Pergamon

PII: S0191-8869(97)00184-0

Person. individ. Diff. Vol. 24, No. 3, pp. 405–419, 1998

Printed in Great Britain
0191-8869/98 $19.00+0.00

ESTIMATING INSPECTION TIME: RESPONSE PROBABILITIES, THE BRAT IT ALGORITHM, AND IQ CORRELATIONS

P. T. Barrett,[1] K. V. Petrides[2] and H. J. Eysenck[3]✠

[1] Ashworth Hospital, Psychometrics Unit, Parkbourn, Maghull, Liverpool L31 1HW and University of Liverpool, Department of Clinical Psychology, The Whelan Building, Brownlow Hill, Liverpool L69 3GB, [2] University of Nottingham, Department of Psychology, University Park, Nottingham NG7 2RD and [3] Institute of Psychiatry, Department of Psychology, De Crespigny Park, Denmark Hill, London SE5 8AF, England

(Received 8 December 1996; received for publication 3 October 1997)

Summary—Bates and Eysenck (1993), used a 3rd-order cubic polynomial curve fitting procedure on correct-response probabilities computed from the trial record of individual research participants ($N = 70$) in an inspection time (IT) task. They demonstrated that this methodology produced estimates of IT that, when correlated with full-scale IQ scores (assessed by Jackson's Multidimensional Aptitude Battery), provided a measure of agreement that exceeded that given by the Barrett BRAT IT algorithm. The correlation between IT computed via the BRAT algorithm and full-scale IQ in this sample was −0.35, that between IQ and the cubic polynomial estimate was −0.35. When removing one outlier observation from the polynomial estimate data, this correlation increased to −0.47. Further, Bates and Eysenck also removed a further 5 cases from the dataset on the basis of "bad fit" of the data by the polynomial function, this had the effect of increasing the correlation to −0.62. However, it is demonstrated in this paper that when systematic, explicit, and quantified, criteria are applied to the outlier analysis, and replication of the results is sought across a further four IT datasets, the correlations between the BRAT algorithm parameters and those produced from 3 curve equation functions are actually equivalent. The average systematic outlier-corrected correlation between IT and IQ for both the BRAT and cubic polynomial estimates is −0.34. Further, the unadjusted correlations between BRAT IT estimates and cubic polynomial estimates all exceed 0.95, across all 5 datasets. It is concluded that given the relative difficulty of producing exact polynomial estimates at 0.76 response probability, the inappropriate use of a cubic polynomial for a function bounded by (0, 1), and the perhaps inappropriate data produced by the BRAT algorithm for this type of approach to IT estimation, the use of the curve-fit procedure is sub-optimal with regard to this particular form of IT estimation algorithm.

Keywords: inspection time, test–retest reliability, intelligence

INTRODUCTION

Inspection time (IT) may be estimated using several different procedures of estimation (Deary and Stough, 1996). The BRAT algorithm (Barrett, submitted) is a heuristic, adaptive staircase procedure, that produces estimates of IT and task performance variability. There are 3 phases in the BRAT algorithm, the first achieves an initial "rough" estimate of IT for an individual, the second refines this estimate within the IT duration "region of interest", and the third attains a maximum 2 msec resolution within the IT duration region of interest. The BRAT algorithm is structured to begin above an individual's IT with large duration step sizes, then after making a rough estimate of where an individual's IT might lie, presents trails that begin 30 msec below this approximate IT duration. Finally, when a more accurate estimate of IT is resolved within this phase of the procedure, the next block of trials begin at a duration 20 msec below that of this second phase estimate of IT. An individual's IT is then "resolved" when they can complete 9 consecutively correct trials in this third phase of assessment. Single day, test–retest reliability of this estimate is 0.75 ($N = 72$), and as reported below, one year duration test–retest reliability is 0.83 ($N = 40$). The median time to complete the latest version of the BRAT task is 3.87 min ($N = 142$).

Bates and Eysenck (1993) examined the IT records of 70 participants who had taken part in an experiment investigating the relationship between ongoing multichannel EEG and performance on

✠Deceased.

a cognitive task (IT). The participants were required to complete the IT task while seated alone in a darkened room, wearing an electrode headcap, and being continually monitored via infrared CCTV. Instead of using the IT estimates provided as part of the BRAT task algorithm, Bates and Eysenck computed an estimate using a third order cubic polynomial function that used the response probabilities calculated from a participant's trial log. These probabilities were formed by noting all the discrete trial durations that were administered to a participant, and summing the number of trials that were responded to correctly at each of these durations by this participant. Then, this number was expressed as a ratio with respect to the total number of trials presented at each particular duration. For example, if 10 trials had been presented at say 60 msec, and 6 were responded to correctly, the response probability would be 0.60. Where these probabilities dropped below 0.5, they were constrained to 0.5. Thus, for every participant, a datafile was created which consisted of a set of durations and the probability of obtaining a correct response at each duration. These data were then fitted with a cubic polynomial function, and IT was estimated for a response probability of 0.76 accuracy. The correlation between IQ (as measured by the Jackson Multidimensional Aptitude Battery, 1984) and the cubic polynomial estimate was -0.35. When removing one outlier observation from the polynomial estimate data, this correlation increased to -0.47. Further, Bates and Eysenck also removed a further 5 cases from the dataset on the basis of "bad fit" or meaningless estimates of the data by the polynomial function, this had the effect of increasing the correlation to -0.62. Given the correlation between BRAT IT and IQ was only -0.35 in this sample, the new procedure seemed to offer a major advantage over and above the algorithm estimate. The size of the value obtained by Bates and Eysenck had significant repercussions for the measurement and substantive theory that underpinned the rationale of the IT task.

However, there were several worrying features of this analysis. Firstly, only 70 of a possible 88 cases were used for the analysis (the other 18 "missing" cases were assessed on the performance IQ test subset only). Secondly, there was no reason offered as to why a cubic polynomial function should be preferred, or if it was even optimal. Thirdly, cases were dropped from the analyses based not upon criteria that were explicit and computationally replicable, but solely upon verbal "judgement" statements. That is, there was no clear, replicable, quantifiable or decision-rule based method of outlier detection reported. Fourthly, no attempt was made to compare the BRAT IT estimate to those provided by the equation function. Fifthly, there was no mention of the fact that some probability estimate for an individual might only be based upon 1 observation at a particular duration. If a participant gets this trial wrong, their response probability is set at 0.5. If they get it right, their response probability is 1.0 (the BRAT algorithm is an adaptive procedure, it is not attempting to make systematic observations at specified levels of durations).

Therefore, it was decided that given the potential significance of the Bates and Eysenck result, a systematic examination of this new procedure was required that addressed the misgivings noted above. In order to examine the first issue, we used the entire dataset available to us for the analyses. With regard to the second issue, we used 6 mathematical functions to compute IT duration estimate. Three of these (negative exponential, logistic, and Gompertz functions) seemed to be better suited to probabilistic estimates where the functions could be explicitly constrained within the 0–1 range of the response probabilities. In respect of the third issue, systematic and explicit outlier detection was undertaken, using a 99% bivariate confidence ellipse on every correlation scatterplot as the method of case removal. The only other procedure for data removal/non-existence was where the equation function estimation procedure was unable to resolve the equation parameters, or where the equation produced negative estimates of IT durations, or values above 1000 msec or more. The fourth issue was examined by computing the correlations between BRAT IT estimates and equation function estimates. Finally, in order to ensure a more reliable and robust set of results, these procedures were computed over 4 extra datasets that were available to us (from the Institute of Psychiatry's Biosignal Lab IQ project database).

METHOD

Participants

Three groups of adult volunteers were used in this study, each group named according to the project in which they had participated. All groups were tested as part of the series of experiments

into the biological correlates of intelligence carried out at the Biosignal Laboratory, Institute of Psychiatry. The TITAN group consisted of 25 males with mean age 33.6 and SD of 11.53 years, and 63 females with mean age 37.03 and SD of 10.13 years. This was the group of 88 participants who provided the IT data in Bates and Eysenck (1993). The BIOCOG2 participant group also consisted of 25 males with mean age 31.60 and SD of 9.80 years and 63 females with mean age 36.79 and SD of 8.94 years (Barrett and Eysenck, 1994). These 88 individuals had not taken part in the previous TITAN study. The third group of volunteers were from the BIOCOG3 experiment, which was a replication of the BIOCOG2 study, carried out 1 year later. The sample consisted of 14 males with mean 35.71 and SD of 8.83 years, and 40 females with mean 38.73 and SD of 8.47 years. These 54 individuals had taken part in BIOCOG2 and thus were able to provide information on the test–retest stability of many of the coefficients assessed. All participants had completed Jackson's Multidimensional Aptitude battery (1984), a group administrable ability test constructed to match as closely as possible the WAIS-R intelligence test. However, within the TITAN group, only 70 participants completed the whole test (performance and verbal subsets), the remaining 18 had only completed the performance IQ subset.

For the latter 2 studies, participants were required to complete the IT task in a separate "performance lab" prior to being connected to an EEG electrode headset, then completing the IT task again in the "EEG lab" whilst their cortical activity was continuously recorded. The TITAN experiment only acquired IT data whilst participants were in the EEG Lab, with their ongoing EEG being recorded.

The IT task

IT was assessed using an inverted U display that was created using light-emitting diodes (LEDs) rather than tachistoscopic illumination of cards. The perceptual images of lines were produced by illuminating the LEDs in predetermined sequences such that 2 bars or "lines of light" could be generated, 1 bar longer on one side of the inverted U than the other. The stimulus presentation, response logging, and all stimulus timing was implemented using dedicated electronics, all housed in a "stimulus box" that was connected to a PC via a direct bus interface. The PC computer implements the IT algorithm and communicates with the stimulus box via the computer bus interface. The LEDs are flush mounted into a matt black panel that forms the front face of the stimulus box. Stimulus timing was discrete, in 2 msec units. Response timing is also discrete, counting in 1 msec units. All timing is measured using hardware clocks within the apparatus rather than relying on software clocks. A software trigger initiates a complete trial sequence within the stimulus box, this sequence controlled entirely by the stimulus box hardware. Two microswitch buttons on 1 m length cables are connected to the back of the stimulus box; these buttons are used by a participant to indicate a response i.e. press the right microswitch if the bar is longer on the right-hand side of the display or press the left microswitch if the bar appears longer on the left-hand side. The LED bar display can be energised in three standard ways, showing a longer red bar on the right side, the left side, and a mask that energises all LED bars that are not illuminated as part of the stimulus. Headphones worn by a participant deliver a warning tone prior to the commencement of each IT trial. The full details of the apparatus dimensions and electronics are given in Barrett (submitted).

The instructions given to a participant in each of the experiments were:

(1) *This task will be measuring how little time you need in order to accurately discriminate between one short and one long bar of light. The long bar will be randomly varied between the left and right positions. Whichever side you see the long bar on, press the button which matches that position (left or right).* ***This is not a reaction time task—you have as much time as you like in which to respond.*** *You can make your response whenever you like.*

(2) *For each trial you will hear a beep from the headphones, and the little light will illuminate on the bar display. Focus on this light. After a short interval, the two lines will be illuminated. You press the button which indicates the longer of the two lines.*

(3) The participant is now given a 510 msec duration practice trial—with the longer bar on the left. The tester then states—*as you can see on this easy trial, the left line was the longer.* Note that *both lines are extended after a short duration, then the whole display is switched off.* The participant is then given more practice trials with feedback—at the tester's discretion.

(4) At this point, the participant is asked if he/she wants more practise—if yes, the participant is given more trials at durations specified by the tester. The tester can continue to give the trials for as long as the participant requests.
(5) When the participant indicates an understanding of what is required the tester states: *When we begin, the trials will be given one after the other with exactly the same sound sounds etc. as in the practice. They will start off very easy but will get more difficult as you proceed. Sometimes you may feel that you just cannot see any difference between the lines because the display was so fast. Don't worry, just guess. However, stay interested because you will find that soon after, you will begin to see the differences again after a few trials.*
(6) *It is important that you maintain your concentration during the trial sequences. The experiment lasts less than 10 min. The computer will not let it continue longer. So don't be worried that you may be doing this task for a long time—it may seem like it but it is pretty short in reality. Remember, don't try to look at your watch or even speak. This task needs all your concentration.*

Within the performance lab test sessions, the IT task was terminated automatically either when the participant's IT was found or when the duration of 10 min **experiment** (not including practice) **test time** is exceeded. Within the EEG lab, for experiments BIOCOG2 and BIOCOG3, the IT task was terminated after 5 min if resolution of a participant's IT was not made earlier. The TITAN time limit was initially set at 15 min, then later to 10 min.

Algorithmic estimation of IT

The BRAT algorithm represents a heuristic approach to the direct measurement of a participant's IT, accurate to 2 msec resolution and to a **measured** minimum 90% consecutive response accuracy. There are three phases in the BRAT algorithm, the first achieves an initial quick estimate of IT, the second refines this estimate within the durations of interest, and the third attains a maximum resolution within the IT duration area of interest for a participant.

Phase 1 rules:

- Begin the algorithm presenting an initial duration of 510 msec. This is the longest possible display duration.
- If the participant gets 4 consecutive trials right, then the next duration selected is 100 msec shorter (410 msec). If the participant gets 2 consecutive trials wrong, then this duration is increased by 100 msec unless already at 510 msec. This process continues until the duration is below 200 msec, then the stepsize is reduced from 100 msec to 20 msec. When the duration is below 100 msec, the stepsize is further reduced to 10 msec.
- The phase 1 "trip time" or initial IT estimate is computed using the following rule shown in Fig. 1 opposite.

Phase 2 rules:

Continuing on from the identification of the Phase 1 time, further trials are presented.

- If the participant gets 5 consecutive trials correct, then the next duration selected is 6 msec shorter. If the participant gets 2 consecutive trials wrong, then this duration is increased by 6 msec.
- The phase 2 "trip time" or initial IT estimate is computed using the rule given in Fig. 2 opposite.

Phase 3 rules:

Continuing on from the identification of the Phase 2 time, further trials are presented.

- If the participant responds correctly to a single trial, then the next duration selected is the same as the current duration. If the participant responds incorrectly to a single trial, then the stimulus duration is increased by 2 msec.
- After 9 consecutive correct responses (no errors) the current duration is assigned as the participant's IT and task is ended.

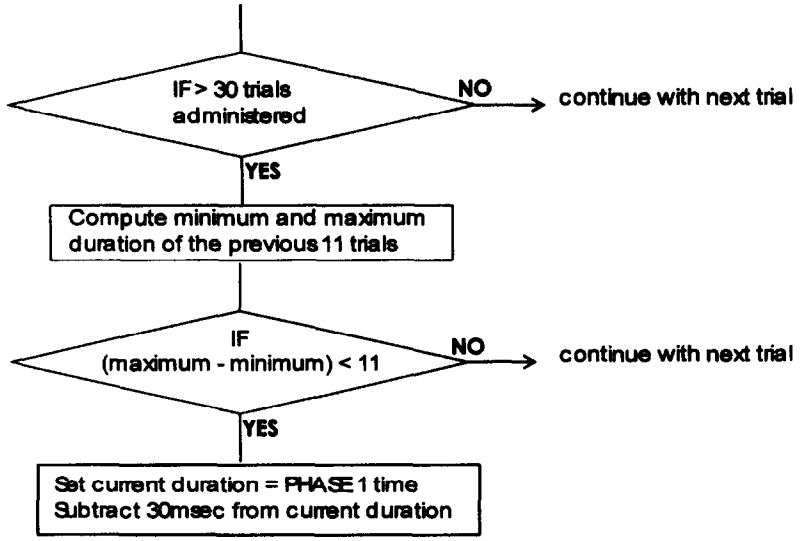
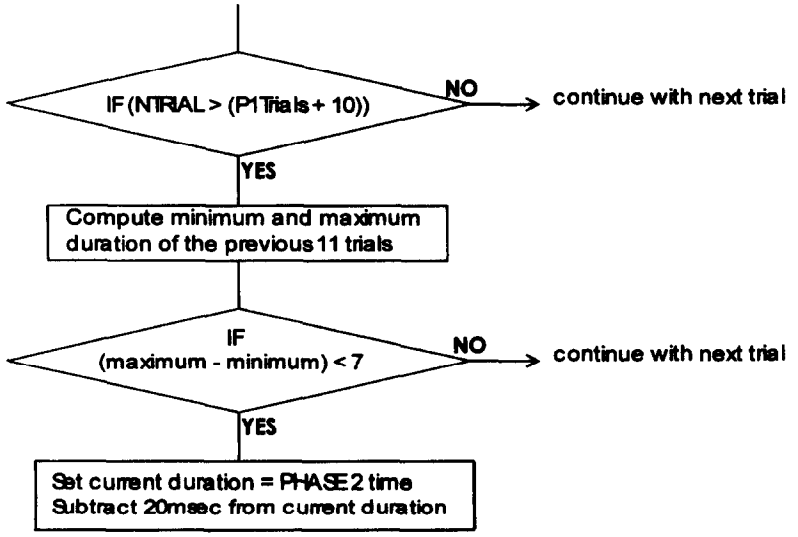


Fig. 1. The decision process required to resolve the Phase 1 time parameter.

Restarts

A correction for the potential error of some participants being pushed into PHASE 3 at a display duration that was still too easy for them is made as part of the algorithm. That is, they make no errors at a duration 20 msec shorter than their PHASE 2 time (the underlying assumption in the algorithm is that this display time should be too short for a participant to respond correctly to the stimulus). This correction is called a "Forward Restart". The "Forward Restart" is operationally defined in Fig. 3 over:

In addition, some participants, when in Phase 3, make so many errors that they are likely to be given too many trials that only decrease in difficulty in small 2 msec steps. Therefore, a correction



Where:
NTRIAL = the number of trials administered so far
P1Trials = the number of trials administered to resolve the Phase 1 trip time

Fig. 2. The decision process required to resolve the Phase 2 time parameter.

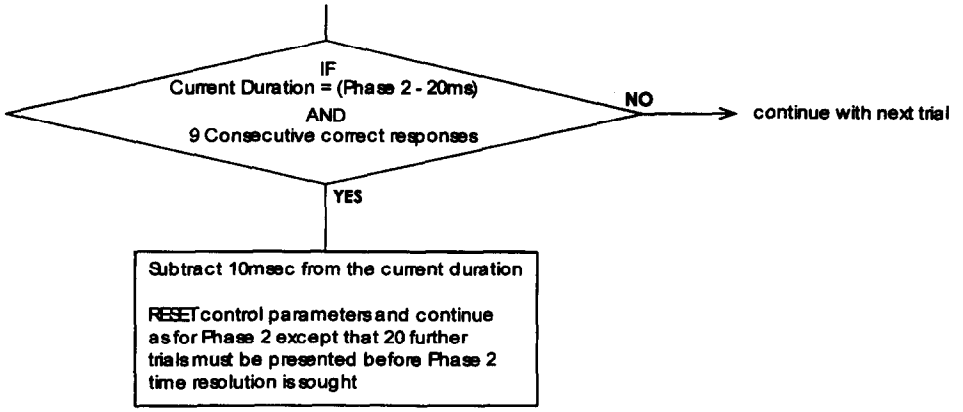


Fig. 3. The BRAT forward restart decision rule.

was applied under certain conditions known as a "Backward Restart". The "Backward Restart" is operationally defined in Fig. 4 below:

The BRAT algorithm is structured to start above the participant's IT, then to overshoot the first approximation (Phase 1 time), then also overshooting the second approximation (Phase 2 time). In all, 31 parameters are computed from the participant's performance on the task. The detailed description of the algorithm, its performance characteristics, and an explanation of all these parameters is given in Barrett (submitted). Within this particular paper, we are interested solely in the estimate of IT itself, which is yielded in Phase 3 of the task. Where a "timeout" occurred, we also have access to Phase 2 times.

Response probability curve fitting

For the initial analysis of the TITAN dataset two sets of response probability files were computed. Set 1 used the observed response probability values, the other (Set 2) constrained probability values of less than 0.5–0.5 (replicating Bates and Eysenck, 1993). The calculation of the observed probability values proceeded by scanning through the entire set of stimulus durations administered to a participant, aggregating the correct/incorrect responses for each discrete duration then finally expressing the response probability as:

$$P_r = \frac{\text{no. correct}}{\text{total no. of trials administered}} \tag{1}$$

Six mathematical functions were then fitted to these data. Curve fitting was in every case implemented

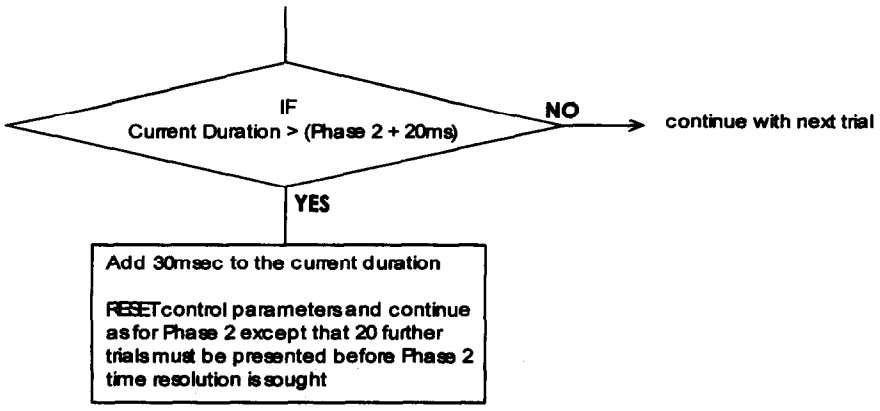


Fig. 4. The BRAT backward restart decision rule.

using the STATISTICA 5 nonlinear module, using primarily the Quasi–Newton and Simplex + Quasi–Newton iterative parameter solutions.

Linear

$$P_r^* = b_0 + b_1 IT_{dur} \tag{2}$$

where b_0 = intercept parameter, b_1 = slope parameter, P_r^* = the estimated response probability, IT_{dur} = IT duration

Quadratic Polynomial

$$P_r^* = b_0 + b_1 IT_{dur} + b_2 IT_{dur}^2 \tag{3}$$

Cubic Polynomial

$$P_r^* = b_0 + b_1 IT_{dur} + b_2 IT_{dur}^2 + b_3 IT_{dur}^3 \tag{4}$$

Negative Exponential

$$P_r^* = b_0(1 - e^{-b_1 IT_{dur}}) \tag{5}$$

Logistic

$$P_r^* = \frac{b_0}{1 + b_1 e^{-b_2 IT_{dur}}} \tag{6}$$

Gompertz

$$P_r^* = b_0 e^{-b_1 e^{-b_2 IT_{dur}}} \tag{7}$$

In the case of the negative exponential, logistic and Gompertz functions, the parameter b_0 should have a value of 1.0 (as the theoretical asymptote of the response probability function). Although it could have been fixed at this value, we allowed the estimation process to estimate this parameter along with the others required in each equation. This had the advantage of providing a cursory check on the validity of the final estimated equation. If this parameter was substantively different from 1.0 (generally outside the limit 1.0 ± 0.05), then this was a clear indication that an improper solution had been calculated. Final parameter estimates for these functions used a fixed value of 1.0 for the asymptote parameter.

In order to replicate Bates and Eysenck (1993), estimated ITs at particular response probabilities were computed at 0.76 for each participant. For interest, estimated ITs were also computed for 0.80, and 0.90 levels of probability. In order to achieve accuracy of estimation, each of the 6 equations above was re-expressed as a function of IT_{dur} using the MATHCAD 4.0 algebraic equation module.

Linear

$$IT_{dur}^* = \frac{0.76 - b_0}{b1} \tag{8}$$

where IT_{dur}^* = estimated IT at 0.76 probability of accuracy of response

Quadratic

$$IT_{dur}^* = \frac{-1.0}{(2b_2)}(b_1 - \sqrt{b_1^2 + 4b_2 0.76 - 4b_2 b_0}) \tag{9}$$

The cubic function estimation produced 3 forms of expression, that each require a whole printed page. These functions can be obtained from the first author.

Negative Exponential

$$IT^{*}_{dur} = -1.0\frac{\ln\left[-1.0\left(\frac{0.76-b_0}{b_0}\right)\right]}{b_1} \tag{10}$$

Logistic

$$IT^{*}_{dur} = -1.0\frac{\ln\left[-1.0\left(\frac{0.76-b_0}{0.76b_1}\right)\right]}{b_2} \tag{11}$$

Gompertz

$$IT^{*}_{dur} = -1.0\frac{\ln\left[-1.0\left(\frac{\ln\left(\frac{0.76}{b_0}\right)}{b_1}\right)\right]}{b_2} \tag{12}$$

The analysis strategy

Stage 1. The analyses implemented here were designed to determine whether using unconstrained (set 1) or constrained (set 2) probability files produced any substantive differences in parameter estimates. Since constraining the data in fact requires "changing" the data, it is worth knowing whether such a constraint has deleterious or advantageous effects upon parameter estimation. Forty cases from the entire set of 88 cases TITAN EEG lab IT data were used here (in fact only 39 were able to be used as one participant (#13) yielded unusable data due to suffering a panic attack in the EEG lab). We had originally planned on calculating parameters over all cases, but sequential analysis of the results indicated that it was unnecessary to proceed beyond the cases analysed.

Stage 2. Using the results in Stage 1, the next analysis question to be answered was which mathematical function is the best, in terms of percentage fit (least-squares criterion) to the data and the case of attaining a numerically stable solution. The TITAN dataset was again used exclusively here.

Stage 3. Having established the optimal function/s for the data, we then focused upon the behaviour of these function estimates in terms of how they replicated across datasets, the size of the correlations between IQ and the IT estimates, and how the function estimates performed with respect to the BRAT algorithm parameters. The five experiment datasets were used here, the TITAN data, BIOCOG2 performance and EEG lab files, and BIOCOG3 performance and EEG lab files. Parameter reliability analysis was also undertaken as 1 year test–retest data was available for a subset of BIOCOG2-3 participants ($N = 54$).

RESULTS

Stage 1

From the calculations made using the 39 cases from the TITAN dataset, using constrained vs unconstrained probability files, the results in Table 1 report the percentage fit of each function to set 1 (unconstrained) and set 2 (constrained) data.

From this Table, it is apparent that Bates and Eysenck (1993) made the correct decision to

Table 1. The percentage fit (least squares criterion) of each of 6 mathematical functions to the TITAN subset data ($N = 39$), for set 1 (unconstrained) and set 2 (constrained) response probability files

Function	Set 1 datafiles	Set 2 datafiles
Linear	17.1	25.2
Quadratic	25.7	36.1
Cubic	31.0	41.8
Negative exponential*	30.9	44.2
Logistic	34.4	43.6
Gompertz	34.2	43.2

* Note: the number of cases for the negative exponential fit was 36 for set 1, and 31 for set 2 cases.

constrain the probability values to a minimum bound of 0.5. This has a marked effect on the fit of all functions to the data. In terms of the differences in parameter estimation between sets 1 and 2, using the estimated IT value computed from each function as the indicator value, the correlations between the estimated ITs at 0.76, 0.80, and 0.90 response probability were all greater than 0.90. The majority were greater than 0.95. From the results in this stage of analysis it was decided to proceed with the further analyses, using only the constrained probability files.

Stage 2

Using the entire TITAN dataset of 88 cases, IT parameter estimates were computed by fitting all 6 mathematical functions to the cases. The percentage fit values follow closely the values presented in Table 1 above, for set 2 data. For example, the 0.76 response probability percentage fit statistics for the cubic polynomial data ($N = 87$ cases, case #13 unusable as noted above) were 44% median fit, minimum of 15%, maximum of 75%, with an interquartile range of 30–55%. From this pattern of results, all functions bar the linear one look acceptable. However, more important was the fact that the linear function was providing a greater proportion of wildly inaccurate estimates (negative durations or impossibly large durations). This was not really surprising as the function to be fitted is essentially non-linear (as shown in Fig. 5 below, and in Bates and Eysenck (1993)) although linear in algebraic form.

The second function to be dropped was the negative exponential. As noted in Table 1, 8 cases could not be fitted out of the initial 39 analysed. In total, 16 out of the 87 available cases could not be fitted (the nonlinear iterative process either refused to converge or produced an invalid parameter

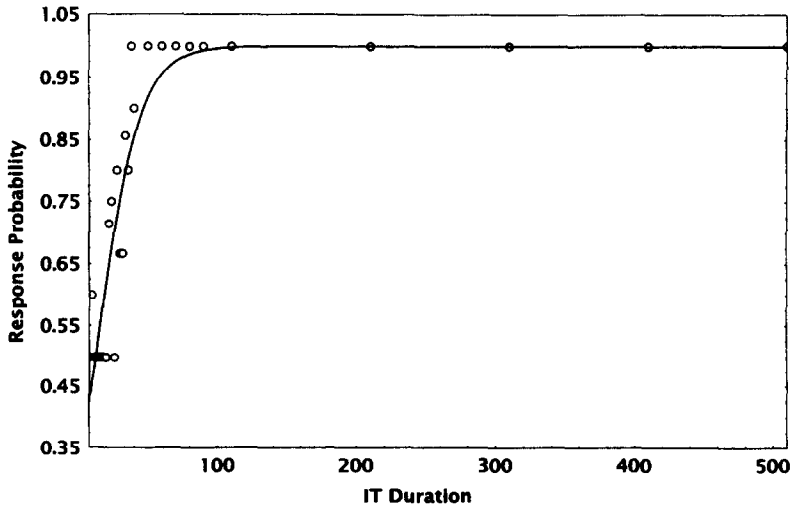


Fig. 5. The logistic function curve-fit for case #47, TITAN data. Using least squares loss, explained variation is 86%.

set, leading to ridiculous duration estimates). Overall, this was seen as a poor analysis tool, even if the percentage fit was comparable to other nonlinear functions; too many cases were simply unable to be estimated. Additionally, correlations between the IT estimates from this function and those from the Logistic and Gompertz equations were all above 0.94 for both the 0.76 and 0.80 response probability levels. Therefore, it was also seen as a redundant tool. Finally, the Gompertz function was dropped, solely upon the basis that it was totally redundant. The correlation between IT duration estimates from the logistic function was 1.0. Note that this redundancy is also reflected in the percentage fit functions. The quadratic function was retained solely on the basis that it was a more simple equation to work with, and, when using more datasets for replication purposes, might be shown to yield estimates comparable with the cubic function, and thus perhaps replace it altogether. The correlation between IT estimates provided by the quadratic and cubic polynomials was greater than 0.96 for the three response probabilities. Thus, from stage 2, we retained the quadratic, cubic, and logistic functions for the final analyses that compare this class of estimated IT parameters with those provided by the BRAT algorithm.

Stage 3

Table 2 below shows the correlations between performance IQ, IT equation function estimates, and BRAT Phase 2 and final IT estimates. The function IT estimates are all for a response probability of 0.80. The correlation between estimates for this level and computed for the 0.76 response probability was virtually 1.0 in all cases. There was also little difference between these estimates and those made at the 0.90 level of response probability, although generally the 0.90 level correlations were slightly lower overall than those computed using the 0.80 level. As will be noted from Tables 2, 3, and 4, below, the numbers of cases for which IT parameter estimates could be computed varies across estimation methods. This is due to the fact that for some cases, either no equation could be computed or no values estimated because of the pattern of response probabilities, or in the case of the BRAT algorithm, the participant failed to complete the IT task to the level of either Phase 2 or Phase 3.

Table 2. Correlations between equation function and BRAT IT measures with performance IQ in 5 datasets

Experiment	Estimate	Actual	With Perform IQ Pass 1	Pass 2	Pass 3
TITAN	2nd Poly	−0.37(87)	−0.42(85)	−0.39(83)	−0.33(82)
	3rd Poly	−0.34(87)	−0.39(85)	−0.34(83)	−0.32(81)
	Logistic	−0.42(86)	−0.36(84)	−0.36(83)	−0.33(81)
	Phase 2 IT	−0.40(86)	−0.34(84)	−0.26(82)	−0.26(80)
	IT	−0.33(73)	−0.29(72)	−0.31(70)	−0.31(68)
BIOCOG2 Performance	2nd Poly	−0.36(83)	−0.47(79)	−0.44(77)	−0.43(76)
	3rd Poly	−0.39(83)	−0.43(79)	−0.46(77)	−0.45(76)
	Logistic	−0.36(82)	−0.41(78)	−0.47(77)	−0.45(76)
	Phase 2 IT	−0.42(87)	−0.29(83)	−0.36(80)	−0.41(79)
	IT	−0.35(82)	−0.31(79)	−0.42(77)	−0.38(74)
BIOCOG2 EEG	2nd Poly	−0.38(82)	−0.23(80)	−0.30(78)	−0.38(76)
	3rd Poly	−0.38(82)	−0.23(80)	−0.36(76)	−0.38(74)
	Logistic	−0.25(82)	−0.21(80)	−0.34(77)	−0.34(74)
	Phase 2 IT	−0.25(85)	−0.23(84)	−0.35(81)	−0.32(80)
	IT	−0.24(71)	−0.18(70)	−0.33(68)	−0.31(67)
BIOCOG3 Performance	2nd Poly	−0.31(52)	−0.33(50)	−0.16(48)	—
	3rd Poly	−0.27(52)	−0.19(50)	−0.13(49)	—
	Logistic	−0.24(52)	−0.16(49)	—	—
	Phase 2 IT	−0.42(54)	−0.16(51)	−0.22(49)	—
	IT	−0.42(50)	−0.20(48)	−0.29(46)	—
BIOCOG3 EEG	2nd Poly	−0.34(53)	−0.31(51)	—	—
	3rd Poly	−0.34(53)	−0.34(51)	−0.40(50)	—
	Logistic	−0.34(53)	−0.32(51)	−0.38(50)	—
	Phase 2 IT	−0.46(54)	−0.29(53)	−0.33(52)	—
	IT	−0.36(43)	−0.42(42)	—	—

Table 3. Reliabilities for 5 IT measures computed using the BIOCOG2 and BIOCOG3 datasets. The equation function values are those computed for a 0.80 level response probability. The datasets are those from the ITs assessed within the PERFORMANCE laboratory

Experiment	Estimate	Actual	With Corresponding BIOCOG3 Measures Pass 1	Pass 2	Pass 3
BIOCOG2 Performance	2nd Poly	0.71 (49)	0.60 (46)	0.46 (45)	—
	3rd Poly	0.72 (50)	0.66 (47)	0.47 (46)	0.44 (45)
	Logistic	0.72 (50)	0.72 (47)	0.73 (44)	0.65 (43)
	Phase 2 IT	0.64 (53)	0.79 (49)	0.73 (47)	0.80 (43)
	IT	0.82 (46)	0.71 (44)	0.83 (40)	—
	ABSRANK	0.61 (54)	0.74 (52)	—	—

Table 4. Reliabilities for 5 IT measures computed using the BIOCOG2 and BIOCOG3 datasets. The equation function values are those computed for a 0.80 level response probability. The datasets are those from the ITs assessed within the EEG laboratory

Experiment	Estimate	Actual	With Corresponding BIOCOG3 Measures Pass 1	Pass 2	Pass 3
BIOCOG2 EEG	2nd Poly	0.68 (51)	0.69 (48)	0.80 (47)	0.85 (44)
	3rd Poly	0.67 (51)	0.76 (48)	0.83 (46)	0.86 (43)
	Logistic	0.39 (51)	0.78 (48)	0.82 (45)	0.74 (43)
	Phase 2 IT	0.12 (53)	0.42 (51)	0.61 (50)	0.61 (48)
	IT	0.35 (37)	0.80 (36)	0.74 (33)	0.74 (31)

The four columns of correlation values are defined as follows:

Actual: the correlations in this column are those computed for all available cases in each particular dataset, for each parameter type. The figures in brackets are the number of cases over which each correlation is computed.

Pass 1: each correlation in the "actual" column was graphically examined using a scatterplot with a 99% bivariate confidence ellipse superimposed over the points. As Stevens (1996) has indicated, this is a reasonably generous bound, outside of which any observation might be considered an outlier. Given the relatively small sample sizes used here, it is wise to check each correlation for potential bias/influence by extreme observations, and remove them systematically according to a fixed rule. In this way, a conservative (and hopefully more robust) estimate of each correlation may be present. The values in this column are those resulting from an initial screen with the bivariate ellipse. Observations lying outside the ellipse were deleted, and correlations recomputed on the remainder.

Pass 2: a second 99% bivariate confidence ellipse was superimposed over the observations plotted in the scatterplot after Pass 1 had been implemented. Any observations falling outside this ellipse were again deleted, and the correlation recomputed.

Pass 3: a second 99% bivariate confidence ellipse was superimposed over the observations plotted in the scatterplot after Pass 2 had been implemented. Any observations falling outside this ellipse were again deleted, and the correlation recomputed. The exercise ceased here as a plateau was reached in the correlations such that either no more observations were able to be removed (consistent with the rule implemented here), or where further removal did not cause a change in the resultant correlation coefficient for any pair of variables.

As can be seen from the results in Table 2, there is little evidence of a consistent difference between the BRAT × performance IQ parameter correlations of Phase 2 IT, IT, and those of the equation function correlations.

Another question that can be asked here is, how do the test–retest reliability estimates computed using the equation functions compare to those computed using the BRAT algorithm? This question can be answered by using the 54 case subset of participants who completed both the BIOCOG2 and the BIOCOG3 experiments, with a year intervening between test sessions. Tables 3 and 4 below provide these test–retest coefficients for the performance and EEG-lab datasets, with outlier analysis implemented in the same way as for Table 2 above.

Looking at these values, it is apparent that the BRAT parameter estimates are more stable than

Table 5. The percentage fit and case IDs of the participants (with a full-scale IQ score) who were rejected as outliers using percentage fit as the criterion, and those who were rejected solely to boost the correlation between estimated IT and IQ

Participants with the Worst % Fits		Participants removed to boost correlation	
Participant	% Fit	Participant	% Fit
17	22.6	7	58.0
19	23.9	14	55.0
32	15.2	50	31.2
42	21.0	56	55.4
43	21.0	58	62.1
72	21.0	63	29.4
Mean Fit = 20.8%		Mean Fit = 48.5%	

Note: The total sample ($N = 87$) mean fit for the 3rd order polynomial is 43%.

those computed using the equation functions, but, for the EEG dataset, the opposite is the case (although the discrepancy is less). Using the correlations and estimates above, there seems little difference between the results provided by one parameter estimation procedure over another.

Interestingly, in Bates and Eysenck (1993), they were able to delete 6 cases from the TITAN dataset (using only those 70 cases who had completed all of the Jackson MAB test) on the basis of bad percentage fit and meaningless parameter estimates. Recomputing the correlation between the remaining cases and full-scale IQ yielded a correlation of -0.62 between IQ and estimated IT (at $p_r = 0.76$). In order to replicate this result more closely, we excluded the 6 worst fitting cases from the reduced participant dataset ($N = 70$) and computed the correlations between the IT estimates (at $p_r = 0.76$) and full-scale IQ. The correlation for the cubic polynomial estimate and full-scale IQ before dropping these 6 cases was -0.35 ($N = 70$). Removing the 6 cases caused the correlation to drop to -0.17. Table 5 provides the details of the participants who were dropped by us, based solely upon the criterion of worst percentage fit of the cubic polynomial, and who also possessed a full-scale IQ score.

Whilst we were engaged in this activity, we decided to see whether we could obtain a similar result to Bates and Eysenck by selecting any 6 cases, regardless of any criterion, with the sole purpose of attaining a correlation nearer that reported by Bates and Eysenck. Table 5 also provides the information on 6 cases, that when removed, yield a correlation between full-scale IQ and the cubic polynomial estimate of -0.60 ($N = 64$). Figure 7 provides the scatterplot with the worst fitting cases, and the excluded cases appropriately marked.

As can be seen from Fig. 7 and Table 5, bad fit (as defined by percentage fit of the function to the data) cannot be the sole criterion for exclusion that was used by Bates and Eysenck.

In comparing BRAT IT estimates and those from the equation function estimates, across all 5 datasets, it is noted from the results in Table 6 that there is no substantive difference at all between the estimates from either procedure.

The coefficients in this table demonstrate that curve fitting of IT data is most unlikely to ever diverge significantly in parameter relationship terms from IT data acquired using the BRAT algorithm. This is important when considering the Bates and Eysenck (1993) result that demonstrated a substantive increase in polynomial estimated IT × IQ correlation, using the three parameter cubic model. The results in Table 6 show that BRAT IT would also have correlated more or less the same with IQ. For example, within the original TITAN dataset, the correlation between polynomial estimated IT and IQ (with the 6 arbitrary cases missing as shown in Table 5) was -0.60. The BRAT IT correlation with IQ, for the same subset of cases, was -0.53 (although N was only 53 instead of 64 due to 11 cases not having an IT estimate due to being timed out on the task). Using the Phase 2 times (as noted in the Method section above), we observed an IT × IQ correlation of -0.59 on 63 out of the 64 subset cases.

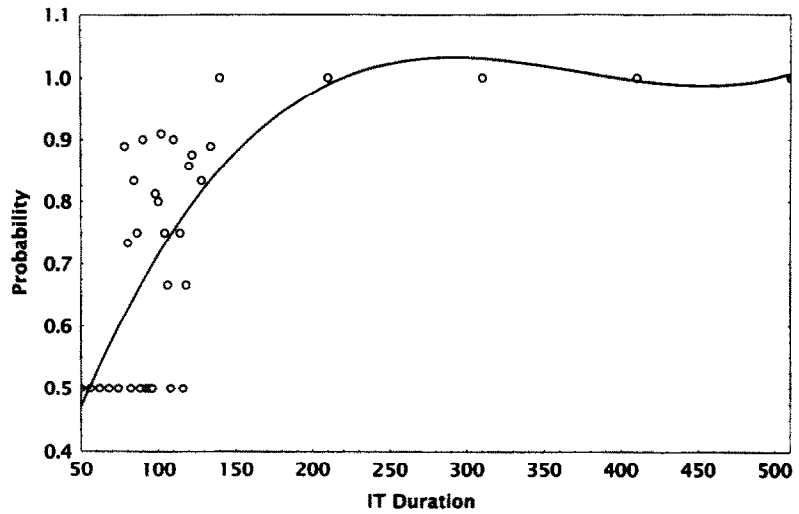


Fig. 6. The cubic polynomial function curve-fit for case #33, TITAN data. Using least squares loss, explained variation is 49%.

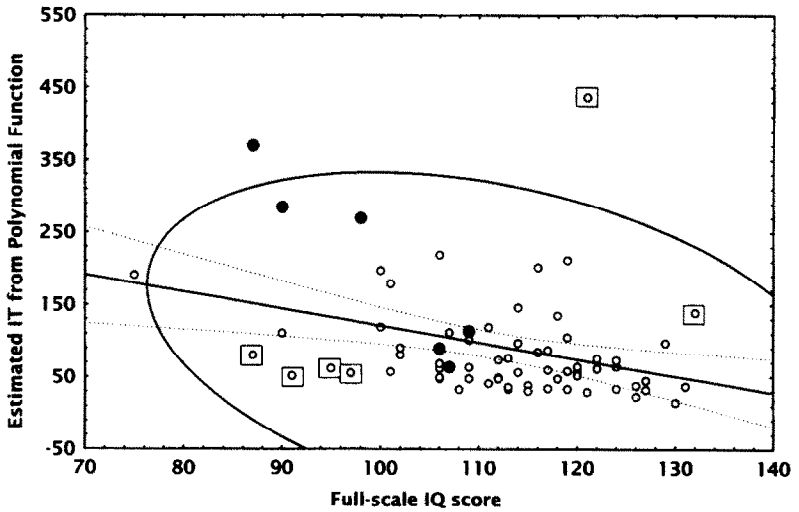


Fig. 7. Estimated IT vs Full-Scale IQ. The solid filled markers are those with the lowest percentage fit values as detailed in Table 5. The markers surrounded by a square correspond to those cases removed in order to maximise the IQ × IT correlations. A 99% confidence ellipse is also plotted on this graph.

Table 6. The correlations between the BRAT IT estimates and those provided by a quadratic and cubic polynomial function, and a logistic curve function (estimated at 0.76 correct response probability). These correlations were computed across all 5 datasets. The figures in brackets are the number of cases over which the correlations are computed. These were all available cases for which there were parameter estimates

Function	Quadratic	Cubic	Logistic
TITAN*	0.95(71)	0.96(72)	0.96(72)
BIOCOG2-Performance Lab	0.96(80)	0.96(80)	0.96(80)
BIOCOG2-EEG Lab	0.98(65)	0.99(68)	0.99(68)
BIOCOG3-Performance Lab	0.95(49)	0.96(49)	0.94(49)
BIOCOG3-EEG Lab	0.90(43)	0.96(43)	0.98(43)

* Note: one case with IT > 440 msec was excluded from this dataset. This case was 7 SDs away from the mean value of IT in this particular dataset.

DISCUSSION

From the rather exhaustive results above, it is apparent that there is little to recommend the use of curve fitting procedures to the estimation of IT, using the BRAT algorithm data. There appears to be no advantage whatsoever in using one form of parameter estimation over another, with regard to the correlations between these parameter estimates and IQ. However, given that the BRAT IT parameter requires no special calculations or nonlinear estimation process, it is simply easier to continue using this parameter than adopt a complex statistical estimation process that generates parameters that appear subsequently to correlate 0.96 with simple BRAT estimates. Further, as indicated in the introduction, it is questionable whether using such curve-fit procedures is supportable on this kind of data—there are too many probability estimates being generated that are based upon fewer than 4 trials. These estimates may be quite misleading. However, given the high relationship between these parameters and the BRAT IT parameter, the decision as to which method of estimation to use seems more to do with practicality of parameter assessment than any particular theoretical argument about the preference for fitting psychometric functions. If one were to recommend an optimal numerical function for curve-fitting the IT data, then the logistic equation would appear to be the theoretically most appropriate and convenient, not a cubic polynomial.

A second, more crucial issue, is that the dropping of cases from an analysis can only sensibly be achieved if based upon clear, empirically-based, and replicable criteria. Degree of fit of a function to the data was clearly not the method used to select outliers within Bates and Eysenck (1993). Quoting from the paper:

> "For 6 of the 70 subjects, this procedure was not applicable. These subjects gave chance performance at stimulus durations longer than those at which they showed a high level of accuracy. Because of this the polynomial bounced between 50% and 100%, and gave a meaningless estimate, as well as a very high mean square error..." (p. 527).

How might a "high level of accuracy" be defined? On page 530 of the same paper, the authors go on to state:

> "The IT-IQ correlation was improved both by filtering the data to remove errors at easy durations and, alternatively, simply excluding subjects whose data showed such errors..."

What are to be considered "easy" durations? Figures 3, 4 within their paper show a raw data plot from a "good" vs "bad" case. Both cases make no errors above 150 msec duration. However, there appears to be no simple way of judging whether or not the "bad" case should be rejected. Bates and Eysenck do not indicate in what way/s the two cases are different from one another, and how this would be specified in such a way that any other researcher could follow their procedures rigidly (actually, with these particular cases, percentage fit can be used as a discriminator, but we now know from the above analyses that this criterion by itself is not sufficient to be used as a data screen in order to achieve maximum correlations between IT parameters and fit). The data presented in Table 5 above, and the results computed from the remaining 64 cases, demonstrate that whatever other criteria were used to de-select cases, percentage fit was not one of the main criteria.

Finally, the reader might question the degree of detail, systematic analysis, and detailed exposition of the BRAT algorithm included in this paper. The reason for this is that Bates and Eysenck (1993) published a result that seemed to indicate that an IT $\times$ IQ relationship could be lifted from a moderate -0.35 to a much stronger value of -0.62, using a cubic polynomial curve-fit procedure, and with judicious selection of "bad-fit" subjects. This substantive change in the degree of relationship matters both at an empirical level and at the theoretical level. It is therefore important (if somewhat numerically tedious) to comprehensively examine the methods of acquiring the raw data, the purported empirical phenomena, and the replicability of any computed results using the Bates and Eysenck computational procedure, in order to determine whether their results were artifactual or substantive.

We conclude that contrary to the results presented in Bates and Eysenck, using a psychometric curve-fit procedure on BRAT algorithm data is irrelevant to final computed parameter and correlation values. The equation estimates are considered both redundant, because of the high BRAT vs equation estimate IT correlation, and fundamentally flawed, because of the perhaps inappropriate

adaptive data acquisition algorithm in the BRAT IT task. Of course, one might question the performance and measurement characteristics of the BRAT algorithm itself, but it does appear (from the data reported in this paper) to be generating reliable measurement that correlates in the expected direction and level with performance on cognitive ability tests.

REFERENCES

Barrett, P. T. (Submitted). The measurement of inspection time: The BRAT algorithm. *Behaviour Research Methods and Instrumentation.*

Barrett, P. T., & Eysenck, H. J. (1994). The relationship between evoked potential component amplitude, latency, contour length, variability, zero-crossings, and psychometric intelligence. *Personality and Individual Differences, 16, 1,* 3–32.

Bates, T. C., & Eysenck, H. J. (1993). Intelligence, Inspection Time, and Decision Time. *Intelligence, 17,* 523–531.

Deary, I. J., & Stough, C. (1996). Intelligence and Inspection Time: Achievements, Prospects, and Problems. *American Psychologist, 51,* 6, 599–608.

Jackson, D. N. (1984). *Handbook of the Multidimensional Aptitude Battery.* Port Huron, Michigan: Research Psychologists Press.

Stevens, J. S. (1996). *Applied Multivariate Statistics for the Social Sciences.* New York: Lawrence Erlbaum.