Person. individ. Diff. Vol. 11, No. 3, pp. 239–257, 1990
Printed in Great Britain. 

0191-8869/90 $3.00 + 0.00


# THE FURNEAUX MODEL OF HUMAN PROBLEM SOLVING: ITS RELATIONSHIP TO REACTION TIME AND INTELLIGENCE

W. FREARSON, H. J. EYSENCK and P. T. BARRETT
Institute of Psychiatry, University of London, London, England

(Received 9 July 1989)

**Summary**—One-hundred and nine adult subjects were tested on two occasions using a set of computer administered letter series problems. Using methods introduced by Furneaux [In Eysenck, H. J. (Ed.), *Handbook of abnormal psychology*. New York: Basic Books, 1961], response times to items were analysed in order to provide estimates of item difficulties based on time to correct solution, rather than on probability of correct solution. Subject parameters of speed, continuance and accuracy were also calculated using both Furneaux (directly and indirectly) and classical methods. These 'time based' parameters were compared with subscale scores from the WAIS-R IQ test and parameters from two reaction time tasks: 1 and 2-bit choice reaction time and the 'Odd-Man-Out' task. It was concluded that the original Furneaux model performed sub-optimally with regard to parameter estimation. However, with some modifications made to these estimation procedures, the Furneaux parameters were shown to be both practical and useful estimates of subject performance on letter series problems.

## INTRODUCTION

The investigation of the role played by individual differences in abilities that have been related to 'speed' has been carried out either by factor analytic studies of test batteries (some of which are more 'speeded' than others) or by the measurement of times to solution of either individual items or blocks of items. The search for a speed factor independent of '*g*' was motivated by the perception that the introduction and extensive use of time limit tests (e.g. the Army Alpha) could be handicapping the 'slow but profound' *S* . Much early work on the *S* was concerned with investigating the differential effect time limits of varying severity had upon individuals' performance. May (1921) reported a high correlation between scores on Army Alpha with both the standard time limit and with double the limit. Recently, Berger (1982) has given a detailed description of the 'speed vs power' controversy.

The introduction of factor analysis enabled experimenters to search for a factor which was separate from *g* and on which the speeded tests of the battery loaded more heavily than the unspeeded. The first of such studies was by Sutherland (1934) who used factor analytic techniques to extract a speed factor independent of *g* from a test battery that included scores derived from the times taken to complete groups of tests. Mangan (1958) performed a similar analysis, again finding a separate speed factor (among others) which he included in a model of individual differences as a subsidiary to *g*.

The existence of such a factor demonstrated that there was a psychometric factor underlying success on the speeded tasks of the battery and less so on the unspeeded. How such a factor was related to the supposed internal properties of the individual *S*s could not be determined. Factor analysis of a battery of tests is, by its nature, unable to explain any more about test taking performance than the experimenter is willing to infer from a knowledge of the properties of the tests. Psychometric validity does not always equate with psychological validity. With the growing sophistication in the use of factor analysis and the availability of computers it was possible for Lord (1956) to rotate a set of factors derived from a battery of tests some 500 times where 'the main guiding principle in all rotations was psychological meaningfulness, as interpreted according to the notions of the writer'.

It was frustration with the methods of traditional psychometrics that led Eysenck (1953) to propose a new form of analysis of test taking performance that had the individual item as the fundamental unit of analysis rather than the test score. For the analysis of the processes underlying the simple total correct score, more information other than simple right or wrong must be derived from each item. The most readily available information is the time to solution (or to failure or abandonment) on each item. He also suggested that IQ scores could be analysed in terms of their three major components (speed, persistence and accuracy) which might be largely independent. Some early work by Thorndike especially had previously indicated the possibility of extracting multiple measures from cognitive test performance. Thorndike, Bregman, Cobb and Woodyard (1927) suggested that two alternative approaches to a *S*'s performance on an item were possible. The first suggestion (and the one which was subsequently pursued to the exclusion of the other) was that a *S* should be scored on the basis of items answered correctly, with the time taken to answer each item being of only secondary importance.

The second suggestion involved the time taken to solve an item. It may be proposed that individuals can differ in the speed at which they can complete an item in a similar way to their differing in probability of giving a correct response. A test can then be imagined in which all individuals can be expected to answer all the problems correctly, measures of individual difference being derived from the time each *S* takes to reach each answer. In such a test the data from an individual item could in fact be used as an indication of intelligence. Analogous to the probabilistic model, the time model allows difficulties to be assigned to items, based either on the average time needed to solve an item for a normal group, or on the time required for a fixed percentage of the group to pass.

A typical battery of ability tests can be viewed in two parallel ways, either taking the number of items an individual passes as the primary data and giving some form of bonus to unusually quick performance, or taking the times to correct solution as the raw data to construct ability scales. The two approaches will lead to two different difficulty scales for the items of the test, one based on the probabilistic model, one on the time model. A relationship between the two sets of difficulties cannot be assumed. Similarly each *S* can be assigned a score or scores based on the probabilistic model, or on the time model. Again no necessary relationship need exist between these two sets of ability scales.

Attempts to measure time based individual differences, and hence simultaneously time based item difficulties were hampered by technical difficulties, both computational and in accurately measuring an individual's performance. Slater (1937) recorded individual times to correct responses and from these derived for each *S* a speed score representing the average deviation of the *S*'s correct response times, from the mean time to solution for each item. Such speed scores from a variety of tests were strongly correlated though they were largely independent from *g*. Tate (1948) performed a similar analysis on individually timed items and again found a *S*'s speed to be independent of test content and item difficulty.

Furneaux (1961) presented a model of human problem solving based upon the concept of a search process with the speed of the search process as one parameter (*speed*), how long the process went on before termination as a second (*persistence*), and how accurately the end of the search could be determined as a third individual difference parameter (*accuracy*). Furneaux presented an algorithm by which the difficulty of an item and the speed of the problem solver could be simultaneously estimated from the times to correct solution of items. Furneaux's main experimental results were concerned with showing that this time-based difficulty remained a constant property of an item and was not dependent upon the characteristics of the sample.

Attempts to replicate Furneaux's work were hampered by the difficulties associated with the technique of simultaneous assessment of item difficulty and *S* speed which meant that accurate estimates of difficulty could be made only at the expense of discarding some *S*s' data. Brierly (1969) in a 'manual' replication, stopped his analysis after only the first estimates of item and *S* characteristics had been made. Berger (1976), with the aid of a computer, could make many more iterative estimates of the required parameters but was unable to provide a full set of *S* parameters and also abandoned his analysis at a relatively early stage.

White (1973, 1982) reformulated the basic Furneaux model, particularly altering the significance of the accuracy parameter and introducing the concept of latent traits. He developed a likelihood equation and by finding the set of item and *S* parameters that maximized the likelihood value of this equation, could directly estimate all the required parameters. However, even when this more efficient estimation technique White failed to provide all parameters for all *S*s.

Recent research indicating a relationship between elementary cognitive–sensory perception tasks (choice reaction time and inspection time) and psychometric IQ has led to new theorizing on the nature of IQ differences (Eysenck, 1982). These new speed theories concentrate more on individual differences in the rate at which elementary tasks can be performed than on differences in how elementary tasks are structured to enable problem solving performance. Of course, while these tasks are described as elementary, this is only with respect to a gross rank order of all possible cognitive based tasks from simple sensory perception to non-linear dynamical systems analysis and say fractal geometry.

The relationship between the speed models which explain the differences in intelligence in terms of differences in physiological processes (e.g. speed or accuracy of nervous conduction, oscillation of synaptic conductivity etc.) and those that address differences in test taking styles (fast but inaccurate, persistent but slow etc.) is unknown. Clearly the models could be linked in a hierarchy (efficient physiological processes leading to fast test taking style) or the relationship could be more complex, with underlying physiological properties defining a variable 'cocktail' of the psychological characteristics.

The present study sets out to perform a Furneaux style analysis on data collected from a sample tested on two separate occasions in order to allow estimates of reliability to be made for both *S* and item characteristics. Attempts will be made to produce Furneaux-like parameters which can be computed for all *S*s to give a full set of *S* parameters. These will then be compared with the results from reaction time tasks and WAIS-R general IQ tests administered to the same experimental group.

## METHOD

### *Subjects*

A group of 109 *S*s was recruited from local Government Employment Centres, by advertisements in local newspapers and magazines and from London Mensa (a restricted 'high IQ only' social organisation). All *S*s were smokers. An approximately normal distribution of IQ was obtained by the recruitment of Mensa members and the pre-testing of some *S*s from the job centre to ensure adequate sampling of both high and low extremes of IQ. The total sample distribution of full scale WAIS IQ is shown in Fig. 1 below. The age range of the 70 male *S*s was from 17 to 60 yr with a mean of 28.21 and an SD of 9.53. The age range of the 39 females was from 18 to 44 with a mean of 26.49 and an SD of 7.21.

### *Design*

All *S*s were tested on two consecutive working days (i.e. about one-fifth of the *S*s were tested on a Friday and then on the following Monday) as part of a study investigating the effects of

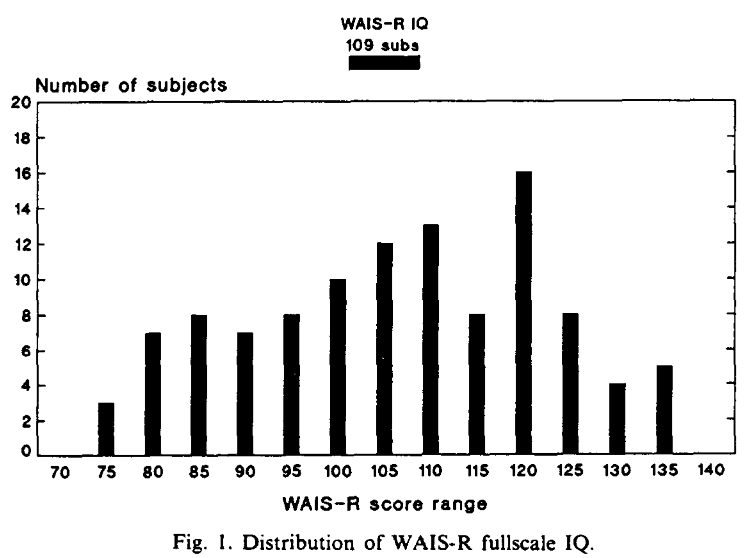


Fig. 1. Distribution of WAIS-R fullscale IQ.

smoking on reaction time tasks and the relation of reaction time performance and IQ, as well as the $S$ of the present paper (further details of experimental design as related to the smoking and reaction time tasks are reported in Frearson, Barrett & Eysenck, 1988).

Each $S$ performed on one of two parallel letter-series tasks on each of the 2 days, the order of the two tasks being counter-balanced so that half the $S$s performed first on Test A and half first on Test B. Each $S$ also performed two reaction time tasks on each day, a version of the Jensen choice reaction time task and the Odd-Man-Out task (Frearson & Eysenck, 1986).

The WAIS-R intelligence test was administered to all $S$s with the sub-tests split over the 2 days of testing. All the $S$s except numbers 1–7 and 9–22 (i.e. 88 out of the 109 $S$s) were also given a 20 min version of the Raven's Advanced Progressive Matrices (Frearson & Eysenck, 1986). Computer administered versions of the Eysenck Personality Questionnaire and the $I_7$ (IVE) (Impulsivity, Venturesomeness, Empathy; Eysenck, Pearson, Easting & Allsopp, 1985) were also given to all $S$s.

*Apparatus*

All experimental control, stimulus presentation and data acquisition was controlled by an ACT Sirius 1 microcomputer. For the two reaction time tasks, signal priming, detection and timing were implemented via a Biodata Microlink unit comprising modules RR8 (8 channel reed relays), two CC8 (8 channel digital inputs) and two TIM (clock module timing in msec, units up to 9999 msec count). The items of the letter series tasks were presented via the Sirius VDU; control of presentation and response detection being implemented directly by the Sirius, using the Biodata TIM module and the Sirius' internal clock to extend the timing range beyond that of the TIM module (up to 9.999 sec then starts again at 0 msec).

*Procedure for the letter series task*

Two parallel sets of letter series problems were created from the Nufferno test (Furneaux, 1955) by taking items of the speed test and the power scale to give a set of 28 items of widely varying difficulty. A second set of items was then created by taking each of the 28 items and repeating the patterning of the letters though starting at a different part of the alphabet. This gave items of the same 'logical' format though physically different. At no point did any $S$ comment on any similarity between any items. The parent and parallel items were assigned at random to the two experimental item sets. Each of the experimental sets was then arranged into cycles of 7 of increasing difficulty using the 'difficulty' assigned in the Nufferno test to the parent. The parent and its parallel item were assigned individually to cycles so they did not appear in the same chronological position in the two sets. Table 1 gives the 56 items along with the difficulties that were assigned to them by the Furneaux procedure (see below).

The $S$ was seated at the Sirius. Each item appeared on the screen along with a display of the alphabet and a message to press either the 'got it' or the 'give up' key as soon as they either knew the answer or wanted to go on to the next item. The 'got it' and 'give up' keys were two of the Sirius' numeric keys with stick-on faces with the legends 'give up' or 'got it'. On pressing the 'got it' key the Sirius put up a new screen asking the $S$ to key in their answer, which was done using the standard QWERTY keyboard. It was explained to the $S$ that this procedure was necessary to "equalize for the effect of some people being more familiar than others with the keyboard". Other than this no mention was made of the fact that $S$s' responses were being timed, $S$s being told only to try their best to get the items right though not to make any 'wild guesses', pressing the 'give up' key if they were doubtful about an answer.

After the $S$s had keyed in their responses the Sirius cleared the screen and displayed a message asking the $S$ to press the spacebar when ready for the next item. This procedure was used to try and stop $S$s taking rest breaks in between items during their nominal solution time, in line with the suggestions of Berger (1982). $S$s who gave up on an item were immediately presented with the 'press spacebar for the next problem' order.

The time between the presentation of the item and the $S$s' pressing either the 'give up' or 'got it' key was recorded, as were the times taken by the $S$ to find the answer key and the length of the rest interval before the $S$ requested the next item.

Table 1 (parts 1 and 2). The 1st and 2nd series of 28 items used in the Furneaux task: their correct solution and the Furneaux derived time based difficulties

| Item | Solution | Difficulty |
|---|---|---|
| *Part 1* | | |
| PQQRRRSSSS? | T | 5.63 |
| DEFUV? | W | 5.45 |
| MNOMNPMNQMNR? | M | 5.61 |
| BEHCFIDG? | J | 6.55 |
| QSTUTUWXYXYABCBC? | E | 7.81 |
| SYXWYXTVUTVUUSR? | Q | 7.15 |
| KLKLLJMLI? | N | 7.72 |
| BWCXBWCXB? | W | 5.99 |
| LZLLYLLLXLLWLL? | V | 5.99 |
| OPQOPQRSTRSTU? | V | 6.03 |
| HZGHZFGHZEFGHZ? | D | 6.46 |
| ZACFJNQS? | T | 7.24 |
| KLMNNOPQRQSTUVTWXYZ? | W | 7.74 |
| LMNJKLMHIJKL? | F | 7.47 |
| MNOMNOPQRP? | Q | 5.74 |
| DCVWBX? | A | 6.18 |
| FXEFXDEFXCDEFX? | B | 6.25 |
| TUVSTUVRSTUV? | Q | 6.03 |
| ZAXZZXZYXZXXZ? | W | 6.63 |
| DPONMEFLKJGHI? | I | 7.42 |
| RSTPQRSNOPQR? | L | 7.38 |
| OPPQRRSTT? | U | 5.88 |
| AXAYBXBYCXCY? | D | 5.73 |
| ·RRSSTUUVVWXXY? | Y | 5.74 |
| LMLMJKNONOJK? | P | 6.53 |
| WEGYIKAMO? | C | 6.98 |
| DEFDFGHIHJKLJLMNO? | N | 7.73 |
| RWSTVUUVWTSR? | X | 8.52 |
| *Part 2* | | |
| GHIGHIJKLJ? | K | 5.54 |
| AYAXBYBXCYCX? | D | 5.93 |
| WXYWXZWXAWXB? | W | 5.92 |
| FYEFYDEFYCDEFY? | B | 6.44 |
| YAWYZWYYWYXWY? | W | 6.67 |
| GHIGIJKLKMNOMOPQR? | Q | 7.40 |
| TUVRSTUPQRST? | N | 7.40 |
| CDDEEEFFFF? | G | 5.67 |
| CDEVW? | X | 5.67 |
| EZDEZCDEZBCDEZ? | A | 6.05 |
| ADGBEHCF? | I | 6.66 |
| WXZCGKNP? | Q | 7.29 |
| XDCBDCYAZYAZZXW? | V | 7.86 |
| QRQRRPSRO? | T | 8.23 |
| HIIJKKLMM? | N | 5.82 |
| AZAAYAAXAAWAA? | V | 5.91 |
| BCDBCDEFGEFGH? | I | 6.17 |
| EFEFCDGHGHCD? | I | 6.56 |
| RTUVUVXYZYZBCDCD? | F | 7.60 |
| BCDEEFGHIHJKLMKNOPQ? | N | 7.57 |
| YZAWXYZUVWXY? | S | 7.39 |
| DUEVDUEVD? | U | 6.03 |
| DOCNBM? | A | 6.34 |
| BBCCDEEFFGHHI? | I | 5.75 |
| HIJGHIJFGHIJ? | E | 6.47 |
| XFHZJLBNP? | D | 7.08 |
| FSRQPGHONMIJK? | L | 7.37 |
| CHDEGFFGHEDC? | I | 7.59 |

*S*s were given a set of 8 practise items, with the experimenter present, to familiarize themselves with the procedure, the experimenter then judging whether the *S* required more practise or could be allowed to continue with the experimental items. The order of presentation of the items was fixed, items being presented in cycles of 7 of increasing difficulty (using the difficulties of the parent items in the Nufferno test as a guide). If a *S* either gave up or keyed in a wrong answer to two consecutive items the nest item presented would be the item at the start of the next cycle of difficulty (i.e. an extremely simple item). This leads to a crude form of tailored testing in which *S*s are only ever presented with items of such difficulty that could reasonably be expected to solve.

*Procedure for Choice Reaction Time task (CRT)*

Decision times (DT) and movement times (MT) were assessed over conditions of 1 and 2 bits of decision information, corresponding to two and four lights on show respectively. Four sets of

10 trials were given, the order of conditions being fixed for all $S$s. The first set of trials was at 1-bit, then 2-bits, then a second set at 2-bits and a second set at 1-bit. Covers were placed over the lights not used in any condition.

The $S$s were seated in front of the response box and used their preferred hand for all button-pressing. The $S$s were given as many practise trials on the 1-bit condition as were required before the $S$ expressed confidence in the task.

Each trial was started by a warning tone of 1000 Hz frequency and 54 msec duration given at approx. 70 dB by the Sirius. The tone was followed by a random delay of 1–4 sec. If the $S$'s finger was on the home button the target light was illuminated. (If the $S$ was not depressing the home button the experimenter received a warning message from the Sirius and could restart the trial after having ensured the $S$ was holding down the home button.) The sequence of positions for the target light were randomized for each $S$. The DT clock was started at the onset of the light and was stopped by the $S$ releasing the home button. This action also started the MT clock which was stopped by the $S$ depressing the target button. The DT, MT and light position were recorded by the Sirius. Further details of the CRT procedure are given in Frearson *et al.* (1988).

*Procedure for Odd-Man-Out task (OMO)*

The Odd-Man-Out paradigm used the same response box as the CRT task. Each stimulus in the OMO task consisted of 3 out of the possible 8 lights being lit. The 3 lights were so arranged that the distance (the number of intervening light positions) between the left light and the centre light was different from the distance between the centre and right lights. Such displays were explained to the $S$ as consisting of a pair of lights (the two closest together) with an 'odd-man-out'. With 8 light positions there are 44 such possible displays. The present study used 12 different displays of which 6 were left–right mirror images of the other 6 (Fig. 2). These 12 were chosen as those which were found to have the best correlation with Raven's matrices score shown by Frearson and Eysenck (1986).

The $S$s were instructed, starting with their finger on the home button, to press the button corresponding to the 'odd-man-out' light. Each display was presented to the $S$ 5 times, making a total of 60 separate trials. The 60 trials were assigned to two groups (three of any display in one group, two in the other). The order of presentation of displays was randomized within each group. After every 15 trials the task was suspended, to allow the $S$ to rest, the task being restarted by the experimenter at the $S$'s instigation.

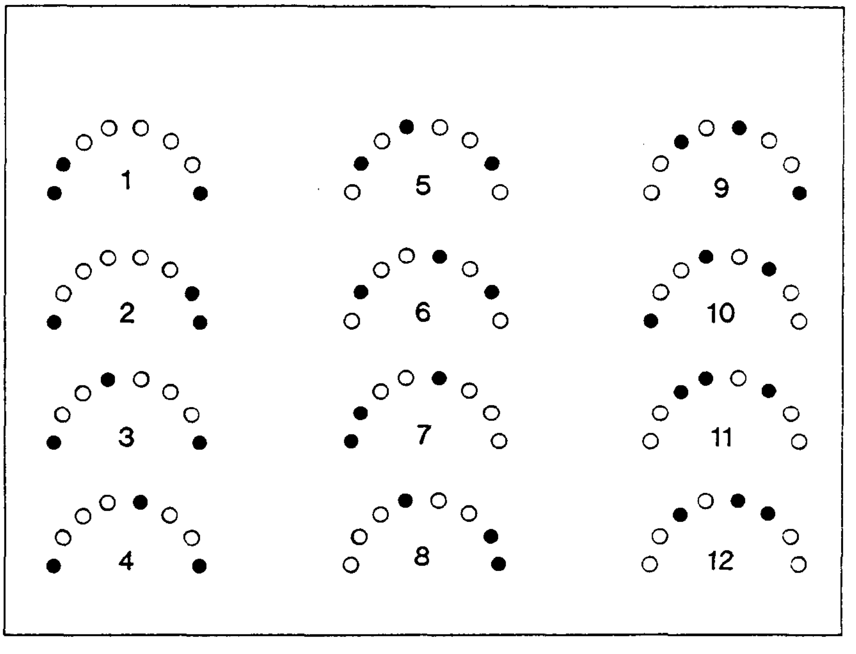


Fig. 2. The 12 Odd-Man-Out displays used.

Trials where the $S$ produced an error (pressed a button other than that corresponding to the odd-man-out) were repeated at the end of each block. (If errors occurred in these repetitions the trial was given again after all the other errors from that block had been repeated.) If on any one block more than 10 errors occurred, the program was halted to allow for more practise. $S$s were instructed in the task and then given a batch of 8–18 practise trials consisting of displays of varying complexity until they expressed confidence in the task. $S$s were told to be as quick as possible in all button-pressing but not to be overly concerned about the possibility of making an error as all errors would be repeated.

The OMO task was implemented in a similar way to the CRT task. Decision times, movement times, type of response (right or wrong) and the chronological sequence of the trial were logged by the Sirius.

The OMO task has been shown to give higher correlations with IQ scores than straightforward CRT (Frearson & Eysenck, 1986) and is hypothesized to involve more of the cognitive factors involved in problem solving. In terms of complexity the OMO lies between the CRT parameters and the simplest of the letter series problems. Further details of the implementation of the OMO task are reported in Frearson, Barrett and Eysenck (1988).

## STATISTICAL ANALYSIS

### *Furneaux analysis of letter series problems*

The Furneaux (1961) model specifies a simplified human problem solver characterized by three parameters: speed, accuracy and continuance (persistence). Furneaux's problem solver is made up of two 'black boxes', a 'problem box' and a 'time-switch'. When a problem is input the time-switch comes on and allows the problem box to start working.

The problem box has two components: a search mechanism and a comparator. The search mechanism sets up 'networks of neural mechanisms' and uses each network to try and solve the problem. The output of each of these 'networks' is tested by the comparator to see if it is an acceptable solution to the problem. If it is not, the search mechanism sets up a new, more complex network. If the solution is acceptable, the problem solver gives it as its answer to the input problem.

Each of the three processes has a parameter associated with it. The 'continuance' of the problem solver is the average length of time the time switch remains on after input before it turns off the problem box (and gives 'give up' as the answer). The speed of the problem solver is related to the rate at which the search mechanism can set up new networks. The 'accuracy' of the problem solver is how 'fussy' the comparator is in accepting or rejecting a possible solution as adequate.

Furneaux specifies that the search mechanism is not operating at random but rather is highly systematic. The first network set up will be comprised of only one neural unit, and the search mechanism will go through all possible one unit mechanisms before it sets up a network where two units are brought into association. Similarly a three unit network will only be tried when all possible two unit networks have been tried and found wanting and so on.

The time for the search mechanism to reach an adequate solution, then, will increase exponentially as the item complexity increases. The function relating solution time, $S$ speed and item difficulty, will be of the form:

$$t = \exp(mD)/\text{rate} \tag{1}$$

where $t$ = solution time; $D$ = item difficulty; $m$ = a constant.
Taking logs

$$\log(t) = mD - \log(\text{rate}) \tag{2}$$

Re-naming $[-\log(\text{rate})]$ as $S$'s speed constant $K$

$$\log(t) = mD + K \tag{3}$$

This relationship, however, will only hold for solution times where the solution is a correct response, indicating that the solution has come from a completed search process. Wrong responses will be the results of truncated searches; abandonments will have been produced by the action of the time switch.

<u>1.Obtain initial estimates of speed and critical difficulty for subjects</u>

```
1.i.    select reference items

1.ii.   select subjects of high continuance.
        split into 3 groups  fast, medium and slow.

1.iii.  calculate difficulty of each reference item.
              D = mean tc      from medium group.

1.iv.   calculate a speed for each subject.
              K = mean (tc - D)    over the reference items.

1.v.    calculate a "can't be less than" estimate of critical difficulty for each subject.
             cD = largest (tc - D) - smallest ta
```

<u>2. Obtain estimates of item difficulties using the subject parameters</u>

```
2.i.    create a matrix of "reduced times"
            each element = tc - K

2.ii.   calculate a difficulty for each item
             D = mean (tc - K)

2.iii.  compare each item's difficulty with each subject's critical difficulty
        discard any element where cD < D
```

<u>3. Obtain a better estimate of subjects critical difficulty</u>

```
3.i.    calculate a difficulty of certain abandonment for all subjects.
             cDa = largest ta - K

3.ii.   calculate mean of the abandonment times greater than cDa for each subject.
             Ta = mean ta (where ta > cDa)

3.iii.  calculate deviations of ta (where ta > cDa) from Ta.

3.iv.   use deviations from all subjects to compute variance of ta (where ta > cDa).

                    Va = mean (Ta-ta)

3.v.    for each subject calculate "effective minimum" abandonment time
            effective minimum ta = Ta - (2*√Va)

3.vi.   for each subject calculate a new critical difficulty.
           cD = largest (tc - D) - effective minimum ta
```

<u>where:</u>

```
tc = log time to correct solution
ta = log time to abandon
 K = subject's speed
 D = item difficulty
Ta = mean of a subject's unambiguous abandonment times
Va = variance of all unambiguous abandonment times
cD = subjects critical difficulty i.e. difficulty at which subject might first abandon
cDa= subject's difficulty of certain abandonment i.e. difficulty at which subject will always abandon
```

Fig. 3. The Furneaux algorithm for simultaneous assessment of item and $S$ parameters.

Assigning a difficulty value based on time to correct solution for each item requires the simultaneous calculation of item difficulties and $S$s' speed constants, as solution time is related to difficulty by only one equation:

$$\log t = mD + K \tag{4}$$

Rather than attempt to solve for all the item difficulties and $S$'s speeds simultaneously, the Furneaux analysis attempts to approximate first one set of parameters and then use these to estimate the other. Figure 3 gives a diagrammatic description of the Furneaux algorithm. A set of 'reference' items is first selected. These are the simplest of the items (i.e. the 8 items that were originally derived from the Nufferno speed scale). Inspection of the percentage of $S$s answering these correctly confirmed that they were the items of the lowest classical difficulty to ensure that there are no (or at least very few) abandonments on any of them. (As abandonments and indeed wrong answers are produced by a different mechanism, the estimation at this stage can only work where there is no chance of an abandonment having taken place.)

Now by looking at their abandonment times $S$s are split into high continuant and low continuant halves (simply on the basis of $S$s' average time to abandonment). The highly continuant $S$s are split into three groups, a fast, medium and slow group, on the basis of their average solution time for the set of reference items.

Now an estimate of the difficulty for the reference items is possible given two assumptions. From equation (4):

*Assumption 1:* $m = 1$. (The value of $m$ is largely a matter of convenience, altering its value will simply multiply all $D$ values by some constant factor.)

*Assumption 2.* The average value of the speed constant for the medium speed group = 0. Again altering this value will simply add a constant to all values of $D$.

So for each item in the reference set:

$$\text{Mean value of} (\log t) = D \tag{5}$$

Now it is possible to calculate for each $S$ in the medium group a speed constant using his response times to the 8 reference items. Each $S$'s speed constant (under the two assumptions above) will be the average deviation of the $S$'s solution time for an item from the problem's difficulty. (The speed constant takes a negative value for faster than average $S$s and a positive value for slow $S$s. In effect it is a 'slowness' rather than a 'fastness' measure. The authors however will follow Furneaux's original usage of the term.)

$$K = \sum (\log t - D)/\text{number of reference items}. \tag{6}$$

Finally a value for the *critical difficulty level* for each $S$ is calculable. This is the difficulty level for any $S$ at which there is first a possibility of his abandoning an item. Clearly for a very large set of items this parameter could be taken to be just the difficulty of the easiest item that was abandoned. For a small set of abandoned items the lowest possible difficulty at which there might be an abandonment must be lower than the lowest actual difficulty of the abandoned items. How much lower is clearly a function of how much variation there is in the $S$s' solution times as it is only when the solution process (of variable duration) is beaten by the 'time switch' (of constant duration) that an abandonment takes place. For an initial estimate it is important only that the value of critical difficulty should not be an over-estimate, as an underestimate will be increased in subsequent iterations whilst an over-estimate will lead to the inclusion of ambiguous data (i.e. data that are being 'filtered' by the action of the time switch) in the estimation of speed and difficulty parameters. To obtain a 'can't be less than' estimate of critical difficulty the largest deviation between an item solution time and its difficulty is subtracted from the smallest of the abandonment times.

It is necessary that the values of these critical difficulties be greater than the difficulties for the reference items in order for the original assumption—that the solution times of this medium speed, high continuance group could be used to estimate item difficulties of the reference items—to be shown to have been justified.

Now by using the same method as above a value of $K$ (speed constant) and $cD$ (critical difficulty level) for each $S$ can be calculated and using these a $D$ (difficulty) for each item. The speed constant ($K$) of each $S$ is subtracted from each correct solution time to produce a matrix of reduced times. Using these reduced times the difficulty value for each item is given by the average value of the reduced times for that item.

Now the new difficulties for the items can be compared with the critical difficulties of the $S$ and any $S$/item pairings where the $S$'s critical difficulty ($cD$) is smaller than the item difficulty are discarded. After a series of iterations, then, a difficulty will have been assigned to each of the items.

The wastefulness of the above process (in terms of the number of $S$s who are being left out of the matrix of reduced times) can now be lessened by obtaining a more accurate estimate (higher) of each $S$'s critical difficulty. First find the critical difficulty for abandonment ($cDa$), i.e. that difficulty above which the item will always be abandoned, by taking the largest abandonment time for each $S$ and subtracting the $S$'s speed constant ($K$). Now all abandoments of items of difficulty higher than the $cDa$ can be taken as unambiguous. The mean for each $S$ of these unambiguous abandonments (expressed as $Ta$) can now be calculated.

Having identified a population of unambiguous abandonments, estimates of the distribution of abandonment times can be made. In particular the variance of abandonments across all $S$s can be calculated.

Furneaux (1961) reports the distribution of abandonments for different $S$s to be uniform enough to justify their combination. However his description of individual statistics is too scant to reasonably evaluate the claim. The combination of individual statistics is necessary to be able to produce any form of distribution statistics given the small number of abandonments. Berger (1976) attempted an individual level analysis and found it impossible to proceed because of the high proportion of $S$s who could not have estimates of their variability of abandonment made.

Given an estimate of the variability of abandoments an estimate of the effective minimum abandonment time can be made, this being taken at the level of two standard deviations below the mean.

$$\text{Effective minimum abandonment time} = Ta - (2^* \sqrt{Va}) \tag{7}$$

This effective minimum is then used to calculate a $cD$ for each $S$ as before using the effective minimum in place of the smallest abandonment time. This will give a higher value for $cD$ and hence allow more data for the calculation of $D$-values in a new set of iterations as above.

*Choice Reaction Time*

The data from the two sets of 10 trials on each condition are combined to give from each day a set of 20 DTs and 20 MTs on each of 1- and 2-bit conditions.

Both DT and MT data were 'corrected' by replacing the largest value within any condition with the mean value of the other 19 DT/MT in that condition, a process first suggested by Barrett, Eysenck and Lucking (1986). Data were also passed through a validity check, such that if a DT was $<140$ msec or longer than 999 msec or an MT was $<90$ msec or $>999$ msec, the DT/MT would be replaced by the mean DT/MT for that condition.

The choice reaction time data were then analysed to give a median DT and MT for each condition and the SD of DT and MT for each condition. In addition a 'slope' measure was obtained by subtracting the median DT obtained for the 1-bit trials from the median DT for the 2-bit trials.

*Odd Man Out*

The median DT and MT of the five correct responses for each display were calculated along with the range which was simply the value of the largest DT/MT minus the smallest. These measures were chosen over the mean and standard deviation because of skewness in the DT/MT data as has been suggested for other DT measures by Brownlee (1975) and Winer (1971), and because of the small number of observations. Also the number of errors made during the task was calculated.

Previous analysis of the data from this study had shown that for each of these four parameters the values for the 12 displays used gave a strong general factor when submitted to factor analysis.

Given the adoption of a general factor solution the arithmetic mean of the 12 individual scores could be taken as a reasonable measure of performance on all 12 different displays. The 'factor scores' were computed by applying unit weights to all variables. The 4 factor scores are then *ODT.FS* for the 12 median decision times, *OMT.FS* for the 12 movement times, *ORRNG.FS* for the 12 ranges of decision times and *OMRNG.FS* for the ranges of the movement times.

## RESULTS

*The simultaneous assessment of item difficulties and* S *speeds and critical difficulties by the Furneaux iteration*

The Furneaux iteration to produce estimates of item difficulties and $S$ speed and persistence parameters was performed on three sets of data. These were the data from all $S$s on day 1, which for half the $S$s consists of responses to the first set of letter-series items, and to the second set of letter-series items for the rest of the $S$s; data from all $S$s on day 2, and data from all $S$s on both days i.e. responses on all items from all $S$s. These three analyses would allow estimates for reliability of both $S$ and item parameters.

The Furneaux procedure as described above consists of a pair of nested iteration loops. The inner loop (moving individual solution times in and out of a matrix of 'reduced' times) was performed three times on each successive run, this being an arbitrary value set by inspection of the reduced matrix on successive runs through this loop both on these data and on a set of simulated data. The form of the matrix appeared to remain stable after the first or second time through the loop.

The outer loop (re-calculating values for critical difficulties and then using these for forming a new matrix of reduced times) was performed 20 times on each run. The progress of the iteration was monitored by (at the end of each outer loop iteration) correlating the 56 item difficulties obtained with the difficulties produced by the last iteration. In all cases this correlation rose to excess of 0.99 within 10 iterations and remained at this level on each successive iteration. The stability of the derived difficulty values is therefore extremely high.

The set of parameters of primary interest at this stage is the item difficulties as it is to be expected that several $S$s will not provide a full set of parameters from this procedure. However with a stable set of item difficulties it will be possible to calculate for every $S$ a variety of speed and persistence parameters.

*Item difficulties*

Each item has been assigned three different estimates of its difficulty, one based on data from the 54/55 $S$s who 'attempted' it on day 1 (though of course by the nature of the tailored testing procedure the number who actually answered an item is variable), one based on the 54/55 $S$s who attempted it on day 2, and a third based on data from all the $S$s, irrespective of when they answered the item.

Comparison of the first two difficulties provides a rigorous test of the generality of the Furneaux derived difficulties; as they represent the case of two unrelated samples doing the same items at different times, any correspondence between difficulties can only be attributed to the item's contents.

The difficulties of the first set of items from day 1 $S$s are incomplete in that item 7 was not answered correctly by any of the sample and so cannot be assigned a difficulty, likewise the day 2 $S$s failed to provide a correct answer for item 28 which again cannot be assigned a value. The correlation between the two sets of data is 0.914 when items 7 and 28 are removed. The difficulties for set two of the letter-series show an even greater correspondence between day 1 and day 2 $S$s (difficulties are correlated 0.956).

The conclusion of these comparisons is that the Furneaux time based difficulties have a reality based solely upon item content. Though the actual values that will be derived from a particular group are arbitrary, a difficult item for one group remains a difficult item for another group.

The choice of actual value to assign to an item for use for further analysis being arbtitrary, it would seem sensible to use the values obtained by the Furneaux iteration that was performed on the whole data set. (These difficulties correlates with day 1 difficulties on set 1 at 0.961 and set 2 at 0.987, and with the day 2 difficulties at 0.991 and 0.968 for the two sets of items.) These difficulties for each item are given in Table 1 above.

*S speed and persistence parameters*

The Furneaux theory assigns to each $S$ a speed constant and a critical difficulty, this being the difficulty at which the $S$ might first abandom an item. The Furneaux iteration requires these constants to be assigned to at least some of the $S$s in order to produce item difficulties. At the end of a perfect analysis then one would expect to have a speed and difficulty constant for each $S$. However, in practise such an ideal is very unlikely to be achieved. $S$s might not produce any abandonments throughout the task (and hence cannot be assigned a critical difficulty), or alternatively might have so low a critical difficulty that they never produce the unambiguous (i.e. responses so quick that there is no possibility of the $S$ abandoning such an item) correct responses needed to assign them a speed. As well as these 'legitimate' failures of the iterative method, the requirements for obtaining an accurate set of item difficulties (i.e. that all correct responses used in estimating item difficulties must be unambiguous) lead to an over rigorous exclusion of *S*s and hence to their not being assigned one or both of the parameters.

In all only 59 of the 109 *S*s were assigned a full set of parameters (i.e. both speed and critical difficulty on day 1, day 2 and when all data is analysed as one). The data for these 59 *S*s however allow the calculation of day to day reliabilities for these two parameters (i.e. the Pearson correlation of the parameter value on day 1 with the day 2 value). The reliability of the speed parameter is 0.542 and for the critical difficulty 0.746. These values are quite low in the context of the usual psychometric measures though it should be pointed out that they are based on an average of $<13$ correct answers on each day.

Comparisons between these Furneaux parameters and other variables will be made using the speeds and critical difficulties obtained when the whole data set is analysed as one. These parameters being based on the largest number of observations are assumed to be the best estimators of the underlying individual traits. The correlation between *S* speed and critical difficulty parameters and WAIS-R subtest scores are given in Table 2.

The Furneaux model predicts that the two parameters of speed and continuance will be largely independent; the correlation between the speed parameter and critical difficulty is 0.498.

*S parameters derived directly from item difficulties*

The difficulties of using *S* parameters derived from the Furneaux iteration necessitates assigning values that can be calculated using the item difficulties as fixed parameters and the original log (solution times) for each *S*. For each *S* we can then calculate a Furneaux speed where

$$\text{Speed} = \text{mean of}\,(\log(\text{solution time}) - \text{item difficulty}) \tag{8}$$

Such a speed can be calculated using *S*s' response times on day 1, day 2 and both together (though note that the item difficulties used are always those derived by the Furneaux iteration using all the data). This allows the finding of a reliability for *S*s' Furneaux speed parameters.

To compare the performance of the Furneaux parameter with other estimates of speed performance two further sets of parameters are computed for each *S*. First is a simple average of all the log (solution time)s, second is an average of only the log (solution time)s on the items that the tailored procedure forced all *S*s to attempt (8 items if the statistic is calculated from 1 day's performance, 16 if 2 days'). Again both these statistics are computed for just day 1, just day 2 and combining all the data.

These three statistics can be computed for all *S*s who answered any question correctly. In fact on day 2 *S* number 7 failed to get any questions correct and so was removed from this part of the analysis. Also in order to have the same group in all subsequent analysis, *S*s who failed to abandon any item and hence would not be able to have any continuance parameters assigned to them were also dropped from the analysis, these being the 'legitimate' failures, i.e. people whose true speeds and continuances lay outside the measuring range of this set of items. This left a sample of 103 *S*s.

Table 2. Correlations between WAIS-R subtest scores and speed and critical difficulty from the Furneaux iteration

| WAIS-R scale | Speed (derived from all data) | Critical difficulty (derived from all data) |
|---|---|---|
| Information | −0.079 | 0.206 |
| Digit span | −0.221* | 0.194 |
| Vocabulary | −0.177 | 0.307** |
| Arithmetic | −0.244* | −0.031 |
| Comprehension | −0.168 | 0.381** |
| Similarities | 0.045 | 0.350** |
| Verbal IQ | −0.234* | 0.307** |
| Picture completion | −0.408** | −0.085 |
| Picture arrangement | −0.092 | 0.168 |
| Block design | −0.204 | 0.337** |
| Object assembly | −0.228* | 0.101 |
| Digit symbol | −0.173 | 0.217 |
| Performance IQ | −0.334** | 0.137 |
| Full scale IQ | −0.309** | 0.264* |

Significant correlations marked with an * ($P < 0.05$ 1-tailed) or ** ($P < 0.01$ 1-tailed). $N = 59$.

Table 3. Correlations between WAIS-R subtest scores and three alternative speed parameters

| WAIS-R scale | Furneaux speed | Mean log(sol'n time) | Mean log(sol'n time) common items |
|---|---|---|---|
| Information | −0.162 | 0.081 | −0.126 |
| Digit span | −0.201* | 0.032 | −0.136 |
| Vocabulary | −0.244* | 0.012 | −0.209* |
| Arithmetic | −0.277** | 0.041 | −0.192 |
| Comprehension | −0.194 | 0.100 | −0.128 |
| Similarities | −0.112 | 0.172 | −0.071 |
| Verbal IQ | −0.274** | 0.069 | −0.212* |
| Picture completion | −0.369** | −0.111 | −0.337** |
| Picture arrangement | −0.063 | 0.190 | 0.006 |
| Block design | −0.155 | 0.190 | −0.152 |
| Object assembly | −0.181 | 0.078 | −0.177 |
| Digit symbol | −0.267** | 0.022 | −0.215* |
| Performance IQ | −0.306** | 0.044 | −0.274** |
| Full scale IQ | −0.310** | 0.068 | −0.257* |

Significant correlations marked with an * ($P < 0.05$ 2-tailed) or ** ($P < 0.01$ 2-tailed). $N = 103$.

The day to day reliabilities of the 3 speed parameters are as follows:

| | |
|---|---|
| Furneaux speed | 0.702 |
| Average of all correct log(sol'n time) | 0.718 |
| Average of common log(sol'n time) | 0.691 |

Correlations between these 3 alternate speed parameters (when calculated using data from both days) and WAIS-R subtest scores are given in Table 3. Similarly an alternative set of continuance parameters can be computed using the *S* speeds derived above. For each *S* a value of Furneaux continuance can be computed where

$$\text{Continuance} = \text{mean}\ (\log(\text{time to abandon}) - S\text{s speed}) \tag{9}$$

(For the calculation it is the Furneaux speed which is taken as the *S*s' speed.) To give a parameter to compare against the performance of this continuance parameter, a simple mean log (time to abandon) is also calculated for each *S*.

In addition to these parameters derived from the times to abandonment for *S*s, parameters based solely upon the difficulties of abandoned items can be computed. These are similar to the critical difficulty for *S*s though without its associated computational problems. Such parameters must always be based upon the difficulties of items of the lowest difficulty that any one *S* abandons as the tailored testing procedure ensures that items of a higher difficulty are not given to the *S*. Two such parameters are calculated for each *S*:

(1) Mean of difficulties of the abandoned item of lowest difficulty from each cycle of items, (i.e. a mean of up to 4 difficulties, one from each cycle, when the parameter is computed for 1 day's data or 8 difficulties when computed for both days' data).

(2) The difficulty of the simplest item abandoned.

All these continuance parameters can be calculated using only the data from 1 day, so allowing an estimate of the day to day reliability of the two measures.

| | |
|---|---|
| Reliability of Furneaux Continuance | 0.900 |
| Reliability of mean log(abandon time) | 0.848 |
| Reliability of mean lowest abandoned item difficulty from each cycle | 0.693 |
| Reliability of lowest abandoned item difficulty | 0.608 |

The correlation of each of the measures of persistence (when calculated using the data from both days) and WAIS-R subtest scores are given in Table 4. How closely each of these derived parameters approximates to the true Furneaux parameters (i.e. those produced directly by the iterative procedure which produces the item difficulties) can be seen from each parameters' correlation with its 'parent' parameter computed for those 59 *S*s who gave a full set of values from the iteration. These correlations are given in Table 5.

Table 4. Correlations between WAIS-R subtest scores and four alternative measures of persistence

| WAIS-R scale | Continuance | Mean (log(abandoned time)) | Mean (lowest abandoned item difficulty from 8 cycles) | Lowest abandoned item difficulty |
|---|---|---|---|---|
| Information | 0.145 | 0.315** | 0.382** | 0.305** |
| Digit span | 0.086 | 0.241* | 0.282** | 0.203* |
| Vocabulary | 0.151 | 0.361** | 0.430** | 0.344** |
| Arithmetic | 0.113 | 0.290** | 0.395** | 0.351** |
| Comprehension | 0.205* | 0.398** | 0.513** | 0.446** |
| Similarities | 0.220* | 0.387** | 0.516** | 0.375** |
| Verbal IQ | 0.186 | 0.426** | 0.538** | 0.454** |
| Picture completion | −0.024 | 0.193 | 0.413** | 0.393** |
| Picture arrangement | 0.222* | 0.336** | 0.458** | 0.408** |
| Block design | 0.327** | 0.521** | 0.668** | 0.556** |
| Object assembly | 0.189 | 0.338** | 0.530** | 0.513** |
| Digit symbol | 0.112 | 0.326** | 0.519** | 0.430** |
| Performance IQ | 0.176 | 0.421** | 0.655** | 0.607** |
| Full scale IQ | 0.198* | 0.462** | 0.632** | 0.565** |

Significant correlations marked with an * ($P < 0.05$ 2-tailed) or ** ($P < 0.01$ 2-tailed). $N = 103$.

*Estimation of S accuracy parameters*

The third $S$ parameter of the Furneaux model is accuracy. The formulation of the accuracy parameter in Furneaux's (1961) study is unclear. White (1982) concludes that it is simply the proportion of right to wrong answers. Certainly the accuracy estimates for $S$s never enter into the calculation of any of the other parameters either for $S$s or items. This is in line with the Furneaux conception of $S$ accuracy being associated with the 'comparator' rather than the 'problem solving' process of the model. Three separate parameters representing accuracy were calculated. Each one was calculated over day 1, day 2 and when both sets of data were added together and treated as one. The three parameters were:

(i) number right/(number right + number wrong)
(ii) number right/(number right + number wrong + number abandoned)
(iii) number right.

Of these three parameters, the closest to the conception of Furneaux's accuracy is the simple proportion of number right to number wrong. Computing each variable over two separate days allows for a measure of a day to day reliability for each of these three parameters. These are:

| | |
|---|---|
| proportion right: wrong | 0.678 |
| proportion right: wrong or abandoned | 0.588 |
| number right | 0.868 |

The values of each of these three parameters calculated over both days were then correlated with the individual WAIS-R sub-test scores. These correlations are given in Table 6.

Table 5. Correlations between 'true' Furneaux parameters and parameters derived from item difficulties and raw data

| | Correlation with speed from Furneaux iteration |
|---|---|
| Mean (log(sol'n time) − difficulty) | 0.998 |
| Mean (log(sol'n time)) | 0.895 |
| Mean (log(sol'n time)) over common items | 0.947 |
| | **Correlation with critical difficulty from Furneaux iteration** |
| Mean (log(abandon time) − speed) | 0.898 |
| Mean (log(abandon time)) | 0.947 |
| Mean lowest 8 abandoned item difficulties | 0.564 |
| Lowest abandoned item difficulty | 0.367 |

Table 6. Correlations between WAIS-R subtest scores and three alternative accuracy parameters

| WAIS-R scale | Proportion right: wrong | Proportion right: wrong or abandoned | Number right |
|---|---|---|---|
| Information | −0.392** | −0.310** | 0.448** |
| Digit span | −0.266** | −0.210* | 0.364** |
| Vocabulary | −0.448** | −0.373** | 0.501** |
| Arithmetic | −0.493** | −0.429** | 0.514** |
| Comprehension | −0.443** | −0.359** | 0.576** |
| Similarities | −0.448** | −0.359** | 0.565** |
| Verbal IQ | −0.518** | −0.422** | 0.632** |
| Picture completion | −0.290** | −0.202* | 0.432** |
| Picture arrangement | −0.447** | −0.358** | 0.552** |
| Block design | −0.498** | −0.351** | 0.679** |
| Object assembly | −0.414** | −0.314** | 0.528** |
| Digit symbol | −0.439** | −0.326** | 0.569** |
| Performance IQ | −0.503** | −0.380** | 0.674** |
| Full scale IQ | −0.554** | −0.440** | 0.707** |

Significant correlations marked with an * ($P < 0.05$ 1-tailed) or ** ($P < 0.01$ 1-tailed). $N = 103$.

## *The selection of S parameters*

The Furneaux model proposes that *S*s should be assigned three largely independent parameters representing speed, persistence and accuracy. Furneaux (1961) gives a method of explicitly calculating speed and persistence for each *S* as a by-product of calculating difficulties for items. The requirements for obtaining good estimates of item difficulties, however, act to ensure that only a sub-set of *S*s can be assigned both speed and persistence values. Alternative speed and persistence parameters have then to be computed if all *S*s are to be given a full set of attributes. These alternative parameters can be judged against three sets of criteria. Firstly there are internal considerations, largely day to day reliability, but also how many *S*s cannot be assigned a value. Secondly, there is the correlation between the derived measure and its 'parent' Furneaux parameter. Finally, there is the pattern of correlations between the parameter and the sub-tests of the WAIS-R and the contribution the variable makes in multiple regressions to predict WAIS-R scores.

On these criteria amongst the speed parameters the mean (log(sol'n time) − difficulty) appears as the most suitable. It has a reliability slightly less than the simple mean (log(sol'n time)) (0.702 compared to 0.718) but it is more closely correlated with the speed derived from the Furneaux iteration (0.998 compared to 0.895) and shows the expected negative correlations with WAIS-R scores.

The near zero correlation between the simple mean (log(sol'n time)) and WAIS-R scores is as might be expected, given that there are two underlying processes operating. High IQ *S*s are solving easy items quickly and hence decreasing their mean (log(sol'n time)), but are succeeding at the harder items which take more time and are hence increasing their mean (log(sol'n time)). The two together ensure no correlation between a simple mean (log(sol'n time)) and IQ measures.

Amongst the derived persistence parameters the mean (log(abandon time)) is preferable to the mean (log(abandon time) − *S*s speed) as although it is less reliable (0.848 compared to 0.900) it is more closely correlated with the Furneaux iteration derived parameter (0.947 compared to 0.898) and is more strongly correlated with WAIS-R scores. The two persistence parameters based on item difficulties only (rather than log times to abandonment) are discounted because of their low reliability. (Although the mean of the lowest difficulty from each cycle shows high correlations with WAIS-R scores, its high correlation with accuracy parameters means it performs less well in multiple regressions to predict WAIS-R scores than the more independent mean log (abandon time) which is chosen in preference to it as the representative persistence parameter.)

The Furneaux iteration does not provide accuracy scores. However, the sense of the Furneaux model entails not taking 'number right' as an accuracy score as 'number right' will be a function of speed, persistence and accuracy. The proportion of number correct to (number correct + number wrong) is both more reliable and more closely correlated with WAIS-R scores than the proportion of number right to (number right + number wrong + number abandoned) and is more clearly appropriate to the model.

*Multiple correlations to WAIS-R scores using speed, persistence and accuracy as predictor variables*

The Furneaux model assigns to each individual a set of three 'atomic' scores: speed, accuracy and persistence. Any cognitive task can be said to depend upon each of these three attributes to a different degree, and also upon a variety of non-intellective factors. If the three Furneaux scores of speed, accuracy and persistence truly are the atomic properties of the problem solver, then they should when combined, using suitable weights, be capable of predicting performance on any cognitive task to an equal level dependent only upon the extent to which performance on the task is unaffected by task specific factors i.e. a task's $g$ loading.

Multiple regressions using the sub-scores from the WAIS-R as dependant variables and the speed, accuracy and persistence parameters as predictors will show both the extent to which performance on a given WAIS-R task depends on each of the supposed 'atomic' properties and how well performance on the task is predicted by the atomic properties. This will give two alternative measures of the adequacy of the model, one qualitative (the apparent validity of the relative weight of each atomic property on the task's performance), and one quantitative (the size of the correlation between a task's '$g$-loading' and the $R^2$ from the multiple regression). In order to be able to legitimately compare the $\beta$-coefficients both between speed, accuracy and persistence and the WAIS-R subtests, the data for all these variables are put into the same measurement scale, i.e. normalized to give a mean of zero and SD of 1.0.

The results of these multiple regressions are given in Table 7. The multiple-$R$s given in Table 7 are not shrunken for bias. With 103 $S$s and only 3 predictor variables capitalization on chance will not inflate the multiple $R$ value. Corrections if applied would only affect the third decimal place.) The $g$-loading of each of 11 WAIS-R subtests was calculated by performing a principal components analysis on the matrix of correlations given in the WAIS-R Manual. The correlation between the amount of variance predicted by the three Furneaux measures (i.e. the $R^2$ in Table 7) and the $g$-loading was 0.05.

The loadings of speed, accuracy and persistence on the WAIS-R subtests are largely as would be predicted from a knowledge of the test contents. Digit symbol and picture completion are strongly dependent on speed whilst picture arrangement is far less so; comprehension, vocabulary and block design depend heavily on persistence; arithmetic is heavily dependent on accuracy, whilst digit span and picture completion are much less so.

Comparison of the multiple $R$ obtained with the WAIS-R scores and the three separate measures derived from the letter series problems and the simple correlation between the simple score and WAIS-R scores (Table 6) show that the separate analysis of performance into three (largely independent) parameters allows for a generally better prediction of the various different subtests of the WAIS-R than is possible from the single heterogenous 'number right' score.

*Correlation between speed, accuracy and persistence scores and Ravens matrices scores*

The correlations are −0.328 for speed, 0.409 for persistence, and −0.470 for accuracy.

Table 7. Results of multiple regressions using three parameters representative of speed, persistence and accuracy (which have been normalized) to estimate each of the WAIS-R subtest scores and verbal, Performance and Full scale IQ (also normalized)

| WAIS-R scale | $R^2$ | Furneaux speed | Average abandon time | Proportion right:wrong |
|---|---|---|---|---|
| Information | 0.509 | −0.310 | 0.341 | −0.260 |
| Digit span | 0.430 | −0.343 | 0.330 | −0.139 |
| Vocabulary | 0.622 | −0.430 | 0.428 | −0.283 |
| Arithmetic | 0.593 | −0.354 | 0.285 | −0.384 |
| Comprehension | 0.612 | −0.391 | 0.456 | −0.265 |
| Similarities | 0.560 | −0.280 | 0.385 | −0.297 |
| Picture completion | 0.552 | −0.519 | 0.348 | −0.159 |
| Picture arrangement | 0.506 | −0.186 | 0.278 | −0.338 |
| Block design | 0.708 | −0.406 | 0.583 | −0.270 |
| Object assembly | 0.548 | −0.344 | 0.374 | −0.269 |
| Digit symbol | 0.611 | −0.439 | 0.396 | −0.287 |
| Verbal IQ | 0.719 | −0.492 | 0.503 | −0.324 |
| Performance IQ | 0.731 | −0.532 | 0.524 | −0.301 |
| Full scale IQ | 0.785 | −0.551 | 0.557 | −0.339 |

$N$ = 103.

Table 8. Correlations between average reaction time parameters from day 1 and day 2 on the choice reaction time tasks and three variables from the Furneaux letter series task representing speed, persistence and accuracy

| Variable | Speed based on Furneaux item difficulties (Speed) | Average abandonment time (Persistence) | Proportion of right:wrong (Accuracy) |
|---|---|---|---|
| DT1 | 0.086 | −0.265** | 0.222* |
| DT2 | 0.085 | −0.195* | 0.183 |
| SLOPE | 0.069 | −0.143 | 0.130 |
| DV1 | 0.156 | −0.288** | 0.280** |
| DV2 | 0.130 | −0.162 | 0.035 |
| MT1 | 0.022 | −0.304** | 0.271** |
| MT2 | −0.046 | −0.316** | 0.261** |
| MV1 | 0.138 | −0.228* | 0.149 |
| MV2 | 0.039 | −0.300** | 0.139 |
| ORT.FS | 0.177 | −0.246* | 0.237* |
| OMT.FS | −0.091 | −0.290** | 0.184 |
| ORRNG.FS | 0.232* | −0.241* | 0.320** |
| OMRNG.FS | 0.088 | −0.153 | 0.186 |
| FULL SCALE IQ | −0.310** | 0.462** | −0.555** |

Significant correlations marked with * ($P < 0.05$ 2-tailed) or ** ($P < 0.01$ 2-tailed). $N = 103$.

*Correlation between speed, accuracy and persistence scores and decision time parameters*

The correlations between the three representative speed, accuracy and persistence parameters and parameters derived from the CRT and OMO tasks are given in Table 8. The most notable feature of these results is the absence of significant correlation between speed and any of the decision time measures. The pattern of correlations is indicative of the correlation betweeen Furneaux parameters and decision time parameters being mediated by their common relationship to IQ rather than any causal link between fast/consistent decision and movement times and fast solving of problems. Certainly there is no support for the suggestion that the correlation between DT and IQ is a function only of the common 'speeded' elements in DT tasks and IQ tests. The failure of DT parameters to correlate with speed is related to the 'test-speed paradox' discussed by Jensen (1982). Jensen suggests that items of a complexity much greater than simple reaction time tasks invoke other processes than the simple 'limited capacity information processing channel' such as short term memory, encoding stimuli etc. The involvement of such processes would mitigate against a large correlation between DT measures and Furneaux style speed measures.

*Correlation between speed, accuracy and persistence scores and personality parameters*

Furneaux (1961) postulates that a relationship should exist between the 'atomic' properties of the individual problem solver (i.e. the *S*'s Furneaux-Serived speed, accuracy and continuance scores) and the individual's personality. Particularly he suggests that continuance and speed will be a function of 'drive' and describes some experiments, where the manipulation of arousal affects *S*s' speed and accuracy differentially depending on their scores on the Guilford S.T.D.C.R. inventory.

Correlations between the three representative speed, accuracy, and persistence parameters and the seven personality parameters from the EPQ and IVE are given in Table 9. From the EPQ only

Table 9. Correlations between three 'Furneaux' parameters representing speed, persistence and accuracy and personality scores from the EPQ and IVE

| Variable | Speed based on Furneaux item difficulties (Speed) | Average abandonment time (Persistence) | Proportion right:wrong (Accuracy) |
|---|---|---|---|
| Psychoticism | 0.035 | −0.009 | −0.184 |
| Extraversion | 0.024 | 0.060 | 0.082 |
| Neuroticism | −0.022 | −0.002 | −0.075 |
| Social desirability | −0.144 | −0.346** | 0.203* |
| Impulsivity | −0.148 | 0.013 | 0.097 |
| Venturesomeness | 0.213* | 0.346** | −0.076 |
| Empathy | −0.102 | 0.164 | −0.138 |

Significant values marked with an * ($P < 0.05$ 2-tailed) or ** ($P < 0.01$ 2-tailed).

social desirability is significantly correlated to any of the 'Furneaux' parameters. This is probably attributable to the marked correlation between social desirability and WAIS-R IQ ($-0.335$ in this sample). This result is surprising in the light of Farley (1966) reporting correlations between both $E$ and $N$ and speed as measured using the Nufferno tests. However, the Nufferno tests were not subjected to the type of analysis that Furneaux (1961) suggested for tests to truly measure the atomic property of speed. The relationship between Nufferno speed and speed as calculated by the Furneaux methods used here is unclear.

From the IVE, venturesomeness is significantly correlated with both speed and persistence. This result mirrors results comparing decision time performance with personality where again it is venturesomeness which shows the greatest relationship with performance measures (Frearson *et al.*, 1988).

## DISCUSSION

The Furneaux model approaches individual differences in problem solving in a new way. Both the properties of items and *S*s are re-interpreted. Items are considered in terms of their difficulty, based not on the probabilistic properties of pass/fail (as in classic psychometrics and item response theory) but rather in terms of difficulty based on time to solve. The relationship between this 'time based' difficulty and the difficulties of classical psychometrics and item response theory remains unexplored, but the meaningfulness of time-based difficulties (i.e. that relative item difficulty is dependent solely upon the item's content not upon properties of the sample's individuals) has been demonstrated.

The classification of *S*s by their speed, accuracy and continuance has also been shown to be reliable and apparently valid. However, the strict adherence to the Furneaux method for the computation of *S* speed and continuance parameters which leads to the failure to assign a full set of parameters to many *S*s (and hence has led to the premature halting of attempts at replication) has been shown to be unnecessary. Parameters for continuance can be computed directly from the times taken to abandon items and such parameters out-perform parameters calculated with a closer adherence to the letter of the Furneaux method. Similarly, *S* accuracy parameters can be computed with no regard to the Furneaux model (indeed it is hard to see how the Furneaux method does calculate accuracy parameters) and appear to perform well.

It is in the calculation of speed parameters that the Furneaux model shows distinct advantages over more crude measures. To investigate individual differences in speed it is clearly imperative to account for differences in item characteristics. A simple summation of the time to complete a set of items can never show the relationship between speed and general ability as it confounds two opposite tendencies: the tendency for bright *S*s to rapidly complete items and to succeed at difficult and time-consuming problems. Attempts to measure speed have therefore concentrated on giving *S*s items only of such low difficulty that all *S*s will succeed on them, measuring only the first of the tendencies. The resulting speed measures applying as they do only to items of a trivial difficulty are of limited validity (and the difficulties in measuring solution times are magnified by virtue of their shortness). The Furneaux method addresses the problem of measuring speed by determining for items a time-based difficulty measure. The use of these difficulties means that *S*s' speeds can be calculated from any set of items of a full range of difficulty. It is this facility that allows the Furneaux method to produce speed measures for subjects that are reliable and apparently valid.

The nature of this speed characteristic of individuals is indicated by the results from comparing Furneaux speed and the WAIS-R scores and decision time data. Significant correlations between Furneaux speed and the unspeeded subtests of the WAIS-R, particularly vocabulary, show that Furneaux speed is not simply a facility for 'working against the clock' but is rather a trait for rapid manipulation of data. The low correlation between speed and decision time parameters shows Furneaux speed not to be a simple measure of ability for rapid response but rather an aspect of more cognitive traits.

Theoretically the poor correlation between the Furneaux speed measure and decision time parameters is counter indicative of a simple hierarchical model where fast decision times lead to fast cognitive speed and hence a high IQ, but rather points to a model of a high IQ 'dispositional set' where responsivity to outside stimuli interacts with internal properties of the nervous system leading to the parallel development of fast decision times, cognitive speed and IQ.

The practical applications of the Furneaux method are important. It has long been felt that the time taken for a *S* to produce a response was as informative as whether it was right or wrong. However, the difficulty in accurately timing each response has made use of such data prohibitive. With the introduction of automatic ability testing by micro computer, the collection of response times becomes trivial. The Furneaux method allows use to be made of these extra data. The results of the multiple regressions using Furneaux speed along with accuracy and persistence parameters to predict WAIS-R subtest scores as well as verbal, performance and full scale IQ shows that the use of response times allows a short test to perform more like a battery of tests than a single one. (Typically any one test will correlate highly with some tests with which it shares common item specific factors and very badly with others. The generally high multiple *R* obtained with the three parameters of the Furneaux test for all the WAIS-R subtests show its greater independence from the effects of item specific factors.)

The supporting evidence for the details of the Furneaux model is weak, indeed it is hard to see what form such evidence could take. However, the function of models of such wide scope is more to be of use in stimulating research than to be literally correct. The Furneaux model highlights an alternative approach to individual difference research distinct from both psychometrics and item response theory. Such an approach appears to give results of acceptable reliability and validity and with changes in techniques of practical testing the over-heads involved in collecting the extra data required become less of an obstacle.

## REFERENCES


Barrett, P., Eysenck, H. J. & Lucking, S. (1986). Reaction time and intelligence: A replicated study. *Intelligence*, **10**, 9–40.

Berger, M. (1976). A study of the Eysenck–Furneaux approach to the analysis of performance on intelligence tests. Ph.D. thesis. University of London.

Berger, M. (1982). The scientific approach to intelligence: An overview of its history with special reference to mental speed. In Eysenck, H. J. (Ed.), *A model for intelligence*. New York: Springer.

Brierly, H. (1969). The intellectual speed and accuracy of patients diagnosed as neurotic. Ph.D. thesis. University of Leeds.

Brownlee, K. T. (1975). *Statistical theory and methodology in science and engineering*. New York: Wiley.

Eysenck, H. J. (1953). *Uses and abuses of psychology*. London: Pelican.

Eysenck, H. J. (1982). *A model for intelligence*. New York: Springer.

Eysenck, S. B. G., Pearson, P. R., Easting, G. & Allsopp, J. (1985). Age norms for impulsiveness, venturesomeness and empathy in adults. *Personality and Individual Differences*, **6**, 613–620.

Farley, F. H. (1966). Individual differences in solution time in error free problem solving. *British Journal of Social and Clinical Psychology*, **5**, 306–309.

Frearson, W. M. & Eysenck, H. J. (1986). Intelligence, reaction time and a new 'Odd-Man-Out' RT paradigm. *Personality and Individual Differences*, **7**, 807–817.

Frearson, W. M., Barrett, P. T. & Eysenck, H. J. (1988). Intelligence reaction time and the effects of smoking. *Personality and Individual Differences*, **9**, 497–517.

Furneaux, W. D. (1955). *The Nufferno tests of speed, accuracy and level*. London: University of London Press.

Furneaux, W. D. (1961). Intellectual abilities and problem solving behaviour. In Eysenck, H. J. (Ed.), *Handbook of abnormal psychology*. New York: Basic Books.

Jensen, A. R. (1982). Reaction time and psychometric '*g*'. In Eysenck, H. J. (Ed.), *A model for intelligence* (pp. 93–132). New York: Springer.

Jensen, A. R. & Munro, E. (1979). Reaction time movement time and intelligence. *Intelligence*, **3**, 121–126.

Lord, F. M. (1956). A study of speed factors in tests and academic grades. *Psychometrika*, **21**, 31–50.

Mangan, G. L. (1958). A factorial study of speed, power and related temperament variables. *British Journal of Educational Psychology*, **29**, 144–154.

May, M. A. (1921). Psychological examination in the U.S. army. *National Academy of Science Memoirs*, **5**, 415.

Slater, P. (1937). Speed of work in intelligence tests. *British Journal of Psychology*, **29**, 55–68.

Sutherland, J. D. (1934). The speed factor in intelligent reactions. *British Journal of Psychology*, **24**, 276–294.

Tate, M. W. (1948). Individual differences in speed of response in mental tests of varying degrees of difficulty. *Educational and Psychological Measurement*, **8**, 353–374.

Thorndike, E. L., Bregman, E. O., Cobb, M. V. & Woodyard, E. (1927). *The measurement of intelligence*. New York: Teachers College, Columbia University.

White, P. O. (1973). Individual differences in speed, accuracy and persistence: A mathematical model for problem solving. In Eysenck, H. J. (Ed.), *The measurement of human intelligence*. Lancaster: TP Press.

White, P. O. (1982). Some major components in general intelligence. In Eysenck, H. J. (Ed.), *A model for intelligence*. New York: Springer.

Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.