ADAPTIVE ABILITY TESTS

User Manual

Selby & Mills Limited
Tregoning Mill
St. Keverne
Helston
Cornwall
TR12 6QE
United Kingdom

Tel: +44(0) 1326 280382
Email: info@selbymills.co.uk
Web site: www.selbymills.co.uk

CONTENTS

1. INTRODUCTION

Ability tests have been used in the UK over several decades and the ability test market has formed by far the major part of the occupational test market in the UK. They have until recently been paper and pencil materials. In the last decade the PC has become an integral part of virtually all work environments. This has opened up new opportunities in the field of occupational testing and especially down the route of ability testing.

The Adaptive Ability Tests consist of three modules, which are offered to the user in software form, and can be loaded onto most PCs. The three modules are:-

1. Language Ability.
2. Numeric Ability
3. Administrative Ability.

The Adaptive Ability Tests were developed to offer the user a number of key benefits over traditional paper and pencil tests.

A. **Each of the three test modules will assess the entire range of ability in a single test administration lasting no more than twenty minutes.**

To do this the Adaptive Ability Tests utilise sophisticated hunting algorithms as part of the software design. These algorithms ensure that the candidate will need to answer fewer questions to indicate their level of ability to the same accuracy as traditional methods. Such tests are called Adaptive Ability Tests. The hunting algorithms work as follows.

The computer continually monitors the speed and accuracy at which the candidate is working. This then enables the computer to select only those items which lie close to the candidates ability level. If the candidate answers a set of questions correctly in a short space of time the computer will present them with a set of questions at a higher level of ability. If the candidate is answering many questions wrongly or is taking a long time, the computer will present them with questions at a lower level of ability.

As the test session proceeds the performance level is re-assessed by the computer on a number of occasions. If the performance of the candidate is consistent then the tests will automatically time out. If it appears that the candidate is operating at a particularly high level of ability or at a particularly low level of ability, the computer will present more items. This ensures that the computer will present more items. This ensures that the tests discriminate well at either end of the ability spectrum, i.e. the tests discriminate well at the tails of the normal curve.

B. **There is no need for Alternative Forms**

Paper and pencil ability tests often require that different forms of a test are used depending on what level of ability the sample has that is being assessed. The Adaptive Ability Tests ensure that the individual administering the test does not need to do their own subjective analysis of their candidates ability level before the test has even begun. Consequently they do not run the risk of ceiling or floor effects occurring because they have used the wrong form. The Adaptive Ability Tests ensure that the whole range of ability is assessed in a single administration.

C. **Reduced administrative burden**

The use of the PC eliminates a lot of the paraphernalia that the traditional paper and pencil users have had to contend with. Answer sheets, test booklets, administration cards, scoring cards, pencils, rubbers and stopwatches are no longer needed. Test administration and scoring are carried out entirely automatically. The candidate answers each question at the computer. Once the candidate has completed a particular module, the results are stored on the computer. The administrator may then produce a hard copy of the results. Scoring is carried out automatically and feedback can be given instantaneously.

D**. Developed specifically for use in the UK and Europe**

The Adaptive Ability Tests were wholly developed within the UK. The initial items were drawn from the extensive results of ability tests which have been used in the regular and territorial armies for over a decade. The tests have also been trialled in a number of UK companies, as well as with an extensive sample of graduates across the UK. Lower levels of the numeric ability module were trialled at schools in Oxfordshire.

E. **Improved reliability**

The modules are entirely administered by the computer. This eliminates the bias associated with human administrations. The computer will guarantee 100% consistency in administration. Similarly the results are scored by the computer and again this will be independent of human error.

F. **Automatic data storage**

As a candidate completes a particular test the results are automatically filed. This means that the results remain accessible at all times and can be used as a

database for producing additional norms. They can be archived for future retrieval.

G. Comprehensive Support and Updates

As these adaptive tests were developed by us, we are able to offer full telephone support and software updates. Refinements in the software will automatically be available. As part of our development programme the user will be regularly updated with new norms and validity studies to add to their manual and users guide.

2. THE ADAPTIVE ABILITY TEST MODULES

2.1 Description

The Adaptive Ability Tests were developed and computerised for use in the selection, training and career development of personnel in the UK. The modules are designed for all types of company personnel, these include:

1. Operatives
2. Clerical/Supervisory Staff
3. Management

The Test Modules include measures of Language, Numerical Ability and Clerical/Administrative skills. The modules have different levels of difficulty. This means that the individual taking the test will only have to answer the questions suited to their level. They do not have to answer a large number of redundant questions which would have to be answered with a paper and pencil test.

2.2 Language Ability

The language module was designed to test the range and clarity of a person's vocabulary as well as their ability to use language. Analogue tests of this type are highly correlated with verbal reasoning intelligence tests. Such tests are used to predict successful performance in examinations and training courses which have a high verbal component. High scores on this test would be essential for positions which require sophisticated communication skills. Similarly, minimum scores are required for operative or supervisory roles where precision is required in giving and receiving instructions. It is to be remembered however that a low score on this test is not an indicator of low intelligence. Many scientists and technologists have poorly developed verbal skills whilst possessing high levels of abstract and spatial ability.

The module covers a range of ability from final year junior or first year secondary school through secondary school up to graduate level.

2.3 Numeric Ability

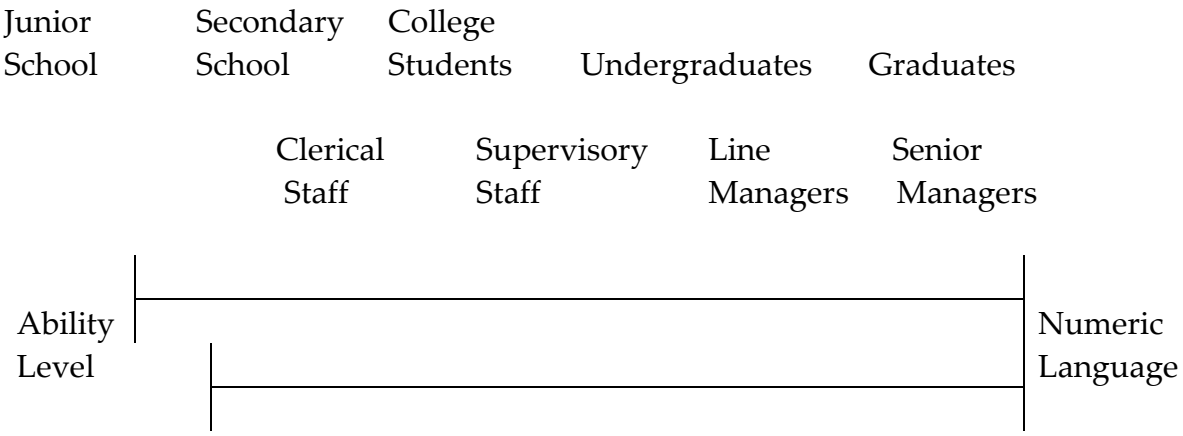
The numeric ability module was designed to measure the aspects of numeracy generally required in work situations. At lowest difficulty level simply 'four rule' numeracy is covered. In the higher difficulty levels the emphasis is on the individual's ability and confidence when using numbers or formulae. Though little numeric knowledge is required the module examine a person's ability to estimate and solve problems with speed and relative accuracy up to

about 'O' or GCSE Level 'C'. High scores on this type of test suggest that a person will cope with numeracy generally used by managerial and supervisor tasks in non-technical settings. It is a key managerial skill and a key deficit among managers and graduate candidates for employment.

This ensures that the module will assess across the whole range of ability (although not the specialist maths skills of a graduate mathematician), and that the module is appropriate to all levels of occupational use.

Diagram 1 below indicates the range of individuals that can be assessed using the language and numeric modules.

DIAGRAM 1 THE RANGE OF INDIVIDUALS WHO CAN BE APPROPRIATELY ASSESSED BY THE LANGUAGE AND NUMERIC MODULES OF THE ADAPTIVE ABILITY TESTS



2.4 Administrative Ability

This module is split up into 3 sections. The administrative ability sections are performance tests which measure the speed and accuracy with which a person approaches clerical tasks. Though this type of measure is predominantly used to select clerical and administrative personnel the ability to organise and perform well on such tasks is a requirement at all levels in modern companies. The modules are highly predicative of work performance on paper work and other administrative tasks.

The three sections in the Administrative Ability Module are:

- NUMBERS
- ADDRESSES
- CODES

Again they are structured to evaluate across a range of difficulty. This includes the lower half of the ability range, which might be associated with manual workers or operatives. Through ability levels which would be typical of clerical and supervisory staff. Up to the top end of the ability range and scores within this level represent considerable performance skill.

The three sections within the administrative ability module provide the user with a breakdown of that candidate's performance on administrative tasks. This will give an indication of where specific aspects of an individuals skill may lie. For example, assessment for an accountancy position will require a high degree of numeric checking whereas other aspects may be of less importance. It should be noted that these sections are not discrete. An individual will only receive a score on these scales as subsets of the administrative test module. These sections should not be given equal weightings as the three main modules, they are three subsections of the administrative module.

3. ADMINISTRATION

The computerised version of the Adaptive Ability Tests are entirely administered by the computer. However it is also important to consider the conditions in which the computer is used. The computer should be set up in a quiet room free from interruptions. It is also important the contrast, brightness and colour of the screen is to the satisfaction of the candidate. The screen should also be set to avoid any glare, for example, from a window, or internal lighting.

Once the candidate has been settled at the screen the session is administered by the computer. At the beginning of each of the modules, there are a set of screens which introduce that particular module. These screens are self paced, i.e. the next screen will not appear until the candidate presses the 'enter' key on the keyboard. These initial screens show the candidate examples of the type of question they will encounter, and how to select an answer.

Timing for a particular test does not begin until the first question appears on the screen.

Each test starts with a short locator test, which determines the level at which the applicant will start answering. Once allocated to a given level in a subject area, the candidate's performance is continuously monitored, and they are shifted to different levels if they are performing above or below average from that level. In this way the candidate has the maximum number of questions around their particular level of ability,

Questions are presented sequentially, one at a time. The candidate will not be presented with the next question until they have answered the question they are on. Timing of the test occurs automatically, the test will be terminated automatically and this may occur at any stage during a test.

A screen will inform the candidate that they have completed the test and that they should inform the administrator.

Section 3.1 below gives an example of how to run the test session, and what to say to the candidate.

3.1 Administration Sheet

1. Ensure that the computer is ready for the candidate to sit down and begin their assessment session (see administration section of user manual).

2. Ensure that there is some rough paper and two pens or pencils. A calculator should be provided if the candidate is to complete the Numeric Ability Test.

3. Settle the candidate in front of the screen and make sure they can read the screen perfectly clearly. If the candidate needs glasses they may wish to fetch them at this point.

4. Once the candidate has been settled, say :

   These test(s) are test(s) of ability. During this session you will complete the following tests. (THEN NAME THE TESTS THAT HAVE BEEN SELECTED). Full instructions are given at the beginning of each test, on the computer screen. You move from one screen to the next by pressing the return key (DEMONSTRATE WHERE).

   The instructions will indicate how to answer and they will also give examples of the types of questions you will have to answer. Do any rough work on the paper provided, and use the calculator if you wish.

   Each test is automatically timed. The time will vary but generally each test will last between 10 and 20 minutes. Do not worry if the test finishes while you are in the middle of a question.

   Timing does not start until you are presented with the first question of the test. Please read the instructions carefully. When you have completed the tests please inform the administrator.

   Are there any questions before you begin? (ANSWER ANY QUESTIONS THAT THE CANDIDATE HAS (SEE SECTION 3.2), AND THEN SAY) - You may begin now.

3.2 Questions and Answers for Administrators

Do I lose marks for incorrect answers?

No, you do not lose marks if you answer any of the questions incorrectly. How well you do on each test will depend on a combination of three factors. Firstly, the number of questions that you answer correctly. Secondly, the difficulty level of the questions that you answer; for example, a hard question will be worth more marks than an easy one. The final factor is the length of time it takes you to answer the questions that you attempt. The computer will not add or subtract anything from your score if you answer a question incorrectly.

How long will the tests take?

Each of the tests takes between 10 and 20 minutes to complete. The timing is carried out automatically by the computer. This means that any test may finish when you are halfway through a question, try not to let this bother you. Also, you may find that different people will take different amounts of time to complete the test. This is because sometimes the computer will take longer to identify the candidates level of ability. This is especially true for candidates who answer erratically.

How many questions will I have to do?

This will depend on how quickly you answer the individual questions. Generally you will have to answer between 30 and 40 questions. If you are answering quite cautiously and therefore slowly you may complete less questions in the time limit allowed. However, you should remember that the questions vary in difficulty. This means that if you answer fewer of the more difficult questions you could get the same score as someone answering a greater number of the less difficult questions. Speed and accuracy are both taken into account.

Are the tests valid for me?

These tests have been developed as the result of extensive trials involving thousands of people. These have included secondary school children, university students, graduates, clerical staff, senior managers and others. The tests have been specifically designed to measure the abilities of all types of people of working age. The reason that you have been asked to complete these tests is that they have been found to be an important factor which is associated with being able to do the job for which you are applying.

NB If a candidate queries the validity of the test for their ethnic or cultural group, reassure them of the integrity and validity of the entire procedure.

I'm a member of a minority group?

These tests will (or should) have been validated for use with respect to the application for which they are being used. This means that if a test measures a candidate's numeric ability, the organisation using it should have determined in advance that this kind of ability is an important requirement for success in the job or career for which the candidate is being considered. Therefore, irrespective of the candidates' ethnic or racial origin, their gender or age, they need to be able to cope with the type of work requirement which is evaluated by the test or tests.

NB When you are administering to someone and you feel the candidate is genuinely uncomfortable, use your discretion to query this with the candidate. Generally believe what candidates tell you in this respect. If you are unsure whether to intervene, and decide not to do so, but feel the candidate is genuinely experiencing discomfort due to the test, make a note of this on the paper and discuss it with the authorised test user in your organisation, or if you are that person, please call for advice. Always err in favour of the candidate's interests.

My mother tongue isn't English?

The general level of language in the tests is at the 11-12 year level of difficulty. The language ability tests is aimed at assessing the range and clarity of an individual's vocabulary. This is because this skill is an important requirement for success in the role for which you are being considered. If you genuinely do not wish to complete the test you must realise that we will not have as much information on your application as we will have for the other candidates who have completed this test. Your application may suffer as a result of this. Please complete the test and discuss any difficulties with me afterwards.

4. SCORING THE ADAPTIVE ABILITY TESTS

Each module of the Adaptive Ability Test is automatically scored by the computer.

The score report that is produced for an individual shows two scores. The first score is the raw score which is a direct result of the number of questions which the individual answered correctly. The second score on the score report is a STEN score which is an interpretation of that individual's raw score against a selected norm group. (See below). The STEN score is also represented diagramatically as a bar chart on the score report.

For the numeric and language modules, the score report consists of a single pair of scores (the raw score and the STEN score). For the administrative test module the score report shows the level of attainment in each of the three sections of that module as well as an overall score on the administrative ability module. The score report therefore looks as follows.

DIAGRAM 2: SAMPLE SCORE CHART

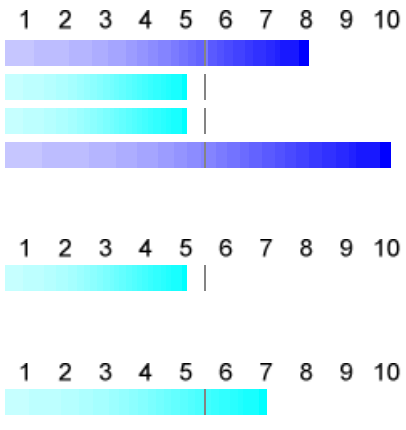
Score Chart Report

Report Date **21/03/2004**

Administrative Ability	Raw Score	STEN
OVERALL	152	8
Numbers	30	5
Addresses	28	5
Codes	94	10

	Raw Score	STEN
Language Ability	46	5

	Raw Score	STEN
Numeric Ability	61	7



Norm group used = General Population

This is THE END of the report.

5. INTERPRETATION - THE USE OF NORMS

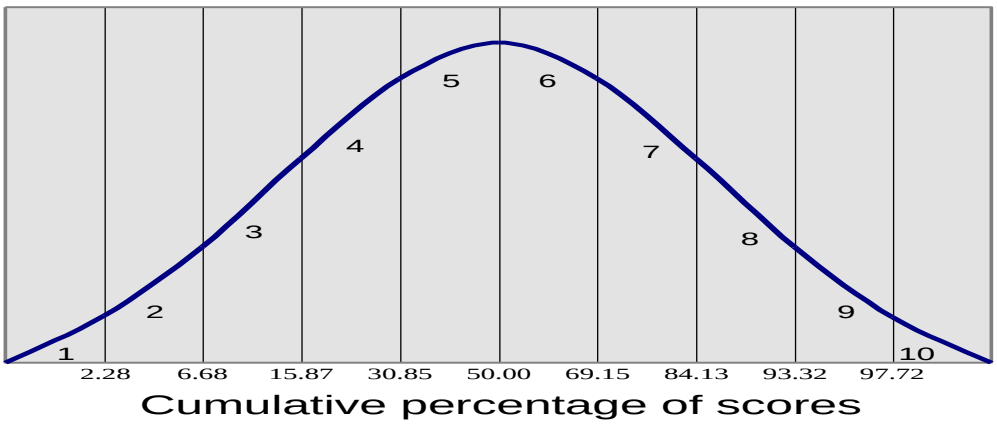
Normative data gives us guidelines against which an individual's scores can be interpreted. To know that an individual's raw score is thirty-seven does not tell us whether this can be regarded as good, bad or indifferent. To interpret a given score correctly we need to compare an individual's scores against the scores of a large group of people. Such a group would be called a norm group.

Norm groups can be drawn from various populations. For example, graduates or clerical staff. Obviously, one can reach entirely different conclusion depending with which norm group an individual is compared. One would expect someone who scores well on the numeric ability module compared with a group of clerical staff to score less well when compared to a group of graduates. For this reason, one should always ensure that the norms being used are relevant to a given situation.

Of the various norm systems that are available, the Adaptive Ability Tests uses one of the standard score systems know as the 'Standard Ten' (STEN). The standard score gives an indication of how far above or below the mean a score is. This interpretation is done in relation to the normal distribution curve. (See Diagram 3). This standardisation of the scores means that a candidate's performance can also be compared between the different ability tests, in terms of norms. That is to say, that an individual that scores a STEN of 7 against a graduate norm on numeracy and a STEN of 4 against a graduate norm on language can be compared directly across these two modules, i.e. that individual is scoring well above average, compared with a graduate group, on numeracy, and below average compared with a graduate group on language.

Another advantage of using the normal distribution as a basis for test norms is that the standard deviation has a precise relationship to the area under the curve. This in turn means we are able to say what percentage of the norm group would achieve a given STEN SCORE. Diagram 3 shows the cumulative percentage of scores. So, for example, if an individual scores a STEN of 4, we would expect 30% of the norm group to score at or below this level.

Diagram 3. STENS and the normal curve



Norms are provided as part of the software package. However, if it possible for the user to generate their own norms. This facility enables the user, for example, to generate norms that would then be available as part of the software. Such norms could be generated for different job groups, for example, technical staff or management trainees. This gives the user norm groups which are directly relevant, hence improving the accuracy of their decision making.

6. AVAILABLE NORMS

This section gives information about the norms that are currently available for the Adaptive Ability Tests. These are provided as part of the software and are reproduced here for reference.

6.1 Clerical Norms

TABLE 1 : NORMS FOR OPERATIVES

Sample Size: 256

Sample Description: Employees from a large securities company in UK

ABILITY MODULE	SAMPLE SIZE	STEN SCORES										MEAN	STANDARD DEVIATION
		1	2	3	4	5	6	7	8	9	10		
LANGUAGE	256	0-22	23-24	25-28	29-35	36-43	44-51	52-68	69-90	91-126	127+	53.77	26.44
NUMERIC	180	0-12	13-14	15-18	19-25	27-37	38-45	46-54	55-60	61-92	93+	42.42	18.70
ADMINISTRATIVE													
NUMBERS	139	0-9	10-13	14-21	22-27	28-34	35-38	39-48	49-52	53-56	57+	37.02	12.62
ADDRESSES	140	0-9	10-13	14-15	16-20	21-24	25-29	30-42	43-49	50-59	60+	30.70	13.82
CODES	140	0-6	7-10	11-14	15-21	22-26	27-34	35-41	42-68	69-84	85+	34.02	20.30
OVERALL	139	0-22	23-29	46-51	52-70	71-04	95-105	106-146	147-174	175-188	189+	104.14	45.37

6.2 Graduate Norm

TABLE 2 : NORMS FOR GRADUATE POPULATION

Sample Size: 445

Sample Description: the sample was drawn from eleven different universities and polytechnics across the UK. the majority of the sample was aged between 20 and 22.

ABILITY MODULE	SAMPLE SIZE	STEN SCORES										MEAN	STANDARD DEVIATION
		1	2	3	4	5	6	7	8	9	10		
LANGUAGE	445	0-42	43-52	53-62	63-72	73-82	83-92	93-102	103-112	113-122	123+	82.696	19.723
NUMERIC	445	0-57	58-65	66-73	74-82	83-90	91-98	99-107	108-115	116-123	124+	90.586	16.821
ADMINISTRATIVE													
NUMBERS	445	0-41	42-46	47-51	52-56	57-61	62-66	67-72	73-77	78-82	83+	61.925	10.638
ADDRESSES	445	0-21	22-30	31-40	41-50	51-59	60-69	70-78	79-88	89-98	99+	60.690	19.135
CODES	445	0-27	28-45	46-64	65-83	84-102	103-120	121-139	140-158	159-176	177+	102.797	36.471
OVERALL	445	0-113	114-141	142-168	169-196	197-224	225-252	253-280	281-308	309-335	336+	225.412	54.556

7. RELIABILITY

This chapter is aimed at giving a brief introduction to the reliability of the Adaptive Ability Tests. Please note that complete details can be found in the technical section of the manual.

Reliability and Validity are the two measures which indicate if a test will be effective for the user. For a test to be valid it must be reliable. (Reliability is a measure of the consistency and robustness of a measure).

Reliabilities are always expressed in terms of a correlation coefficient which will line between -1 and +1. We can consider a coefficient of .65 or above as good, and a coefficient of .9 or above as exceptional.

There are several different methods of assessing the reliability of a test. The table below presents the test, re-test coefficients of the three modules of the Adaptive Ability Tests.

TABLE 3 : TEST, RETEST CORRELATIONS OF THE ADAPTIVE ABILITY TESTS
N = 5 COMPOSITE GROUP

	Module	Second Administration		
		Numeric	Language	Administrative
First Administration	Numeric	0.67		
	Language		0.95	
	Administration			0.98

8. BACKGROUND AND DEVELOPMENT

8.1 General

The Adaptive Ability Test modules have been developed using the results of more than 86,000 tests over the past ten years in the regular and territorial armies.

The Regular Army has for many years used aptitude tests to select and allocate personnel to the many specialisations available to entrants. The Army Entrance Tests, used at present have been developed over a number of years, and many thousands of recruits have provided validation information. A parallel version of these tests were designed to enable the TA to recruit at the same level as the Regular Army.

Items included in the Adaptive Ability Tests were developed as part of the above project and trialled with 18,000 applicants prior to inclusion in the Adaptive Ability Tests. Further studies were carried out in schools and UK companies to locate the levels of difficulty and examine their use in industry and commerce. The Army tests were presented in pencil and paper format. Further studies were set up to standardise the tests to be administered and scored by computer. Items and the general format of the tests have been altered to suit this mode of presentation and a simple and pragmatic method of tailoring tests adopted. Further data collection is at present ongoing enabling the use of more widely accepted probability model methodology.

8.2 Background

The Adaptive Ability Tests were developed in a series of stages. From an initial item pool containing items previously tried in the Army Entrance Test project items were selected to represent appropriate difficulty levels and content. These items were trialled in pencil and paper format using samples of personnel employed by some of our clients. It was this first sample which was used to set up the difficulty levels one, two and three on each of the modules. A total sample size of 180 people consisted of:

95 new entrants
85 middle and senior management volunteers

From the initial analyses of these data it became apparent that for the language and numeric modules there may have been a problem discriminating performances at the lower end of the ability range. It was for this reason that a further sample of school children in two Bicester schools

were used to locate the Level 0 Language and Numeric tests. A total sample size of 68 consisted of:

27	St Mary's Junior School 4th year pupils	(age 10 - 11)
41	Bicester Community College 1st year pupils	(age 11 - 12)

The pencil and paper forms of the tests were item analysed using a split half method designed by Nuttall and Skurnik. The method compares favourably with the Kuder-Richardson Formula 20 (KR20) method, and provides additional statistics on each item to enable the comparisons of scores between the 3 Levels of the test. The results of the item analysis of the pencil and paper version are given in Table 4 and the Level 0 schools analysis in Table 5 below.

TABLE 4 : PENCIL AND PAPER ITEM ANALYSIS RESULTS
LEVELS 1, 2 & 3

Test	Sample	Items	Sample Size	Hi/Lo Split	Mean	Standard Deviation	Measure-ments	Reliability Coefficient	Standard Error of Measurement	Sent
Language1	Entry	35	94	25	22.1	8.8	8	0.93	2.4	2.3
Language1	Managers	35	85	22	26.6	5.4	5	0.83	2.2	2.3
Language1	All	35	179	48	23.7	7.6	7	0.90	2.4	2.4
Language2	Managers	35	85	22	18.2	5.6	5	0.80	2.5	2.4
Language3	Graduates	40	19	5	16.0	5.4	5	0.79	2.5	
Numeric 1	Entry	35	95	25	21.8	5.8	5	0.83	2.4	2.4
Numeric 1	Managers	35	85	22	25.0	4.4	4	0.73	2.3	2.3
Numeric 1	All	35	180	48	23.2	5.3	5	0.80	2.4	2.4
Numeric 2	Managers	35	85	22	18.0	7.0	6	0.87	2.5	2.5
Numeric 3	Graduates	50	19	5	19.6	5.4	6	0.84	2.2	
Administrative										
Number 1	All	60	179		42.2	3.1	6	0.82*	1.3	
Number 1	Entry	60	93		40.3	3.6	6	0.82*	1.5	
Number 1	Managers	60	85		45.9	1.9	7	0.85*	0.8	
Number 2	Managers	60	41		44.7	4.9	7	0.85*	1.9	
Number 3	Managers	60	44		38.0	3.7	6	0.82*	1.5	
Address 1	All	60	180		38.4	5.7	7	0.84*	2.3	
Address 1	Entry	60	94		35.8	5.7	7	0.84*	2.3	
Address 1	Managers	60	85		44.2	5.6	6	0.81*	2.4	
Address 2	Managers	60	41		42.3	6.8	6	0.81*	3.0	
Address 3	Managers	60	44		36.5	5.2	7	0.83*	2.1	
Codes 1	All	98	180		46.0	3.0	6	0.79*	1.4	
Codes 1	Entry	98	94		41.6	3.3	6	0.79*	1.5	
Codes 1	Managers	98	85		48.8	2.8	4	0.76*	1.4	
Codes 2	Managers	63	41		40.8	7.7	4	0.76*	3.8	
Codes 3	Managers	48	44		38.9	2.1	3	0.70*	1.2	

* corrected using the Spearman-Brown formula.

TABLE 5 : SCHOOL ITEMS ANALYSIS RESULTS LEVEL 0

Test	Mean	Standard Deviation	Measure-ments	Reliability Coefficient	Standard Error of Measurement
Language All	28.9	4.6	4	0.76	2.3
Language Bicester	29.8	4.0	4	0.72	2.2
Language St Mary's	28.2	5.5	5	0.83	2.3
Numeric All	39.2	8.6	8	0.91	2.6
Numeric Bicester	43.4	6.1	6	0.85	2.4
Numeric St Mary's	35.3	9.1	8	0.91	2.8

8.3 Measuremeters

The levels in the Ability Tests have been derived by using the measuremeter statistic described by Nuttall and Skurnik (1974). The statistic provides an indication of how reliable the test is. The higher the measuremeter the higher the reliability of the test. In addition a further characteristic of this statistic is the relation of one measuremeter to another. A test taken by a group producing four measuremeters could be described as twice as reliable as a test taken by that same group with two measuremeters. This linearity of scale is not possessed by the reliability coefficient, and provides a pragmatic method of linking the three levels of the Adaptive Ability Tests. Nuttall and Skurnik's method of converting scores to grades by dividing by the appropriate measuremeter was adapted to produce a method of weighting scores on more difficult tests to produce higher scores.

This method has been adopted as an interim while development of a fully tailored testing system is developed using Rasch item scaling or other appropriate probabilistic models.

Table 6 shows the measuremeters for the three modules of the Adaptive Ability Tests. Please note that the table gives results for each of the blocks that were developed as part of the computerised version (see below).

The table shows the measuremeter calculations for establishing levels of difficulty The low sample sizes in some cells have produced some misleading results and therefore in the case of block 1 level 3 in the numeric and language module estimates have been made to maintain consistency across levels. The level 0 items were taken by a different sample and were not comparable with the occupational data. These estimated figures will be replaced as further data becomes available.

The average measuremeters are used to adjust the scores prior to comparison with appropriate normative groups. In each case a sub score for each block at

each level is multiplied by the measuremeter for that block. The results of each block are then added to produce a final score.

TABLE 6 : LOCATING LEVEL MEASUREMETERS

Language Measuremeters

	Level 0	Level 1	Level 2	Level 3
Block				
1	1.0*	2.2	3.4	3.0*
2	1.0*	2.1	3.1	4.3
3	2.0*	3.0	3.2	4.6
Total	4.0*	7.3	9.7	11.9
Average	1.0	2.4	3.2	3.9

Locate items 0.65 Measurements 3.0

Numeric Measuremeters

	Level 0	Level 1	Level 2	Level 3
Block				
1	1.0*	3.1	4.0	1.0*
2	1.0*	2.0	3.6	4.4
3	2.0*	1.8	3.2	4.2
Total	4.0	6.9	10.8	9.6
Average	1.0	2.3	3.6	3.2

Locate items 0.62 Measurements 3.0

*Estimated figures

Administrative

Number Measuremeters

Level 1	Level 2	Level 3
0.5 (2.8)	1.0 (3.7)	1.5*

Address Measuremeters

Level 1	Level 2	Level 3
0.5 (1.8)	1.0 (2.2)	1.5*

Codes Measurements

Level 1	Level 2	Level 3
0.5 (3.4)	1.0 (3.7)	1.5*

* Indicates estimate figures.

Figures in brackets are actual, Measuremeters levels are adjusted.

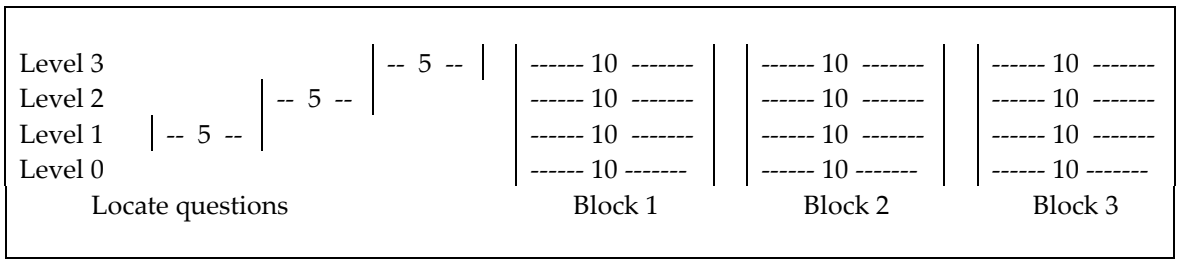
8.4 Computerised Version

From these studies it was possible to construct a computerised version of the tests where each could be sub-divided into three levels in the case of the Administrative Ability tests and four levels in the case of Language and Numeric Ability tests. Within these levels blocks of items were separately tested in groups of 15 to establish whether individuals could be re-allocated to levels during testing. Individuals are allocated to a given difficulty level on the basis of their performance on a series of locator questions at the beginning of the test. The sample used to locate level 0 was different to that used to locate levels 1, 2 and 3. This means that currently the locator questions operate at level 1, 2 and 3. The starting algorithm will be changed as data becomes available to include locator items at level 0. If a person was allocated to Level 1 at the beginning of the test and then performed better than the criterion for that level, then that person could be moved up a level for the second block of items. Similarly if it was then clear that he was struggling at the new level he would drop down for the final block of items.

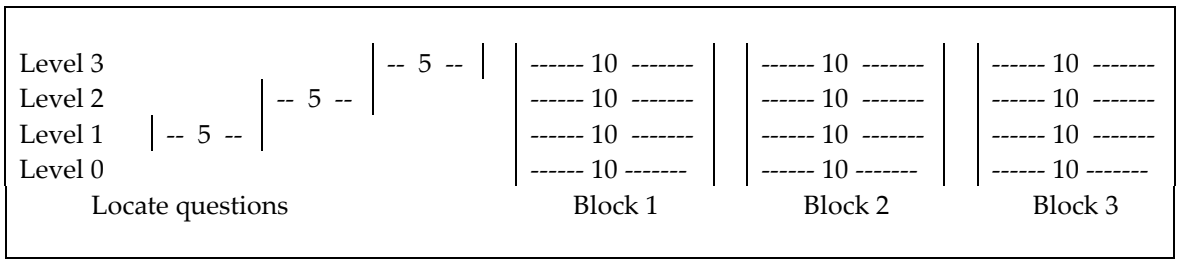
In practice this system was only practical for the Language Ability and Numeric Ability tests. For the Administrative Ability tests there was a problem using these blocks of items. When the candidate changed blocks they also had to receive additional instructions. As a result, they would actually achieve a lower score by being unable to maintain performance at the new level. Therefore, for the Administrative Ability tests, once the candidate is allocated to a particular level they receive all the items at that level. For the Numeric Ability and Language Ability tests, this partial tailoring of testing was trialled in the next stage. Diagram 4 illustrates the structure of the three test modules.

DIAGRAM 4: STRUCTURE OF THE MODULES

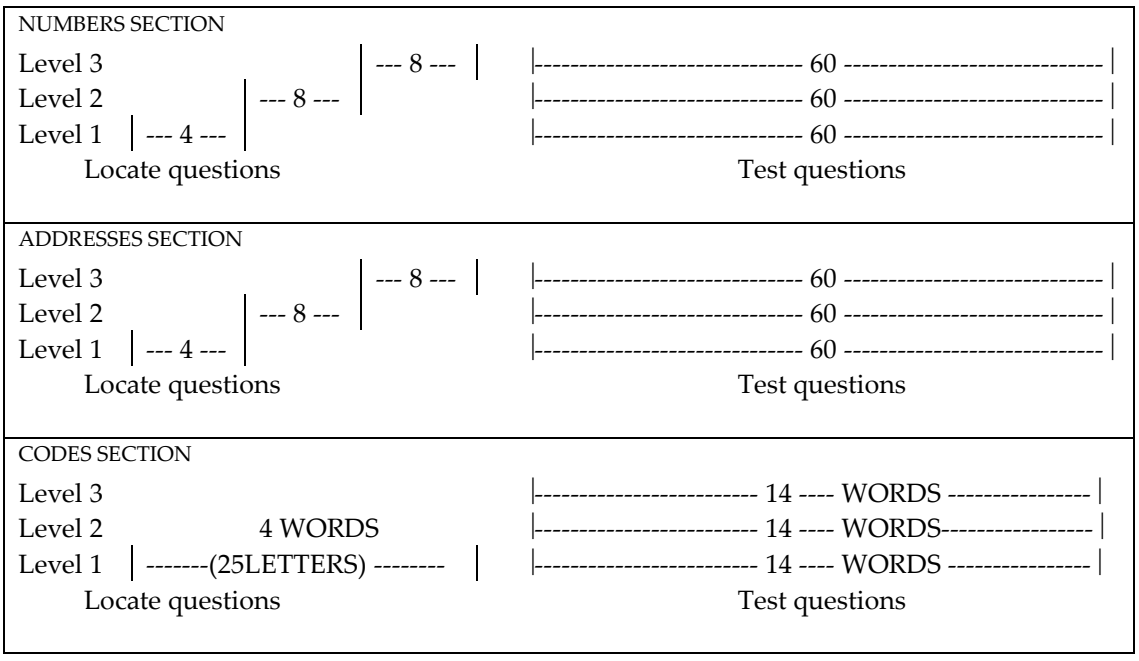
LANGUAGE ABILITY STRUCTURE



NUMERIC ABILITY STRUCTURE



ADMINISTRATIVE ABILITY STRUCTURE



A sample of 260 operatives, clerical, supervisory and management staff were tested on the computerised version. The results of the item analysis for language and numeracy are given in Table 7. The tests were analysed by block of items at each level. Each block consisted of a subtest of 15 items at each of the three levels of supervisory tasks and despite the suitability of the Nuttall technique for small samples results below a sample size of 10 must be considered to be unreliable.

TABLE 7 : ADAPTIVE ABILITY TESTS
THE INTERNAL CONSISTENCY OF TEST SCORES
COMPOSITE OPERATIVE, CLERICAL, SUPERVISING AND MANAGEMENT STAFF
N = 260

LANGUAGE MODULE ITEM ANALYSIS

LEVEL	BLOCK	SAMPLE SIZE	MEAN	STANDARD DEVIATION	RELIABILITY COEFFICIENT	STANDARD ERROR OF MEASUREMENT
1	1	152	6.1	3.16	0.87	1.2
1	2	19	3.5	1.88	0.48	1.4
1	3	30	2.3	1.18	0.14	1.3
2	1	24	6.6	2.79	0.82	1.2
2	2	126	3.7	1.98	0.56	1.3
2	3	35	1.2	0.99	0.04	1.0
3	1	4	2.0	0.82	0.56	1.0
3	2	11	2.8	1.84	0.72	1.0
3	3	7	1.0	0.82	0.28	0.7

NUMERIC MODULE ITEM ANALYSIS

LEVEL	BLOCK	SAMPLE SIZE	MEAN	STANDARD DEVIATION	RELIABILITY COEFFICIENT	STANDARD ERROR OF MEASUREMENT
1	1	235	7.9	1.68	0.49	1.2
1	2	51	5.5	2.51	0.70	1.4
1	3	85	3.8	2.18	0.62	1.4
2	1	18	7.8	1.63	0.53	1.1
2	2	181	4.0	2.33	0.63	1.4
2	3	64	2.4	1.54	0.42	1.2
3	1	NA	Sample too small			
3	2	19	3.5	1.71	0.55	1.2
3	3	19	3.2	2.29	0.72	1.2

Time trials were carried out to establish cut off times for the locator items of the test. The results of these trials are shown in the Table 8.

TABLE 8: LOCATING LEVEL TIME STUDY (TIMES IN SECONDS)

LANGUAGE LEVEL	MEAN	SAMPLE SIZE	STANDARD DEVIATION
1	92	119	47.65
2	95	119	40.45
3	95	16	34.74
NUMERIC LEVEL	**MEAN**	**SAMPLE SIZE**	**STANDARD DEVIATION**
1	188	181	107.56
2	286	181	159.87
3	275	29	152.40
ADMINISTRATIVE TEST	**MEAN**	**SAMPLE SIZE**	**STANDARD DEVIATION**
NUMBERS	159	141	50.76
ADDRESSES	226	141	74.66
CODES	138	141	58.94

8.5 Standardisation of the Tests

A normative study was carried out and cut off scores were calculated based on the performance of the three groups in the sample: managerial, supervisory and operative. These cut off scores are shown in Table 9, for the paper and pencil versions of the test.

TABLE 9: NORMATIVE DATA CUT OFF SCORES

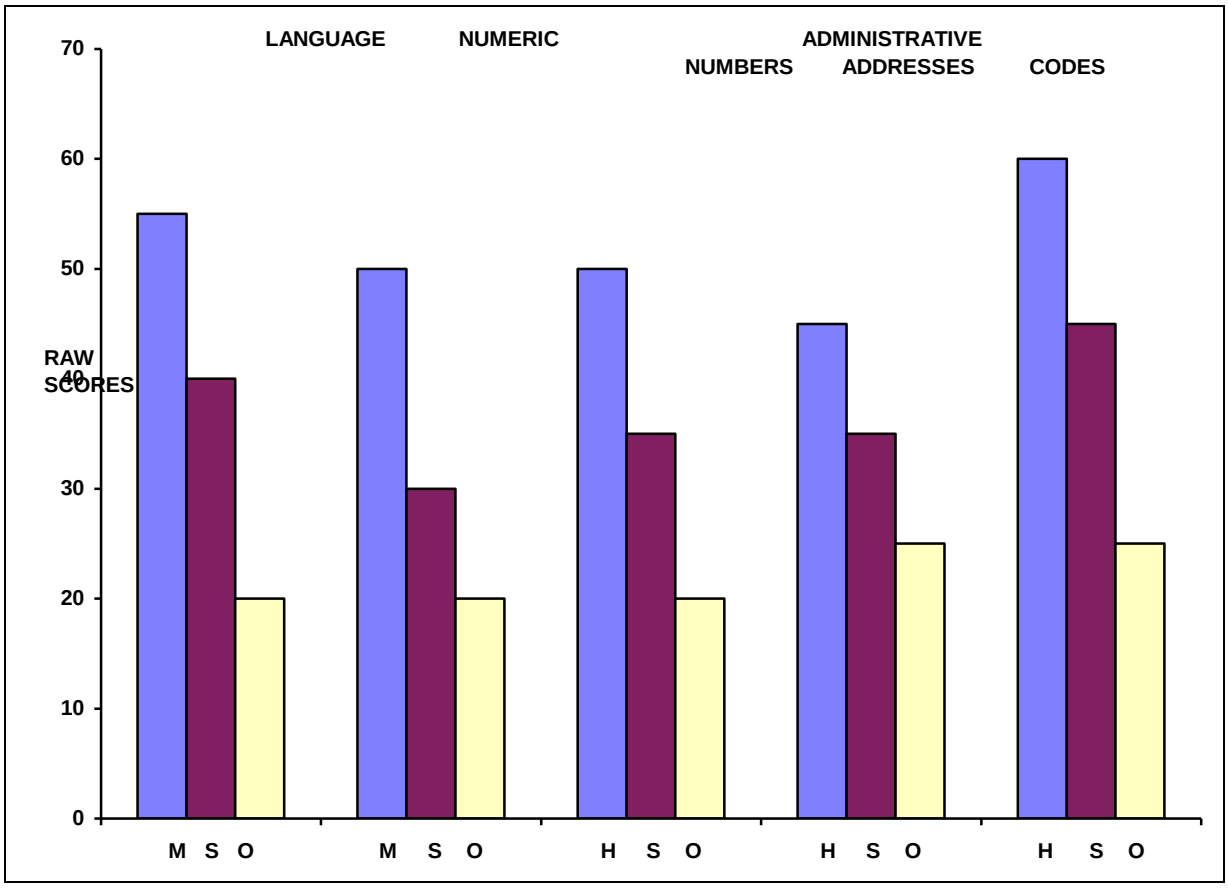
TEST	MANAGERIAL	SUPERVISOR	OPERATIVES
LANGUAGE	50 +	28 - 49	27 - 0
NUMERIC	40 +	26 - 39	25 - 0
ADMINISTRATIVE	*46 +	28 - 45	27 - 0
NUMBERS	*42 +	28 - 41	27 - 0
ADDRESSES	*61 +	35 - 60	32 - 0
CODES			

Scores marked with a * indicate that this is not a management cut off score rather it reflects a very high level of performance. Individuals who score high on the other tests but low on the administrative module are usually preoccupied with precision. In the development of the Differential Aptitude

Test (Psychological Corp.) a group of low performance, high aptitude scorers were retested with new instructions. They were asked to work as fast as possible without worrying about precision. All members of the group improved their scores radically without any appreciable loss of accuracy. You would expect that a low score in these cases has more to do with the work style of the individual rather than the level of performance.

Diagram 5 shows graphically the cut off scores giving an indication of the type of profile you might expect in each case from each of the three groupings. Precise normative information is included in the Appendix.

DIAGRAM 5: TYPICAL PROFILES FOR EACH GROUPING



KEY

M MANAGERS
H HIGH PERFORMANCE GROUP
S SUPERVISORY
O OPERATIVES

Table 17 (see Appendix) gives a summary of the pilot study trials for management, supervisory, operative and school groups in the paper and pencil versions.

Tables 18 to 23 (see Appendix) show frequency tables for each of the test modules from this initial computerised version pilot.

9. RELIABILITY

As with all instruments that are used for measuring, tests always include some error of measurement. Generally this error will arise from three different sources.

i The conditions under which the test is taken may vary from one situation to another.

ii The way the candidate is feeling will vary, e.g. their health.

iii The contents of the test may favour certain individuals.

It is due to these inconsistencies that we need to have statistics which indicate the robustness of a particular measure. In psychological testing this robustness is known as reliability. There are three main types of reliability measure, however, the underlying assumption is the same. If the test is measuring an individual's ability in a reliable manner then that individual's scores will be consistent. Reliability is expressed as a numeric value between -1 and +1. A correlation of +1 describes perfect consistency. The coefficient is calculated by comparing two sets of scores which are achieved by an individual. The source of these scores will vary depending upon which source of inconsistency is being assessed.

A. Test - Retest Reliability

This is a measure of the consistency of the scores that a group of individuals achieve if they complete the tests of two different occasions. This particular type of reliability is aimed at assessing the influence of external factors which may affect a candidates performance. This includes factors such as the conditions under which the test is taken, and how an individual 'feels' on the particular day they take the test. The resulting co-efficient therefore gives information the stability of test scores over time.

It should be noted that the independent variable in test - retest reliability is the length of time that lapses between each test administration. The results may therefore be contaminated by factors such as memory and motivation among the group. However, in the Adaptive Ability Tests, because they are adaptive tests, the questions the candidate answers will change if they are performing at a higher or lower level of ability. This ensures that memory will, at best, only have a marginal effect on a particular candidates performance.

This particular form of reliability does not assess spurious variance which may occur, due to the items chosen to represent the quality being measured. To do this one needs to use parallel forms of a test.

In the parallel - forms procedure, two tests that are equivalent are administered to the same candidates. The tests should be equivalent in the sense that they contain similar items, of equal difficulty, but not the same candidates. Unlike the test re-test procedure, the parallel forms procedure takes into account error variance produced by using different sample of items.

The most desirable way of calculating reliability would be to correlate the scores obtained from a test form on one particular administration, with a parallel form administered to the same group on a subsequent occasion. In this way it is possible to assess errors due to the sample of items used, and those due to the different conditions of administrations.

B. **Internal Reliability**

Internal reliability assessed error that is produced due to the specific set of item that the test contains. Internal reliability can be assessed using parallel forms. It should be noted however, that parallel forms are expensive and frequently difficult to construct.

Typically internal consistency is assessed by splitting the test into two halves and comparing results on each of the split halves, by correlation. Here one can consider each half as being a parallel form. The test results can be split in various ways. For example, the first half may be correlated with the second half, or the results of all the even numbered questions may be compared with the results of all the odd numbered questions. Different splittings of the test may result in slightly different estimates of consistency of an individuals performance on each item of the test are often used.

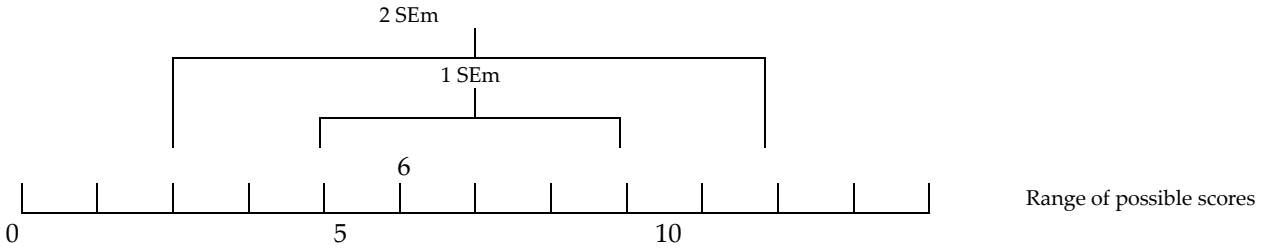
9.1 Standard Error of Measurement

When interpreting an individual's test score knowing that the test has a reliability of 0.8 does not help in deciding for example if a borderline score should fall in one direction or another. The Standard Error of measurement is a statistic which helps to make this decision. The Standard Error of measurement provides a helpful way of interpreting scores. Each score should be thought of as having a margin of error above and below which takes account of the possible error of the scores. This band of uncertainty can be quantified by testing a sample of individuals who are like those you wish to test, and estimating the extent to which their scores varied.

The Standard Error of Measurement (SEm) is estimated by using the formula:

SEm = SD ❾ (l - r) where SD is the Standard Deviation and r is the Reliability

DIAGRAM 6



In the example above one could imagine that an individual has taken a test and achieved a score of 6. From the standard deviation and reliability of this test, as calculated through tests on numerous other individuals it is possible to generate a number which represents the Standard Error of Measurement for this test. This will give a range within which the true score of the individual will fall.

For example:

If SEm = 2 marks then 68% of the time the individual's true score will lie between 4 and 8 i.e., 6 ± 2 marks.

As with any normal distribution, if we were to allow two SEm's either side of the score i.e., 6 ± 4 marks then this is a 95% probability that the individual's true score will lie within this range. In this example the score of six will be affected by factors such as the individual's true score will lie within this range. In this example the score of six will be affected by factors such as the individual's health, the nature of the test administration and whether the test items suited the individual. Hence, the individual's true score independent of these factors, will, 95% of the time, be between 2 and 10.

The use of the SEm is most obvious in borderline cases. However, it is good practice to recognise that the scores you obtain relate to human beings and at all levels score evidence should be sought from other sources such as tests, interviews and so on to take account of this error.

9.2 Standard Error of Differences

In much the same that individual test scores can be differentiated using the Standard Error of Measurement and evaluated to some extent using measuremeters, a more accurate method of determining whether a score on one test is more significant than a score on another test would be to combine the SEm on both.

The Standard Error of Measurement indicated to what extent a test score is a true score for that individual. Similarly the standard error of difference indicates the extent to which, given two scores on two different test, this score difference is an indication of true ability differences. Thus, for example, if an individual gets a score of 6 on the first test and a score of 8 on the second test and the SEd is 2 score units, then you can be sure that 68% of the time there is a true ability difference.

So for example we could use the following formula to give the SEd in score units:

SEd = ❾ (SEm 12 + SEm 22)

The resulting SEd gives us a score value to show the precise difference between a score obtained on one test and the score obtained on one test and the score on another. As with the SEm, 1 SEd suggests that there is a 68% probability that a score obtained on one test is different from a score obtained on another test. 2 SEds provide a probability of 95% that a true ability difference exists. The Standard Error of Differences on the Adaptive Ability Tests are given in Table 10. The lower the score the stronger the relationship between the tests.

TABLE 10 : STANDARD ERROR OF DIFFERENCE BETWEEN TESTS

		LANGUAGE1	NUMERIC1	LANGUAGE2	NUMERIC2	NUMBERS1	ADDRESS1	CODES1	NUMBERS2	ADDRESS2	CODES2	NUMBERS3	ADDRESS3
NUMERIC	1	3.44											
LANGUAGE	2	3.44	3.44										
NUMERIC	2	3.46	3.46	3.44									
NUMBERS	1	2.75	2.75	2.75	2.78								
ADDRESS	1	3.32	3.32	3.32	3.34	2.60							
CODES	1	2.80	2.80	2.80	2.73	3.99	2.65						
NUMBERS	2	3.08	3.08	3.08	3.11	2.00	2.95	2.95					
ADDRESS	2	3.83	3.83	3.83	3.85	3.22	3.72	3.27	3.52				
CODES	2	4.49	4.49	4.49	3.78	3.99	4.40	4.03	4.24	4.76			
NUMBERS	3	2.86	2.86	2.86	2.90	2.00	2.74	2.07	2.45	3.34	4.08		
ADDRESS	3	3.24	3.24	3.24	3.26	2.49	3.11	2.55	2.86	3.65	4.34	2.64	
CODES	3	2.69	2.69	2.69	2.72	1.73	2.54	1.81	2.27	3.18	3.95	1.93	2.43

9.3 Reliability of the Adaptive Ability Tests

Test retest correlations are shown in Table 11. These are preliminary results and it should be noted that the sample size is small. On average the period between each test administration for the sample was a fortnight.

TABLE 11 : INTERNAL CONSISTENCY OF THE ABILITY TESTS TEST, RETEST CORRELATIONS N=16 COMPOSITE GROUP

	Module	Second Administration Numeric	Language	Administrative
First Administration	Numeric	.65		
	Language		.93	
	Administration			.87

10. INTER-RELATIONSHIPS OF THE ADAPTIVE ABILITY TEST MODULES

The intercorrelations of the three modules of the Adaptive Ability Tests are shown below. Additional correlations are provided for the three sections of the administrative ability module.

TABLE 12 : Intercorrelation matrix locator items
The results are shown in Table 6
Language, Numeric & Administrative modules
Graduates (N=445)

	Numeric Module	Administrative Module
Language module	32	34
Numeric module		48

TABLE 13 : Intercorrelations matrix
Administrative module & Administrative module sections
Graduates (N=445)

	Addresses Section	Codes Section	Administrative Module
Numbers Section	35	39	57
Addresses Section		54	78
Codes Section			93

TABLE 14 : Intercorrelation matrix
The three ability test modules & administrative module sections
Language, Numeric & Administrative modules
Graduate (N=445)

	Numeric Module	Administrative Module	Numbers Section	Addresses Section	Codes Section
Language Module	32	34	23	33	27
Numeric Module		48	30	40	43
Administrative Module				57	78
- Numbers section					35
-Addresses section					54

11. VALIDITY

Broadly speaking, the validity of a test is an indication of how much a test is measuring what it is supposed to be measuring. Again there are a number of different types of validity depending on what aspect the user is looking at. The most common forms of validity are as follows:-

A. Face Validity

This is a measure of how much a test appears to measure what it is supposed to measure. Ability tests will generally have a high face validity, for example a candidate will clearly understand that a question about percentages will be assessing their numeric ability. Face validity is important with respect to how the candidate deals with the tests and the thoughts they have during the test sessions. If a candidate perceives a test as having low relevance this can lead to lower motivation and also scepticism.

B. Content Validity

This is a measure of whether the test items are representative of the domain that is being measured. For example, a language ability test containing only items on spelling would be poor content validity if the purpose of the test was to measure general language ability. If the contents of the test are well designed, then the candidates responses to the items on the test would be representative of what their responses would be to the domain of situations sampled by the test.

C. Predictive and Concurrent Validity

Both these types of validity are concerned with the relationship between the test scores and a criterion measure (a standard or variable against which test performance can be evaluated). Typically in occupational testing the criterion will often be successful in the job. Needless to say this approach requires us to be able to define what can be regarded as success in the job, and also that we are able to measure this accurately.

Predictive validity takes a longitudinal approach to this and is the most relevant to the selection process. In this case, test scores are correlated with some measure of the criterion which will not become available until some time in the future. Such criteria may be ratings of an individual, for example, from a supervisor, or they may be a performance measure, for example, sales figures. Concurrent validity is a cruder method of assessing a criterion related validity. In this instance the test is administered to a group who are already

'in post'. Typically the sample would be split between good and poor performers. The validation would assess how well the test scores fit the job performance of these individuals. The problem with concurrent validity studies is that the sample which is assessed will not be typical of the sample upon which the tests would normally be used. A sample of individuals who are already working in a job will already have undergone some self selection, e.g. individuals who were not happy in the job would have left, and all the sample will be higher performers by virtue of having been in the job for a period of time.

D. **Construction Validity**

This type of validity is aimed at relating test scores to the underlying characteristic or trait that is being measured. This will generally be done by relating the test to a measure of behaviour where the construct is thought to be an important variable. For example a test that measures a certain skill would be correlated with a behaviour which could be regarded as typical for that skill. Here, one needs to consider all types of validity including concurrent, predictive and content. Generally the construct validity of a test is established through a slow laborious process of gathering evidence from many experiments and observations on how the test is functioning.

Construct validity can be also be assessed by correlations with other tests or measures of the same characteristic. Thus, one particularl language ability test should correlate with another test of language ability.

11.1 Factors Influencing the Utility of a Test

As well as reliability and validity there are other factors which help us to assess the usefulness of a selection device. These include the appropriateness of the criterion as well as characteristics of the group who are applying for the job and the group who are currently in the job.

A. **Criterion**

The criterion that is used should be relevant an reliable. The criterion should be reliable in that it should not be affected over time, it should also be able to discriminate appropriately e.g. between high and low performers. The criteria should also be relevant in that it is valid for its purpose This is to say, to what extent is the actual criteria that is being used e.g. sales figures relevant to the ultimate criterion e.g. success as a sales manager.

In real terms, the test user also needs to consider how acceptable a criteria is, how much it will cost to measure, how representative it is and how measurable it is, among other factors.

B. **Group Characteristics**

There are two factors that the user needs to consider with respect to the group from which the selection is being made, and the group who are currently 'in post'.

The selection ratio is the ratio of the number of openings to the number of applicants. Whenever this ratio is less than one then the employer can be selective i.e. There are more applicants than openings. The general principle is that the lesser the selection ratio is then the better will be the quality of the selected employees, as long as the relationship between the selection test and the criterion of success in the job is a value greater than zero. This means that if one hundred applicants apply for a single job, then a test of low validity e.g. 0.1 will still be perfectly valid given that the test is reliable. The relationship between test validity and selection ratio to percentage of selected applicants who will be successful can be found in McCormick & Ilgen, (Industrial Psychology, 7th Edition, 1980, p133).

The Base Rate is the proportion of individuals who can be regarded as currently successful in the job for which selection is going on. For a given validity, the fewer people who are currently successful in the job, the larger percentage increase in satisfactory applicants will occur with the introduction of a predictor.

11.2 Relationship with Other Tests

The relationship between the Adaptive Ability Tests and other tests are given in the tables below.

Initial data is given for three of the Personnel Test Battery tests produced by Saville and Holdsworth Ltd in the UK. They are :

A. **PTB Numerical Computation NP2**

This test measures the four basic arithmetic operations of addition, subtraction, division and multiplication. Items include fractions, decimals and percentages, but avoids more complex mathematical areas such as geometry and algebra.

B. **PTB Verbal Meaning VP5**

The VP5 is primarily concerned with the meaning of words and the relationship between them. The vocabulary is non-specialist and is aimed in a broad sense at verbal communication skills.

C. **PTB Checking CP3**

This test was designed to measure speed and accuracy in checking written information. The task requires the candidate to check a printed list again a hand-written list and to spot any errors between the two lists. The list contains words, numbers, letters and symbols.

TABLE 15 : Correlations with other tests
Group = General Populations

Module	Other Test	Correlation Coefficient	Group		N
Numeric	NP2	.55	General	Population	21
	VP5	.03	"	"	20
	CP3	.07	"	"	19
Language	NP2	.77	"	"	21
	VP5	.66	"	"	20
	CP3	.47	"	"	19
Administrative	NP2	.56	"	"	19
	VP5	.05	"	"	18
	CP3	.52	"	"	17

NB the decimal point has been omitted.

11.3 Relationship between the Adaptive Ability Tests and Educational Examination Attainment

This section of the technical manual gives data about the relationship between performance of a group of graduate level candidates and their GCE 'O' level and 'A' level results.

11.3.1 GCE 'O' Levels

The correlations between the Adaptive Ability Test and GCE 'O' levels is given in Table 15. The sample was a group of second and third year undergraduates who had taken their 'O' levels four or five years earlier in most cases. The Ability Test modules were correlated with 'O' level points score (where a grade A = 3 points, grade B = 2 points, and grade C = 1 point) the grade at which 'O' level maths was passed and the grade at which 'O' level English language was passed.

TABLE 16 : CORRELATION COEFFICIENTS BETWEEN THE ADAPTIVE ABILITY TESTS AND GCE 'O' LEVEL EXAMINATIONS

Module	'O' level Points Score	Maths Grade	English Language Grade
Numeric	.26	.31	.09
Language	.26	.09	.21
Administrative	.25	.33	.16
- Numbers	.16	.26	.17
- Addresses	.23	.27	.10
- Codes	.22	.30	.15

11.3.2 GCE 'A' levels

The correlation between the Adaptive Ability Tests and GCE 'A' levels is given in Table 17. The sample was used. This means that the results of the ability tests have been correlated with 'A' level performance taken, on average two or three years earlier. Results were correlated with the number of 'A' levels points score where grade A = 5 points, B = 4 points, C = 3 points, D = 2 points and E = 1 point, and the number that were passed to grade 'A'.

TABLE 17 : CORRELATION COEFFICIENTS BETWEEN THE ADAPTIVE ABILITY TESTS AND GCE 'A' LEVEL EXAMINATIONS

Module	'A' Level Point Score	Number at Grade A
Language	.12	.12
Numeric	.19	.18
Administrative	.18	.13
- Numbers	.12	.11
- Addresses	.13	.07
- Codes	.17	.13

Table 18 : Pilot Study Trials For Management, Supervisory, Operative And School Sample Groups On Paper And Pencil Versions

Test	Sample	Mean	Sample Size	Variance	Standard Deviation
NUMERIC	Management	47.42	7	37.62	6.1
	Supervisory	30.95	63	10.05	3.2
	Operatives	16.76	110	34.57	5.9
LANGUAGE	Management	62.20	15	132.31	11.5
	Supervisory	35.53	73	30.81	5.6
	Operative	19.36	169	27.78	5.3
NUMBER	Management	53.54	35	32.55	5.7
	Supervisory	36.83	69	14.51	3.8
	Operative	20.44	33	33.83	5.8
ADDRESS	Management	50.87	33	63.55	7.9
	Supervisory	33.93	30	22.18	4.7
	Operative	20.25	74	26.69	5.2
CODES	Manager	74.10	19	84.54	9.2
	Supervisory	39.63	46	34.46	5.9
	Operative	20.44	75	114.60	10.7
SCHOOLS					
NUMERIC	Combined	40.58	68	72.78	8.5
	Bicester	44.19	41	37.86	6.2
	St Mary	35.11	27	77.64	8.8
LANGUAGE	Combined	29.29	67	23.52	4.9
	Bicester	29.90	40	15.43	3.9
	St Mary	28.40	27	35.17	5.9

TABLE 19 : FREQUENCY DISTRIBUTION COMPUTERISED LANGUAGE MODULE

Score	Frequency	Cumulative Frequency	Percentile	T Score	Z Score	Sten	
175	2	256	99	105.30	5.53		
174	1	254	98	95.47	4.54		
173	1	253	98	95.09	4.50		
170	2	252	98	93.95	4.50	10	
146	1	250	97	84.88	3.48		
128	1	249	97	78.07	2.80		
126	1	248	96	77.31	2.73		
124	1	247	96	76.56	2.65		
123	1	246	95	76.18	2.61		
118	2	245	95	74.29	2.42		
110	1	243	94	71.26	2.12		
106	2	242	94	69.75	1.97	9	
104	1	240	93	68.99	1.89		
100	2	239	93	67.48	1.74		
97	1	237	92	66.34	1.63		
94	5	236	92	65.21	1.52		
93	2	231	90	64.83	1.48		
88	3	229	89	62.94	1.29		
82	1	226	88	60.67	1.06		
81	3	225	87	60.29	1.02		
80	2	222	86	59.91	0.99		
76	2	220	85	58.40	0.84		
74	4	218	85	57.64	0.76	8	
73	3	214	83	57.27	0.72		
72	2	211	82	56.89	0.68		
71	2	209	81	56.51	0.65		
70	6	207	80	56.13	0.61		
69	3	201	78	55.75	0.57		
68	2	198	77	55.38	0.53		
66	4	196	76	54.62	0.46		
65	3	192	75	54.24	0.42		
64	3	189	73	53.86	0.38		
63	3	186	72	53.48	0.34		
62	4	183	71	53.11	0.31		
61	6	179	69	52.73	0.27	7	
60	2	173	67	52.35	0.23		
59	2	171	66	51.97	0.19		
58	2	169	66	51.59	0.15		
57	3	167	65	51.22	0.12		
56	1	164	64	50.84	0.08		
55	1	163	63	50.46	0.04		
54	1	162	63	50.08	0.00		
53	7	161	62	49.70	-0.02		
52	5	154	60	49.32	-0.06		
51	5	149	58	48.95	-0.10		
50	8	144	56	48.57	-0.14		
49	8	136	53	48.19	-0.18		
48	1	128	50	47.81	-0.21	6	
47	3	127	49	47.43	-0.25		
46	11	124	48	47.05	-0.29		
45	6	113	44	46.48	-0.33		
44	10	107	41	46.30	-0.36		
43	6	97	37	45.92	-0.40		
42	7	91	35	45.54	-0.44		
41	5	84	32	45.16	-0.48		
40	6	79	30	44.79	-0.52	5	
39	4	73	28	44.41	-0.55		
38	6	69	26	44.03	-0.59		
37	4	63	24	43.65	-0.63		
36	3	59	23	43.27	-0.67		
35	3	56	21	42.89	-0.71		
34	4	53	20	42.52	-0.74		
33	4	49	19	42.14	-0.78		
32	3	45	17	41.76	-0.82		
31	5	42	16	41.38	-0.86	4	
30	4	37	14	41.00	-0.89		
29	5	33	12	40.63	-0.93		
28	4	28	10	40.25	-0.97		
27	5	24	9	39.87	-1.01		
26	5	19	7	39.49	-1.05	3	
25	2	14	5	39.11	-1.08		
24	5	12	4	38.73	-1.12		
23	4	7	2	38.36	-1.16	2	
22	3	3	1	37.98	-1.20		

MEAN 53.77 SAMPLE SIZE = 256

VARIANCE 699.3602 STANDARD DEVIATION : 26.44

TABLE 20 : FREQUENCY DISTRIBUTION COMPUTERISED NUMERIC MODULE

Score	Frequency	Cumulative Frequency	Percentile	T Score	Z Score	Sten
103	1	180	99	82.39	3.63	
100	2	179	99	81.74	3.17	
99	1	177	98	81.19	3.11	10
95	1	176	97	79.01	2.90	
93	1	175	97	77.92	2.79	
86	1	174	96	74.09	2.40	
77	1	173	96	69.17	1.91	
75	1	172	95	68.08	1.80	
72	1	171	95	66.44	1.64	
70	1	170	94	65.35	1.53	
68	1	169	93	64.25	1.42	9
66	2	168	93	63.16	1.31	
63	4	166	92	61.52	1.15	
61	4	162	90	60.43	1.04	
59	11	158	87	59.34	0.93	8
57	4	147	81	58.24	0.82	
55	7	143	79	57.15	0.71	
54	4	136	75	56.60	0.66	
52	9	132	73	55.51	0.55	
50	6	123	68	54.42	0.44	7
48	6	117	65	53.33	0.33	
46	4	111	61	52.23	0.22	
45	8	107	59	51.69	0.16	
43	10	99	55	50.59	0.05	
41	4	89	49	49.50	-0.04	6
39	12	85	47	48.41	-0.15	
37	6	73	40	47.31	-0.26	
36	4	67	37	46.77	-0.32	
34	7	63	35	45.68	-0.43	
32	4	56	31	44.58	-0.54	
30	3	52	28	43.49	-0.65	5
28	4	49	27	42.40	-0.75	
27	5	45	25	41.85	-0.81	
25	5	40	22	40.76	-0.92	
23	6	35	19	39.66	-1.03	
21	5	29	16	38.57	-1.14	4
19	5	24	13	37.48	-1.25	
18	7	19	10	36.93	-1.30	
16	4	12	6	35.84	-1.41	3
14	6	8	4	34.75	-1.52	
12	2	2	1	33.65	-1.63	1

MEAN 42.42 SAMPLE SIZE = 180
VARIANCE 349.7 STANDARD DEVIATION : 18.7

TABLE 21 : FREQUENCY DISTRIBUTION COMPUTERISED ADMINISTRATIVE MODULE, OVERALL

Score	Frequency	Cumulative Frequency	Percentile	T Score	Z Score	Sten
228	1	139	99	77.73	2.77	
227	1	138	99	77.07	2.70	
218	1	137	98	75.09	2.50	
195	1	136	97	70.02	2.00	10
189	1	135	97	68.70	1.87	
187	1	134	96	68.26	1.82	
186	1	133	95	68.04	1.80	
185	1	132	94	67.82	1.78	
183	1	131	94	67.38	1.73	
182	1	130	93	67.16	1.71	9
181	1	129	92	66.93	1.69	
179	1	128	92	66.49	1.64	
177	1	127	91	66.05	1.60	
175	1	126	90	65.61	1.56	
172	1	125	89	64.95	1.49	
170	1	124	89	64.51	1.45	
168	1	123	88	64.07	1.40	
166	1	122	87	63.63	1.36	
164	1	121	87	63.19	1.31	
163	1	120	86	62.97	1.29	
162	3	119	85	62.75	1.27	8
161	1	116	83	62.53	1.25	
159	1	115	82	62.09	1.20	
156	1	114	82	61.42	1.14	
155	1	113	81	61.20	1.12	
154	1	112	80	60.98	1.09	
150	1	111	79	60.10	1.01	
148	1	110	79	59.66	0.96	
147	1	109	78	59.44	0.94	
145	1	108	77	59.00	0.90	
140	1	107	76	57.90	0.79	
135	1	106	76	56.80	0.68	
132	1	105	75	56.13	0.61	
129	1	104	74	55.47	0.54	
126	1	103	74	54.81	0.48	
125	1	102	73	54.59	0.45	
123	1	101	72	54.15	0.41	
119	1	100	71	53.27	0.32	
118	1	99	71	53.05	0.30	
117	1	98	70	52.83	0.28	7
115	1	97	69	52.39	0.23	
114	1	96	69	52.17	0.21	
113	1	95	68	51.95	0.19	
112	4	94	67	51.73	0.17	
111	1	90	64	51.51	0.15	
110	1	89	64	51.29	0.12	
108	1	88	63	50.84	0.08	
107	3	87	62	50.62	0.06	
106	2	84	60	50.40	0.04	
105	4	82	58	50.18	0.01	
104	3	78	56	49.96	-0.00	
103	3	75	53	49.74	-0.02	
102	1	72	51	49.52	-0.04	
101	3	71	51	49.30	-0.06	6
100	3	68	48	49.08	-0.09	
99	3	65	46	48.86	-0.11	
98	3	62	44	48.64	-0.13	
97	1	59	42	48.42	-0.15	
96	1	58	41	48.20	-0.17	
95	2	57	41	47.98	-0.20	
94	2	55	39	47.76	-0.22	
93	1	53	38	47.54	-0.24	
92	2	52	37	47.32	-0.26	
91	1	50	35	47.10	-0.28	
90	2	49	35	46.88	-0.31	
89	1	47	33	46.66	-0.33	
88	1	46	33	46.44	-0.35	
87	1	45	32	46.22	-0.37	
85	1	44	31	45.78	-0.42	
84	1	43	30	45.56	-0.44	
83	1	42	30	45.33	-0.46	5
81	1	41	29	44.89	-0.51	
80	1	40	28	44.67	-0.53	
79	1	39	28	44.45	-0.55	
78	1	38	27	44.23	-0.57	
77	1	37	26	44.01	-0.59	
76	1	36	25	43.79	-0.62	
75	1	35	25	43.57	-0.64	
74	1	34	24	43.35	-0.66	
72	1	33	23	42.91	-0.70	
71	1	32	23	42.69	-0.73	
70	1	31	22	42.47	-0.75	
68	1	30	21	42,.03	-0.79	
67	1	29	20	41.81	-0.81	
65	1	28	20	41.37	-0.86	
63	1	27	19	40.93	-0.90	
61	1	26	18	40.49	-0.95	
60	1	25	17	40.27	-0.97	
59	2	24	17	40.04	-0.99	
58	1	22	15	39.82	-1.01	
57	2	21	15	39.60	-1.03	4
56	1	19	13	39.38	-1.06	
55	1	18	12	39.16	-1.08	
54	1	17	12	38.94	-1.10	
52	1	16	11	38.50	-1.14	
51	1	15	10	38.28	-1.17	
49	1	14	10	37.84	-1.21	
48	1	13	9	37.62	-1.23	
45	1	12	8	36.96	-1.30	
42	1	11	7	36.30	-1.36	
37	1	10	7	35.20	-1.47	
35	1	9	6	34.75	-1.52	3
33	1	8	5	34.31	-1.56	
30	1	7	5	33.65	-1.63	
29	1	6	4	33.43	-1.65	
28	1	5	3	33.21	-1.67	
25	1	4	2	32.55	-1.74	
24	1	3	2	32.33	-1.76	
22	1	2	1	31.89	-1.81	2
20	1	1	0	31.45	-1.85	1

MEAN 104.14 SAMPLE SIZE = 139
VARIANCE 2058.5588 STANDARD VARIATION : 45.37

TABLE 22 : FREQUENCY DISTRIBUTION COMPUTERISED ADMINISTRATIVE MODULE, NUMBERS SECTION

Score	Frequency	Cumulative Frequency	Percentile	T Score	Z Score	Sten
171	1	139	99	76.91	2.69	
170	1	138	98	76.12	2.61	
67	1	137	97	73.74	2.37	
57	2	136	97	65.82	1.58	10
56	3	134	95	65.03	1.50	
55	4	131	93	64.24	1.42	9
54	1	127	90	63.44	1.34	
53	4	126	90	62.65	1.26	
52	5	122	87	61.86	1.18	
51	2	117	83	61.07	1.10	8
50	4	115	82	60.27	1.02	
49	3	111	79	59.48	0.94	
48	2	108	77	58.69	0.86	
47	1	106	75	57.90	0.79	
46	1	105	75	57.10	0.71	
45	2	104	74	56.31	0.63	
44	1	102	72	55.52	0.55	
43	3	101	72	54.73	0.47	
41	6	98	69	53.14	0.31	7
40	3	92	65	52.35	0.23	
39	6	89	63	51.56	0.15	
38	11	83	59	50.76	0.07	
37	11	72	51	49.97	-0.00	6
36	3	61	43	49.18	-0.08	
35	7	58	41	48.39	-0.16	
34	5	51	36	47.60	-0.23	
33	2	46	32	46.80	-0.31	
32	3	44	31	46.01	-0.39	
31	3	41	29	45.22	-0.47	5
30	3	38	27	44.43	-0.55	
29	2	35	25	43.63	-0.63	
28	2	33	23	42.84	-0.71	
27	2	31	22	42.05	-0.79	
26	2	29	20	41.26	-0.87	
25	5	27	19	40.46	-0.95	
24	4	22	15	39.67	-1.03	4
23	2	18	12	38.88	-1.11	
22	2	16	11	38.09	-1.19	
21	2	14	10	37.29	-1.27	
18	1	12	8	34.92	-1.50	
17	2	11	7	34.12	-1.58	
15	2	9	5	32.54	-1.74	3
14	3	7	5	31.75	-1.82	
13	3	4	2	30.95	-1.90	
9	1	1	0	27.29	-2.22	1

MEAN 37.02 SAMPLE SIZE = 139
VARIANCE 159.29 STANDARD DEVIATION : 12.62