test ratings is low, the first thing to examine is the ratings themselves. If scores are bunched around the middle range, the ratings have not been used to discriminate properly between one trainee and another. Obviously, then, they will not be predictive of performance for the strongest and weakest trainees. The tendency to use middle ratings on the part of only a minority of instructors would affect the correlation, so even if in general the ratings appear to have been properly used it is worth looking at the range of ratings used by individual instructors to see whether one or more of them is failing to use the extreme scores.

This problem was observed during the pilot study for the fork lift truck trainability test. Instructors kept to the middle range of the ratings, and for some groups of trainees ratings were not significantly related to results on the competence test by which they were independently assessed. In this case the trainability test error scores proved so valid that ultimately they were used as the sole criterion for selection.

An alternative explanation of poor predictive ability on the part of the ratings might be that the instructors, or some of them, gave help to trainees during the test itself or gave long practice periods. As long as any instructor is giving such help the predictive ability of the test will be badly affected, and it cannot justifiably be used as a means of selection.

If the correlation between performance criteria and the trainability test error scores is low, the possible reasons for it are various. The cause could again be that instructors are giving help during the test, or there could be practical difficulties about recording errors while the test is in progress. This should have emerged from the comments of instructors during the pilot study and corrective action would involve making a change in the way the test is administered.

Yet another possibility is that the task is too easy and that consequently most trainees make the same few mistakes as each other. This could be established by scrutiny of the error scores which in these circumstances would be low. If the task is difficult enough the number of errors should range from a very few for a minority of trainees, through a medium number for the majority, to a very high number for the minority at the poor end of the ability range.

Provided that the test task is at the right level of difficulty, error scores can generally be expected to be good predictors of performance.
The trainability test in use

Acceptable and unacceptable test scores

Once the pilot study is completed and the test has been adjusted where necessary, it can be put into use as a method of selecting trainees from among all applicants for training. Hitherto no decisions have been made about an applicant's suitability on the basis of the test. Now that this is to be done, a decision must be reached about the level of score an applicant must have in order to be accepted for training: instructors will need a simple way of 'reading' an applicant's result in order to decide whether to accept or reject him.

The level of trainability test score at which an applicant is acceptable will obviously be related to those criteria of performance with which the trainability test scores were themselves compared. In other words, an acceptable trainability score will be that which predicts satisfactory performance according to these other criteria (in most cases performance as measured at the end of training). A way of interpreting an applicant's results in the light of his expected performance is by using a simple chart called an expectancy chart. The chart gives information about the percentage of applicants with each trainability test score expected on the basis of independent criteria of performance to do well, to be acceptable, or to be unacceptable. Data of past test scores and performance scores are required for this purpose. Examples of expectancy charts and instructions about how to produce them are given in Appendix 3.

The organisation may change its policy of admission from time to time according to its number of vacancies and the number of applicants applying for them. At a time when it has few vacancies policy may be to accept only applicants with A or B scores. If, however, there is a shortage of suitable applicants, applicants with C scores may also be accepted despite the fact that a proportion of such applicants will probably turn out to be poor performers: as for example in the expectancy chart for lockstitch machinists where 20 per cent. of C applicants are expected to be of below average performance.

Discussion with instructors

Sessions should be organised for instructors so that they can be given a full report of the pilot study and told about the procedures to be followed when the trainability test is in use as a method of selection. The opportunity can also be taken to re-emphasise the importance of their role for the proper administration of the test and to correct any errors of administration which revealed themselves during the pilot study.

The test procedure will be different in certain respects now that it is in use as a selection method:

(i) Because applicants may be rejected on the basis of the test results it is particularly important to reassure them before they take it.
This will help them to do their best. Emphasis should be laid on the fact that they will be shown what to do and that they can ask as many questions as they like during this learning period. Instructors should also point out that the test gives the applicant a chance to find out whether he likes the job.

(ii) What the instructor learns about the applicant’s ability during the test can be used to decide the most appropriate training method. The information will be mainly in the form of the test error scores and ratings, but the instructor will also have made observations of the manner in which the applicant goes about the test task and these may help him to decide on the kind of training required.

(iii) Instructors will need to know about the meaning and use of the expectancy chart. The organisation’s recruitment policy—and thus its current decision about what is an acceptable test score—should also be explained.
Trainability tests, as we have seen, are a very pragmatic way of approaching the selection of applicants who have the capacity to learn a job under the conditions in which that job operates. If a candidate shows signs of being able to learn the job under those conditions, then it is probable that he will thrive in it and prove satisfactory in the long-term.

Trainability tests make no presumptions about aptitudes and the components of skills, but instead take an holistic picture of the competence of the candidate to learn a job in all its various aspects. And, because this condensed learning task is something of a microcosm of the job itself, there is less chance of making an incorrect inference from success in the test to success on the job than in the case of an aptitude test, for instance. Moreover the practical requirement of establishing the validity of the test does not involve an analysis of why one trainee is succeeding and another is failing.

While theoretical issues are not dominant during the design and administration of a trainability test, this form of selection method is closely related to the specific training methods used.

The Industrial Training Research Unit has developed a theoretical model known as CRAMP on which training decisions can be made. (Refs. 10 and 11). Briefly summarised, different types of learning require different training methods, and any one individual may be good at one form of learning while being poor at another. Thus an individual’s ability to learn a given skill will depend upon the particular form of learning he is required to engage in. In addition he may be influenced by his own interest in what he has to learn and by the relationship between trainer and trainee. The ability to progress in learning, therefore, is often specific to particular situations. By using a trainability test we recognise the complexity of learning and attempt to limit the errors to which all attempts at prediction are subject by reproducing as closely as possible the task and the environment with which the selected trainee will be engaged.

The successful application of this approach depends to a large extent upon the importance which is attributed both to selection and training within the individual organisation. Success in administering the test depends upon the quality of the instructor, and professionalism on the part of the instructor depends upon how he is viewed within his organisation. This was shown during trials conducted by
the 50 companies taking part in the Knitting, Lace and Net Industry Training Board project (Ref. 4).

Where training conditions are favourable, organisations can expect to introduce trainability tests successfully. In other cases it may first be necessary to improve the training methods used. There are now a number of Industrial Training Boards equipped to help firms in the development of their own trainability tests.

Let the last word be said by G. T. Minshull, the Chief Inspector at Pye Telecommunications Ltd., where the trainability test for electronic assemblers was developed:

It has been recognised for some time that our ability to select Female Operators for training leaves much to be desired. The method of selection has been by interview by the Personnel Department plus a simple test of memory. Figures showed that a large number recruited left in the first week of training. We clearly required some form of test at interview that would on the one hand show us whether the applicant was suitable for training and, on the other, give the applicant some insight into the type of work for which she would be trained. Both parties could then arrive at a decision.

Advice was sought from the Industrial Training Research Unit in designing a test that would:

a. Show the applicant the type of work
b. Enable us to assess her ability to be trained.

The ITRU Trainability Test does both of these things. The results of its application have contributed to:

a. Reduction in labour turnover
b. A better operator
c. Reduced training time, and therefore, reduced cost.
## Appendix 1: References

<table>
<thead>
<tr>
<th></th>
<th>Author(s)</th>
<th>Title</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Deubert L.W; Smith M C; Downs S; Jenkins L.C.B; and Berry D.C</td>
<td>The Selection of Dental Students: a pilot study of an assessment of manual ability by practical tests <em>British Dental Journal</em> Vol. 139, No. 9 Nov. 1975.</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Local Government Training Board</td>
<td><em>Driver Training Performance Monitoring Test</em> 1976.</td>
<td></td>
</tr>
<tr>
<td>JOB</td>
<td>ESSENTIAL COMPONENTS OF SKILL</td>
<td>TRAINABILITY TEST TASK</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Sewing machinist</td>
<td>1. Cloth handling including cloth positioning.</td>
<td>Machining three seams to join two pieces of cloth to make an open bag. Three identical bags are made.</td>
<td></td>
</tr>
<tr>
<td>Fork lift truck operator</td>
<td>1. Smooth use of controls involving co-ordination of hand, eye and foot.</td>
<td>To start the truck and drive it forwards. To make a left-hand turn round a drum and pick up a pallet. The pallet is then placed in a marked area and the truck is reversed back round the drum to the starting position and parked.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Ability to relate the steering lock to the vehicle position.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Ability to approach a load squarely and centrally.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Electronic assembler</td>
<td>1. Soldering.</td>
<td>Solder three tag strip joints and three wrap round joints. Select three coded wires and insert them in their correct positions on the reverse side of a printed circuit board.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2. Positioning components and wires.</td>
<td>Turn the board over, bend and solder wires to the circuit path.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3. Control of tools, especially pliers and cutters.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4. Selection of correct components.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix III: Statistical validation of test results

1. Product moment correlation and its interpretation

In order to use product moment correlation it is necessary to have all scores in numerical form. Ratings, for example, can be expressed by numerals 1–5 in place of the letters A–E. There need, however, be no correspondence in the range of the sets of scores used—the range of two sets of scores, for example, does not have to be identical. A comparison may be made between error scores including high numbers and a short range of scores from measures of performance say, only 1–5. The other necessary condition is that each set of scores should be normally distributed. That is to say there should be a few high scores and a few low scores with the majority clustered in the middle of the range.

Product moment correlation between trainability test error scores and scores resulting from measurement of performance.

If we have less than say 80 results we would work out the correlation called ‘r’ as follows:

Write down each individual’s error score on the trainability test (X) and his performance score (Y) in two parallel columns in their pairs so that the pair of readings in each row belongs to the same person. e.g.

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Jones</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Brown</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>Singh</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

Next calculate two more columns headed $X^2$ and $Y^2$ by squaring the figures in the first two columns: —

<table>
<thead>
<tr>
<th></th>
<th>X</th>
<th>Y</th>
<th>$X^2$</th>
<th>$Y^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith</td>
<td>7</td>
<td>2</td>
<td>49</td>
<td>4</td>
</tr>
<tr>
<td>Jones</td>
<td>12</td>
<td>2</td>
<td>144</td>
<td>4</td>
</tr>
<tr>
<td>Brown</td>
<td>15</td>
<td>4</td>
<td>225</td>
<td>16</td>
</tr>
<tr>
<td>Singh</td>
<td>6</td>
<td>3</td>
<td>36</td>
<td>9</td>
</tr>
</tbody>
</table>
Then find a fifth column, headed \( XY \), by multiplying together corresponding \( X \)'s and \( Y \)'s in the first two columns:—

<table>
<thead>
<tr>
<th></th>
<th>( X )</th>
<th>( Y )</th>
<th>( X^2 )</th>
<th>( Y^2 )</th>
<th>( XY )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith</td>
<td>7</td>
<td>2</td>
<td>49</td>
<td>4</td>
<td>14</td>
</tr>
<tr>
<td>Jones</td>
<td>12</td>
<td>2</td>
<td>144</td>
<td>4</td>
<td>24</td>
</tr>
<tr>
<td>Brown</td>
<td>15</td>
<td>4</td>
<td>225</td>
<td>16</td>
<td>60</td>
</tr>
<tr>
<td>Singh</td>
<td>6</td>
<td>3</td>
<td>36</td>
<td>9</td>
<td>18</td>
</tr>
<tr>
<td>( S(X) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>40</td>
</tr>
</tbody>
</table>

Then add up, or sum = \( S \), all the figures in each of the five columns so that we have \( S(X) \), \( S(Y) \), \( S(X^2) \), \( S(Y^2) \), \( S(XY) \). The remaining term required for the formula is \( N \).

\( N \) = The total number of people with trainability test and performance scores.

\( S(X) \) in the example given would be 40, and \( N \), the total number of people, would be 4. In practice, of course, the numbers of people involved would be many more than this.

The correlation coefficient can be found from the following formula:—

\[
r = \frac{S(XY)}{N} - \frac{S(X)}{N} \times \frac{S(Y)}{N}
\]

\[
= \sqrt{\left(\frac{S(X^2)}{N} - \left(SX\right)^2\right)\left(\frac{S(Y^2)}{N} - \left(SY\right)^2\right)}
\]

Two comparisons must be made using this method: between the set of trainability test error scores and the set of scores representing performance, another between the set of trainability test ratings and the same performance scores. The number which results in each case is called the correlation coefficient.

By interpreting it correctly we can answer the question: is there a tendency for the individual who scores high (or low) on one set of scores (the relevant trainability test scores) to score high (or low) on the performance scores. If the relationship is perfect, then the correlation coefficient will be \( +1.00 \). If there is no relationship at all, it will be 0.00. If there is a perfect negative relationship (that is an individual who scores high or low on the trainability test scores low or high on performance criteria) the correlation coefficient will be \( -1.00 \).

In practice the correlation coefficient is likely to range between, say, 0.1 and 0.8, and the nearer it approaches to \( +1.00 \) (0.75 for instance), the stronger is the relationship between the two sets of scores.

However, confidence that a correlation coefficient of 0.75 really does represent a strong relationship depends upon enough people having been tested. Broadly speaking, the smaller the number of
people being tested, the higher must the correlation coefficient be. If a very large number of people have been tested, a lower correlation coefficient might still represent a strong relationship between the two sets of scores. An apparent relationship in the scores of two people might be a matter of chance. If the same relationship appears for a very large number of people the chance element is much reduced. Whether or not the correlation coefficient is really strong is usually referred to in terms of its ‘significance’.

The degree of significance of any correlation—the amount of confidence you can place in it—can be ascertained from a table in a statistical handbook. Some computer programmes automatically give the level of significance of the results. Any correlation which would occur only one occasion in twenty on the basis of pure chance is usually accepted as being reasonably significant. This significance level is represented as \( p < .05 \). If the correlation coefficient between one of the sets of trainability test scores and the performance scores was at this level, consideration might reasonably be given to using the trainability test as a means of selection. It is preferable, however, to depend upon a correlation coefficient only if it would occur by chance once in 100 occasions. This is represented as \( p < .01 \). An even more impressive correlation coefficient would be one in which the relationship could have occurred by chance only once in 1000 occasions. \( p < .001 \).

2. Producing an expectancy chart

An expectancy chart is a means of showing in diagrammatic form the relationship between test scores and the proportions of applicants for each score who have been found to perform well or badly according to those independent criteria of performance in which the organisation has confidence. Scores which are used in the chart should be ones which have been shown to be significantly related using a product-moment correlation.

In the example given opposite of an expectancy chart for sewing machinists using a lockstitch machine, each score on the trainability test grade is represented by the percentage of machinists gaining that score who were found to be above average, average or below average performers. The information in the chart is based on 72 pairs of results.
TABLE 1

*Expectancy chart included in the report on 1973/75 Research Project conducted by the Knitting, Lace and Net Industry Training Board.*

<table>
<thead>
<tr>
<th>Grade</th>
<th>Awarded</th>
<th>Performance in job</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D+E</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The use of the chart depends upon the assumption that if 86 per cent of applicants who gained grade A in the test were subsequently found to be above average performers a similar proportion of those gaining grade A in future will do equally well. Provided that the relationship between the original two sets of results is statistically significant, this is a realistic expectation.

Two versions are given of an expectancy chart produced for fork truck operators. Table 2 has been produced in the same way as that for the lockstitch machinists except that in this case the trainability test score is given in terms of the number of errors.
TABLE 2

Expectancy chart showing how to gauge the expected performance of a trainee on the competence test from his trainability total error score.

(Based on 263 results)

<table>
<thead>
<tr>
<th>Trainability errors</th>
<th>Expectancy (per cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>67%</td>
</tr>
<tr>
<td>5-8</td>
<td>46%</td>
</tr>
<tr>
<td>9-12</td>
<td>12%</td>
</tr>
<tr>
<td>13-14</td>
<td>11%</td>
</tr>
<tr>
<td>15 or more</td>
<td>0%</td>
</tr>
</tbody>
</table>

Table 3 includes the same information as Table 2 but in this case it was placed at the end of the instructor’s assessment form. Thus the instructor can see at a glance when he fills in the error score, a trainee’s expected standard of performance and both pieces of information are conveniently placed together.

TABLE 3

<table>
<thead>
<tr>
<th>Total Errors</th>
<th>Tick as Appropriate</th>
<th>Expectancy (per cent)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Very Good</td>
</tr>
<tr>
<td>0-4</td>
<td></td>
<td>56</td>
</tr>
<tr>
<td>5-8</td>
<td></td>
<td>23</td>
</tr>
<tr>
<td>9-12</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>13-14</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>15 or more</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
### PREVIOUS BOOKLETS IN THIS SERIES

<table>
<thead>
<tr>
<th>#</th>
<th>Title</th>
<th>Author(s)</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Design of Instruction</td>
<td>Dr Sheila Jones</td>
<td>15p</td>
</tr>
<tr>
<td>2</td>
<td>Identifying Supervisory Training Needs</td>
<td>Dr P B Warr and M W Bird</td>
<td>55p</td>
</tr>
<tr>
<td>3</td>
<td>The Challenge of Change to the Adult Trainee</td>
<td>Miss D N Newsham</td>
<td>22½p</td>
</tr>
<tr>
<td>4</td>
<td>Improving Skills in Working with People: the T-Group</td>
<td>Dr P B Smith</td>
<td>23½p</td>
</tr>
<tr>
<td>5</td>
<td>The Discovery Method of Training</td>
<td>Dr R M Belbin</td>
<td>30p</td>
</tr>
<tr>
<td>6</td>
<td>Task Analysis</td>
<td>Dr J Annett, K D Duncan, R B Stammers and M J Gray</td>
<td>25p</td>
</tr>
<tr>
<td>7</td>
<td>Improving Skills in Working with People: Interaction Analysis</td>
<td>Damien A Dyar and W John Giles</td>
<td>31p</td>
</tr>
<tr>
<td>8</td>
<td>Selecting the Younger Trainee</td>
<td>M C Killcross and W T G Bates</td>
<td>52p</td>
</tr>
<tr>
<td>9</td>
<td>Selecting and Training Coloured Workers</td>
<td>Dr M A Pearn</td>
<td>£1.00</td>
</tr>
<tr>
<td>10</td>
<td>Hurt or Helped? A study of the personal impact on managers</td>
<td>Cary L Cooper and David Bowles</td>
<td>£1.00</td>
</tr>
</tbody>
</table>