INTELLIGENCE **14**, 309–325 (1990)

Reaction Time Variability and Intelligence: A "Worst Performance" Analysis of Individual Differences

GERALD E. LARSON
DAVID L. ALDERTON
Navy Personnel Research and Development Center

Reaction time distributions from 303 subjects were partitioned into 16 fast-to-slow latency bands. Average latencies for the 16 bands were then correlated separately with various indices of mental ability. The slowest bands (or "worst trials") were by far the best predictors of intelligence and working memory performance. The slowest bands also drove the individual differences in variability, and the variability/intelligence correlation. An error-based production model for worst trials was tried and abandoned.

An axiom of cognitive psychology is that the "global" aptitudes measured by psychometric tests can be decomposed into more elementary processes or components. However, a product of any decomposition can itself be viewed as global in subsequent studies, and there are, indeed, occasions where one wishes that a further partitioning had been undertaken. It is in this light that the current investigation was begun. We were interested in the numerous published reports of a significant relationship between intelligence and variability on "elementary" reaction time tasks, but we did not accept the view that variability is itself an "elementary" measure. Rather, we treated it as a global variable that required decomposition.

BACKGROUND

Jensen (1982), discussing his reaction time (RT) experiments, noted that trial-to-trial variability (the standard deviation of each subject's reaction times) frequently surpassed response speed as a predictor of intelligence. That is, low aptitude individuals were excessively variable from one RT trial to the next,

---

The opinions expressed in this paper are those of the authors, are not official and do not necessarily reflect the views of the Navy Department.

Correspondence and requests for reprints should be sent to Gerald E. Larson, Testing Systems Department, Navy Personnel Research & Development Center, San Diego, CA 92152-6800.

relative to brighter subjects. Currently, numerous studies suggest that his observation was correct and that variability has a robust statistical relationship to intelligence (e.g., Barrett, Eysenck, & Lucking, 1986; Carlson & Jensen, 1982; Carlson, Jensen, & Widaman, 1983; Jensen, 1987; Larson & Saccuzzo, 1989; Nettlebeck & Kirby, 1983; Smith & Stanley, 1983; Vernon, 1983; Vernon, Nador, & Kantor, 1985). In few of these studies, however, is the relationship explored.

One problem is that variability is a global score that summarizes data from tens or even hundreds of individual RT trials. To discover the exact nature of the variability/intelligence relationship it may be necessary to decompose the global variability index into the distribution of its actual constituent RTs, that is, the raw data. Studies that have done this reveal that the RT distributions of low aptitude groups (especially retardates) are mainly aberrant at the slow end, that is, their fastest response nearly match those of brighter subjects, but their slowest responses are uniquely poor (Baumeister & Kellas 1968; Jensen, 1982). Thus, the distinguishing feature of "low aptitude" data, and a source of high variability, is the severity of the deficit exhibited on occasional trials. This suggests a "worst performance rule," such as, "*The worst RT trials reveal more about intelligence than do other portions of the RT distribution.*"

More important, the analysis of variability in terms of individual trials allows us to test specific hypotheses about the origins of the aberrant trials. In the current paper, we (a) verify that the worst trials help drive the variability/intelligence correlation, and (b) explore an error-based production model for the worst-trial deficit of lower mental ability subjects.

Subjects

Subjects were male Navy recruits (N = 332; mean age 19.8 years) selected at random from groups in basic training at the Recruit Training Command, San Diego, CA.

COMPUTER ADMINISTERED TESTS

The computer tests were presented on Hewlett—Packard Integral Personal Computers. All equipment was standard except for a keyboard overlay that exposed only the necessary response keys. The order of test presentation was selected for each subject according to prearranged random sequences. The computer tests were developed by the first author.

Reaction Time

The Arrows Test. In the Arrows test, subjects were instructed to fixate on two small circles (the lowercase letter "o") presented 0.5 horizontal inches apart in the center of the CRT screen. For each trial, one of the circles was replaced by an arrow, and, depending on the arrow's direction and position, the subject responded by pressing either a right or left key on the microcomputer keyboard. If the arrow pointed down, its position indicated the appropriate response. For example, if a down-pointing arrow replaced the right circle, the right key was pressed. If an arrow pointing right or left was presented, however, then its direction became the relevant cue while position became a distractor. For example, if an arrow appearing on either side pointed right, the right key was pressed. The position and direction of the arrow were varied randomly. The test involved 82 trials; 41 with downward arrows and 41 with right-left arrows. Based on our prior research, the test software rejected reaction times greater than 2 s and presented new items to maintain a constant 82 trials per subject. Experience has shown that the 2-s ceiling is generous and produces clean data, since RTs longer than 2 s often reflect events such as interruptions, questions to the proctor, or an occasional computer hardware problem. Final scores are median RT and the standard deviation of the 82 RTs.

In examining the methodology in previous reports of posterror RT slowing, we discovered that subjects typically received no feedback during the task (e.g., Brewer & Smith, 1984; Kiger & Glass, 1981; Rabbitt & Rodgers, 1977). To be consistent with previous research, our subjects received no feedback regarding errors.

Working Memory: Machine-Paced Tests

Machine-paced tests present the subject with a rapid series of video frames, with each frame containing information critical to ongoing cognitive operations. To be successful, subjects must be able to process information in working memory at the rate the frames are presented. These are not reaction time tests—subjects answer by making an untimed, multiple-choice selection, scored as correct or incorrect. Consequently, there is little method overlap between the machine-paced working memory tests and the reaction time test.

Mental Counters. In the Mental Counters test, subjects must keep track of the values of three independent "counters," which change rapidly and in random order. The counters are represented as lines on the video monitor (three side by side 1.0-inch horizontal dashes in the center of the screen). The initial counter values are zero. When a small target (a .25-inch box) appears *above* one of the lines, the corresponding counter must be adjusted by adding "1." When the target appears *below* one of the three lines, the corresponding counter must be adjusted by subtracting "1." The maximum and minimum counter values used in the test were +3 and −3, respectively. The test items vary both in the number of adjustments and the rate of presentation. There were two levels of counter adjustments (five and seven) and two levels of rate of presentation (fast and slow). The actual test involved a total of 40 trials. On 20 trials, five adjustments were required. Seven adjustments were required on the remaining 20 trials. On twenty trials, adjustments were required at the rate of one every .75 s. On the remaining twenty trials, adjustments were required at the rate of one every 1.33 s. Number of adjustments and presentation rate were completely counterbalanced. It was a multiple choice test; the subject was required to chose the correct set of final counter values from the four alternative sets presented. Total correct was used as the summary score. The test is depicted in Larson, Merritt, and Williams (1988) and Larson and Saccuzzo (1989). The difficulty of the task comes from having to simultaneously hold, revise, and store three counter values under severe time pressure. Slow execution of counter adjustments leads to a general breakdown on the task.

Sequential Memory. The Sequential Memory test is presented in two parts. In the first part, the subject is presented with either three, four, or five dots in the center of the screen. Each dot is said to "stand for" a particular single digit number, shown immediately below the dot. For example, in the frame below, the dots stand for 4, 3, and 6, respectively.

Left	Center	Right
.	.	.
4	3	6

Subjects are given 3 s to commit the numbers to memory, then the numbers are erased from the screen. Next, in random order, the dots are briefly changed into an "X." The appearance of the "X" (approximately one "X" every 600 ms) is said to "call" the number previously below the dot, and the subject must remember the sequence of calls. Each item involved five calls, and a unique set of numbers. For example, using the above frame, if the five Xs replaced the dots in this order, center-right-right-left-center, the correct answer would be 3-6-6-4-3. In summary, the subject must remember both (a) the numbers assigned to the dots, and (b) the sequence in which the numbers were called. There were 30 items; 10 each with three, four, and five numbers that could be called. It was a multiple-choice test; the subject was required to chose the correct set of calls from five alternatives presented.

The second half of the test follows the same format, but adds an additional transformation. After the numbers have been assigned to the dots and the sequence of calls has been completed, subjects are told to convert the numbers-in-memory to specified new values, while maintaining the correct order. There were 30 items; 10 each with two, three, and four numbers that could be called. Total correct across both halves of the test was used as the summary score. A sample item is shown in Figure 1. The difficulty of the test comes from having to

Frame	What the subject sees			Correct Answer
	.	.	.	
1	4	7	6	
2	X	.	.	4
3	.	X	.	4,7
4	.	X	.	4,7,7
5	.	.	X	4,7,7,6
6	X	.	.	4,7,7,6,4
7	"4 becomes 2"			2,7,7,6,2
8	"6 becomes 3"			2,7,7,3,2
9	"7 becomes 5"			2,5,5,3,2 Final Answer

FIG. 1. Sample Item from The Sequential Memory Test

simultaneously process the original and revised number sequences in working memory.

PAPER-AND-PENCIL TESTS

Armed Services Vocational Aptitude Battery (ASVAB) scores were gathered from the recruits' personnel records. The ASVAB is a set of 10 tests (listed in Table 1, p. 314) used for selection and classification of military applicants. The tests are scaled to a mean of 50 and a standard deviation of 10 in an unselected, nationally representative sample. In addition, the Raven Progressive Matrices (RPM) Test, Advanced (Raven, 1962), was administered with a 40-min time limit. The Raven is a nonverbal test designed to measure general intelligence.

RESULTS

Descriptive statistics for the major variable are shown in Table 2 (p. 315). The ASVAB means and standard deviations, despite some truncation at the lower tail, are fairly close to those found in unrestricted military applicant samples. The mean Raven score of 18.85 corresponds to an IQ of approximately 104, based on

TABLE 1
Tests[a] in ASVAB Forms 11, 12, and 13

Tests	Abbreviation	Description
General Science	GS	A 25-item test of knowledge of the physical (13 items) and biological (12 items) sciences—11 min
Arithmetic Reasoning	AR	A 30-item test of ability to solve arithmetic word problems—36 min
Word Knowledge	WK	A 35-item test of knowledge of vocabulary, using words embedded in sentences (11 items) and synonyms (24 items)—11 min
Paragraph Comprehension	PC	A 15-item test of reading comprehension—13 min
Numerical Operations	NO	A 50-item speeded test of ability to add, subtract, multiply, and divide one- and two-digit numbers—3 min
Coding Speed	CS	An 84-item speeded test of ability to recognize numbers associated with words from a table—7 min
Auto and Shop Information	AS	A 25-item test of knowledge of automobiles, shop practices, and use of tools—11 min
Mathematics Knowledge	MK	A 25-item test of knowledge of algebra, geometry, fractions, decimals, and exponents—24 min
Mechanical Comprehension	MC	A 25-item test of knowledge of mechanical and physical principles—19 min
Electronics Information	EI	A 20-item test of knowledge of electronics, radio, and electrical principles and information—9 min

[a]Reported as Navy Standard Scores, having a mean of 50 and a standard deviation of 10 for a representative sample of 19–23-year-old American youth.

conversion tables derived by Jensen, Saccuzzo, and Larson (1988). RT and *SD* are reported in seconds. For response accuracy, mean correct was 78.53 out of 82, or about 96%. While subjects were generally accurate, enough errors were committed to allow statistical analyses.

Development of Composites

The design of the study allows us to relate RT variables to three ability constructs: general intelligence, working memory, and clerical speed. The clerical speed dimension is included to establish divergent/convergent validity for the RT

TABLE 2
Descriptive Statistics

Variable	*M*	*SD*
GS	53.05	7.42
AR	51.91	7.51
WK	52.84	5.59
PC	53.32	6.58
NO	53.37	6.66
CS	52.50	6.61
AS	53.71	8.45
MK	52.10	7.27
MC	54.57	8.16
EI	52.38	8.37
Raven	18.85	5.50
Sequence	38.51	8.67
Counters	31.39	8.67
RT	0.574	0.091
SD	0.146	0.059
RT Correct	78.53	3.02

Note. The ASVAB tests are standardized with a mean of 50 and a standard deviation of 10.

measures, as described below. To develop composite measures of the three constructs, we first separated the eight ASVAB power tests from the two clerical speed tests (NO and CS in Table 1). The two clerical speed tests were then standardized and summed to form an equally weighted clerical speed composite. The 8 ASVAB power tests, which are primarily knowledge based, were standardized and averaged to form a "crystallized ability" index. To scale general intelligence, the crystallized ability score was averaged with the standardized Ravens Matrices score (a fluid intelligence test) to produce a composite index of "*Gf/Gc*." This derivation is based on the fluid/crystallized model of intelligence proposed by Cattell (1971). To form a working memory composite we averaged the standardized values for total correct from the Mental Counters and Sequential Memory tests, respectively.

The intercorrelations between the composites are shown in Table 3 (p. 316), which also presents correlations for the median and standard deviation of the choice RT test. More complete intercorrelations for the various tests are shown in the Appendix. Inspection of Table 3 indicates significant relationships between all measures. The .55 correlation between Gf/Gc (the psychometric intelligence index) and the working memory composite is particularly noteworthy. Theoretically, as argued elsewhere, it suggests that psychometric intelligence scores are, in part, related to a basic ability to orchestrate information "in the mind" (Larson & Saccuzzo, 1989). (See Woltz (1988) for further discussion of working memory.)

The significant correlation ($r = -.36$, $p < .01$, $N = 301$) between RT

TABLE 3
Pearson Correlation Coefficients

	Gf/Gc	Memory	Clerical	RT	*SD*
Gf/Gc	1.00				
Memory	0.55**	1.00			
Clerical	0.10*	0.27**	1.00		
RT	−0.28**	−0.31**	−0.19**	1.00	
SD	−0.36**	−0.36**	−0.17**	0.68**	1.00

Based on complete cases (n = 303)
*$p < .05$
**$p < .01$

standard deviations and intelligence scores (as well as the working memory composite) is central to our study. Later, the correlation will be untangled after the RT distributions have been appropriately partitioned.

Worst Performance Analysis

As noted earlier, high standard deviations commonly reflect a handful of inordinately slow responses; the frequency and magnitude of these responses is related to intelligence. Therefore, to understand why variability correlates with intelligence we sought to uncover the origins of the aberrant trials that skew the RT distributions. First, however, we wished to verify that there was, indeed, something of special interest in the slowest or "worst" part of the RT distribution. Our method was to rank-order each individual's 82 response times, from fastest to slowest. We then omitted the single fastest and single slowest RTs in each individual's ranked distribution so that the remaining RTs could be split into equal groupings. The remaining 80 responses were partitioned into 16 consecutive bands, with five RTs per band. The five RTs in each band were then averaged, producing 16 averages. These averages were of course ranked fastest to slowest, given their derivation. In the following analyses, these averages are labeled BAND1 (the average of the five fastest RTs) to BAND16 (the average of the five slowest RTs). BAND1 to BAND16 were then correlated (using Spearman's rho) with other measures, and the results are shown in Table 4. (The Spearman rank-order correlation coefficient was chosen to reduce artifactual effects from differences in within-band variability.) The table can be read as follows: Across all subjects, the rank-order correlation of intelligence with the average of the five fastest RTs (i.e., "BAND1") is −.20. The rank-order correlation of intelligence with the average of the five slowest RTs ("BAND16") is −.37, and so forth. Correlations for intermediate bands are interspersed.

The table reveals a striking trend: The slower end of the distribution is where the better predictions of mental ability are found. Overall, the position of RTs on a fastest-to-slowest RT ordering is almost perfectly correlated with their ability to predict both psychometric intelligence and working memory scores (Spearman's rho is −1.0 and −.99, respectively [X, Y entries were band numbers and column

TABLE 4
Descriptive Statistics and Validities for Fastest (Band1) to Slowest (Band16) *RTs*

		Spearman Correlations with Band Measures				
Band	***M/SD***	***Gf/Gc***	***Memory***	***Clerical***	**RT**	***SD***
B1	406/055	−.20**	−.21**	−.10*	.80**	.33**
B2	444/061	−.23**	−.25**	−.13*	.88**	.45**
B3	467/067	−.23**	−.27**	−.15**	.92**	.50**
B4	491/071	−.24**	−.28**	−.16**	.95**	.55**
B5	510/075	−.24**	−.28**	−.17**	.97**	.59**
B6	528/080	−.25**	−.29**	−.19**	.98**	.62**
B7	546/085	−.26**	−.30**	−.18**	.99**	.65**
B8	564/089	−.27**	−.31**	−.18**	.99**	.67**
B9	583/092	−.28**	−.31**	−.18**	1.0**	.70**
B10	603/097	−.29**	−.32**	−.20**	.99**	.73**
B11	624/102	−.29**	−.32**	−.20**	.98**	.76**
B12	648/109	−.30**	−.32**	−.20**	.97**	.79**
B13	679/120	−.31**	−.33**	−.19**	.95**	.83**
B14	718/139	−.33**	−.35**	−.20**	.93**	.85**
B15	776/167	−.34**	−.36**	−.20**	.90**	.89**
B16	903/232	−.37**	−.36**	−.17**	.82**	.95**
		−1.0††	−.99††	−.75††	.06	1.0††

Note: The numbers in the bottom row are correlations between the Band numbers and the values in the remaining columns. Also, the reader should observe that "RT" refers to each subject's median RT, which by definition falls in the middle Band of the RT distribution.

*$p < .05$, n = 303
**$p < .01$, n = 303
††$p < .01$, n = 16

correlations, $N = 16$]). Since working memory is a construct theoretically related to intelligence (Jensen, 1982; Larson & Saccuzzo, 1989; Woltz, 1988), the parallel findings are a form of convergent validity. The results are illustrated graphically in Figure 2 (p. 318). The figure also shows that the trend is not nearly so pronounced for the clerical speed composite. Results for clerical speed are important insofar as they suggest that the Worst Performance rule applies specifically to "intellectual" tests, rather than originating in some statistical artifact (such as greater intersubject standard deviations in the higher bands) that inflates correlations in an across-the-board manner.

The importance of band position is further illustrated in Figure 3 (p. 318). We selected, from the total sample of subjects, individuals who fell in the upper and lower quartiles of ability based on the Gf/Gc general intelligence composite. We then plotted, for each ability group, the average RT for each consecutive (fastest-to-slowest) band. As can be seen, there is little difference between ability groups on their fastest responses, but the groups increasingly diverge towards the slow end of their distributions. The slowest band manifests the greatest group difference.

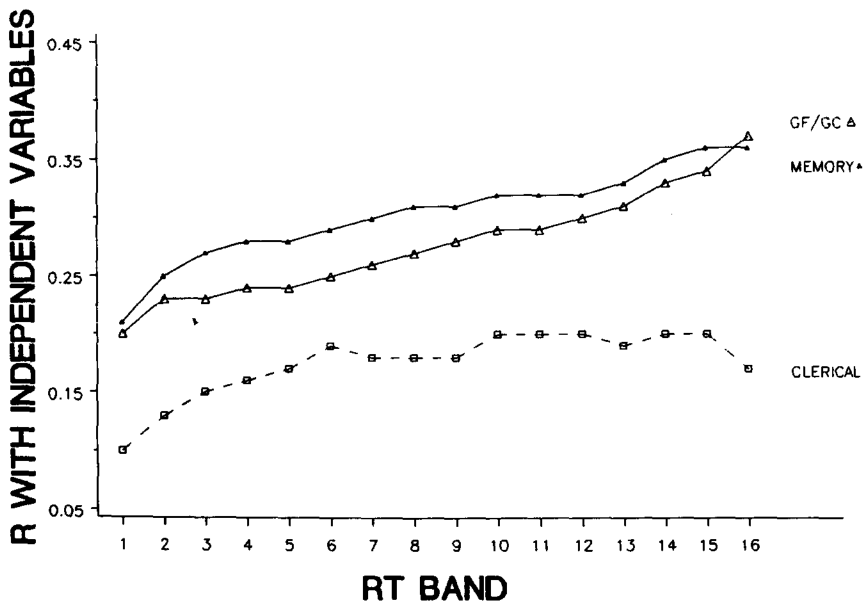

FIG. 2. Correlations of fast-to-slow RT bands with Gf/Gc, working memory, and clerical aptitude.

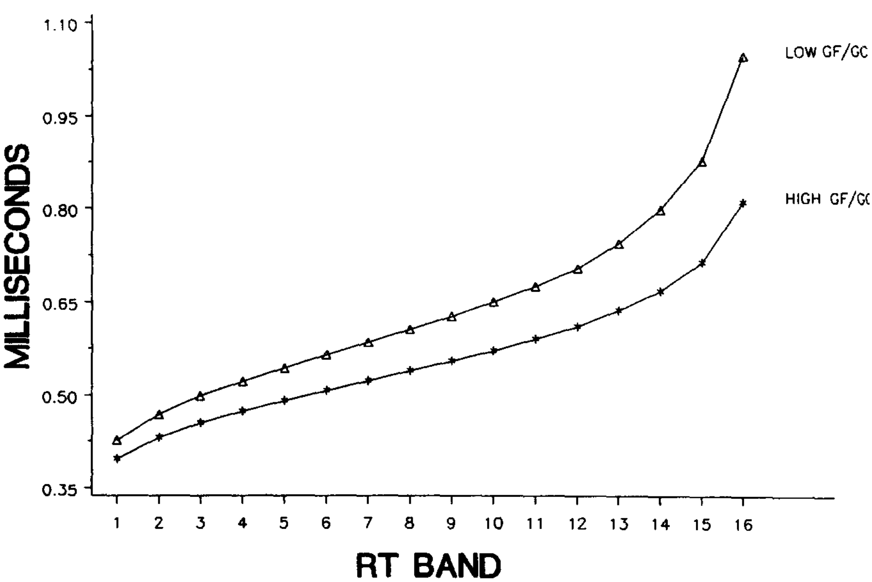

FIG. 3. Means of fast-to-slow RT bands for high and low Gf/Gc groups.

Worst Performance and RT Variability

Returning to Table 4, the far right column indicates a pronounced association between RT band position and the correlation with the overall RT standard deviation. Clearly, the magnitude of final bands, representing each subjects' longest response times, is practically redundant with the standard deviations. Considering the nature of these variables, one might argue that correlation implies (statistical) causation, that is, intellectually important differences in variability originate in the magnitude of each subjects' slowest RTs. Thus the −.36 correlation between variability and psychometric intelligence drops to −.04 (ns) when Band16 is partialled out. Returning to the purpose of our study, the results suggest that we can understand or model the global variability score if we can determine the origins of the longest RTs.

A Note On Confounds

Before presenting the final analyses, several possible sources of confound will be addressed. First is the mistaken criticism that our emphasis on "worst trials" simply exploits the disproportionate effect of outliers on correlations. As noted throughout the paper, our analysis involves *within-subject* partitioning of RT distributions, so that *every* individual achieves a score for *every* part of his RT distribution. The result is a strong linear relationship between the position on these within-subject continuums and the prediction of intelligence. Our finding that certain *aspects of each subjects performance* drive the correlations is entirely different from instances where certain *extreme cases* account for the statistical relationships. Moreover, the fact that the relationships in Table 4 were systematic and linear suggests that nearly the same results would be found if we threw out Band16, or even Band 15 and Band16. The data are lawful; outliers are not.

A second criticism is that intersubject variability increases steadily from Band1 to Band16, possibly creating the parallel increase in external correlations with intelligence and other variables. This, of course, was the reason we report Spearman's rho, a statistic less dependent on the magnitude of sample variances than is the Pearson product–moment correlation coefficient. More important, however, is that the increase in intersubject variability is not a confound, but an answer. If the question is, "On what aspect of the data should we focus our study of individual differences?", the answer is, "Obviously, to the place where those differences emerge." The fact that there is more variance around the means of the final RT bands justifies the investigation of the source of that variance.

To summarize the preceeding analyses, we began by viewing variability as a global score that could be decomposed into a distribution of consecutive fast-to-slow RT bands. Analysis of the bands reveals that the greatest deficits of lower aptitude subjects occur in the slowest bands, and that these slow bands are of fundamental importance to individual differences in variability. Thus, hypotheses about variability can be framed as hypotheses about the origins of the worst trials. In the next section, we investigate whether slow RTs follow response errors.

Are Worst Trials a By-Product of Errors?

Various studies show that posterror reaction times are abnormally slow (Brewer & Smith, 1984; Kiger & Glass, 1981; Kleiter & Schwarzenbacher, cited in Crano, 1988; Rabbitt & Rodgers, 1977). The pattern is exaggerated in retardates, who also make more errors overall (Brewer & Smith, 1984). Frequent and extreme posterror slowing, by itself, could substantially explain why retarded subjects have larger RT standard deviations than do normals (see Baumeister & Kellas 1968; Jensen, 1982). If the phenomena were pervasive, posterror slowing might also explain the variability/intelligence correlations found with normal adults.

Before presenting the analysis, however, current views on the nature of posterror slowing should be noted. Apparently, subjects continue to process data even after they have made a response, if they are uncertain of the response's accuracy. According to Welford (1980), such continued processing can occupy the cognitive machinery to the detriment of new information, thus slowing a posterror response via interference. An interference hypothesis is also presented by Rabbitt and Rodgers (1977), who note that if a response error has been committed, a following response that would have been correct on the preceeding trial will be relatively fast, as if orders for that response were still operating. The "excited" response unit from the previous trial is said to interfere with other responses. Finally, Brewer and Smith (1984) suggest that posterror slowing is due to criterion shifts by which subjects attempt to maintain an appropriate level of speed and accuracy. We wish to note, however, that our study was not designed to contrast these hypotheses.

We first selected only those subjects who had committed at least one error ($N = 273$), then we calculated three means per subject: (a) their average RT for trials preceeding an error, (b) the average RT for their error trials, and (c) their average posterror RT. Table 5 shows the distribution of each subject's three averages as a function of their *own* fast-to-slow RT bands. Next, the percentage of subjects whose pre-error, error, and posterror RTs fell into each of the 16 RT categories was computed. These data are presented in Table 5. For example, 0.4% of all subjects had their pre-error RT mean fall in their fastest RT band (BAND1) while 3.7% of all subjects had their mean pre-error RT fall in their slowest RT category (Band16). These obtained frequency distributions (Table 5) were tested against a null distribution of equal band frequencies. The chi-square tests for these data rejected the null (equal frequency) hypothesis for pre- (chi-square [15] = 80.965, $p < .001$) and posterror (chi-square [15] = 137.697, $p < .001$) times only. The entries in the table suggest that pre-error RTs are bunched in the middle of the distribution, while posterror RTs are skewed toward the slower bands. Thus, posterror trials are, indeed, associated with slower responding. From this, one would expect that subjects with more errors, and thus more frequent posterror slowing, would have greater variability. This was indeed the case; a significant relationship between number of errors and variability ($r = .26$, $p < .001$) emerged.

TABLE 5
Distribution of Pre-error, Error, and Post-error RTs ($n = 273$) and Chi-square Tests Against a Uniform Distribution

Band	Percentage of Trials Falling in Bands		
	Pre-error	Error	Posterror
B1	0.4	3.3	0.4
B2	0.7	5.9	0.0
B3	3.3	6.6	0.7
B4	7.7	6.2	1.5
B5	6.2	6.6	1.1
B6	7.3	5.1	2.2
B7	7.7	7.0	3.7
B8	9.5	9.5	7.7
B9	10.6	7.7	5.9
B10	13.6	5.5	9.5
B11	6.2	5.1	10.6
B12	8.8	4.0	12.8
B13	4.4	6.2	11.4
B14	5.1	8.8	9.9
B15	4.0	5.5	7.7
B16	3.7	5.9	11.0
Chi-Square	80.965*	16.494**	137.679*

*$p < .001$
**$p = .350$

Have we thus explained the slow bands (and high variability) of lower intelligence subjects? Unfortunately, the answer appears to be no. Consider the following.

1. First, while the number of errors was correlated .26 with variability, it was unrelated to intelligence ($r = .08$, NS). Therefore, low aptitude subjects do not exhibit higher variability because they have more instances of posterror slowing per se. It is nevertheless possible that while low aptitude subjects do not make more errors, the deficit following *each* error could be greater. An examination of our data, however, revealed little support for this hypothesis. The intelligence score was only correlated $-.17$ ($p < .01$) with the relative magnitude of the posterror slowing. In other words, the posterror slowing of low intelligence subjects (relative to their own mean performance) was only slightly greater than posterror slowing of brighter subjects. Finally, we noted earlier that the correlation of psychometric intelligence with RT variability was $-.36$ ($p < .001$). If we now partial out response accuracy, the correlation only drops to $-.35$. Partialling out the relative magnitude of posterror slowing produced exactly the same result. These analyses suggest that error-

related variance *on elementary RT tasks* is not intellectually important. (Obviously this should not be generalized to more complicated tasks.)
2. On a percentage basis, posterror trials comprise only about 6% of the Band15 and Band16 RTs. This is far too meager to explain the entire set of findings for these bands.

DISCUSSION

The present study was undertaken to clarify the widely reported relationships between RT variability and intelligence. The important aspects of the study were the method and the hypothesis. The method was to treat RT variability as a global index that should be decomposed and studied in terms of the distribution of its constituent RTs. This allows the isolation of the subset of RTs that, for example, inflate the variabilities of low aptitude subjects. Once the key RTs have been located in the distribution, hypotheses about their origins (and about the nature of individual differences in variability) can be tested. An analysis of our subjects' RT distributions suggested (as expected) that the worst trials, or slowest bands, were of great importance to variability and to the variability/intelligence relationship. There was, in fact, a near perfect correlation between the position of RTs in a fast-to-slow ranking and the correlation of the RT with psychometric intelligence. The same pattern emerged when working memory tests were used as criteria. The pattern was not nearly so strong, however, when the criterion was a clerical speed index. The latter finding demonstrates the construct and divergent validity of the "worst performance" method.

We proceeded to evaluate an error-based production model for the worst-trial deficit of lower mental ability subjects. When the aftereffects of errors were examined, we found, in agreement with the literature, posterror slowing of RTs. Posterror trials were primarily distributed in the slower RT bands, and the total number of errors was significantly related to RT variability ($r = .26, p < .001$). However, there were no particularly impressive relationships between error data and mental ability. Thus, posterror slowing contributes to variability, but that relationship may be orthogonal to the relationship between variability and intelligence. The latter relationship, therefore, remains essentially unexplained.

Though the *hypothesis* was not supported, we are nevertheless convinced that the *method* is correct. That is, the pronounced skew observed in the RT distributions of lower mental ability subjects will probably only be explained when global variability scores are abandoned in favor of some trial-by-trial form of analysis and hypothesis testing. The question of "What happens on the *worst* trials?" is central to this line of research. Further hypotheses about those trials are necessary.

Presumably, the variability/intelligence relationship could emerge either because variability reflects a genuine cognitive deficit, or because of some stillhidden artifact in RT measurement. The deficit view has been proposed by Jensen (1982), who suggests that variability is a product of the oscillatory nature of the central nervous system. Larson and Saccuzzo (1989) take a position potentially compatible with Jensen's; they argue that variability implies lapses in the chaining of working memory processes. That is, during problem solving, the representation and transformation of problem elements "in the mind" is intermittent; this impairs efficiency, and limits mental capacity. Since mental objects/processes stem from pooled neural functioning, Larson and Saccuzzo conclude by arguing for a neural assembly view of the variability/intelligence relationship. As they note, however, the view is speculative.

Finally, a comment on RT methodology is warranted. It has become standard procedure to delete error trials from RT data and assume that the remaining trials are unblemished. Our data (and other literature) suggest that this assumption is wrong, since error trials are far more representative of all trials than are posterror trials. Investigators who want "clean data" should carefully examine their rationale as well as the posterror latencies.

REFERENCES

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

Baumeister, A.A., & Kellas, G. (1986). Distributions of reaction times of retardates and normals. *American Journal of Mental Deficiency, 72,* 715–718.

Brewer, N., & Smith, G.A. (1984). How normal and retarded individuals monitor and regulate speed and accuracy of responding in serial choice tasks. *Journal of Experimental Psychology: General, 113,* 71–93.

Carlson, J.S., & Jensen, C.M. (1982). Reaction time, movement time, and intelligence: A replication and extension. *Intelligence, 6,* 265–274.

Carlson, J.S., Jensen, C.M., & Widaman, K.F. (1983). Reaction time, intelligence, and attention. *Intelligence, 7,* 329–344.

Cattell, R.B. (1971). *Abilities: Their structure, growth, and action.* New York: Houghton Mifflin.

Crano, W. (1988). Austrian research on the cognitive and psychophysiological consequences of errors. *European Science Notes Information Bulletin (88-06).* London: Office of Naval Research Branch Office.

[illegible]en, A.R. (1982). Reaction time and psychometric "g". In H.J. Eysenck (Ed.), *A model for intelligence.* New York: Springer-Verlag.

[illegible]en, A.R. (1987). Process differences and individual differences in some cognitive tasks. *Intelligence, 11,* 107–136.

[illegible]en, A.R., Saccuzzo, D.P., & Larson, G.E. (1988). Equating the standard and advanced forms of the Raven Progressive Matrices. *Educational and Psychological Measurement, 48,* 1091–1095.

[illegible]r, J.I., & Glass, A.I. (1981). Context effects in sentence verification. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 688–700.

[illegible]on, G.E., Merritt, C.R., & Williams, S.E. (1988). Information processing and intelligence: Some implications of task complexity. *Intelligence, 12,* 131–147.

[illegible]n, G.E., & Saccuzzo, D.P. (1989). Cognitive correlates of general intelligence: Toward a process theory of *g*. *Intelligence, 13,* 5–31.

Nettlebeck, T., & Kirby, N.H. (1983). Measures of timed performance and intelligence. *Intelligence, 7*, 39–52.

Rabbitt, P. & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *Quarterly Journal of Experimental Psychology, 29*, 727–743.

Raven, J.C. (1962). *Advanced Progressive Matrices. Set II*. London: H.K. Lewis.

Smith, G.A., & Stanley, G. (1983). Clocking *g*: Relating intelligence and measures of timed performance. *Intelligence, 7*, 353–368.

Vernon, P.A. (1983). Speed of information processing and general intelligence. *Intelligence, 7*, 53–70.

Vernon, P.A., Nador, S., & Kantor, L. (1985). Reaction time and speed-of-processing: Their relationship to timed and untimed measures of intelligence. *Intelligence, 9*, 357–374.

Welford, A.T. (1980). The single-channel hypothesis. In A.T. Welford (Ed.), *Reaction times*. New York: Academic.

Woltz, D.J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General, 117*, 319–331.

Appendix: Complete Intercorrelations of Criterion Measures

	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	Raven	Sequenc	Counters	*SD*	RT
AR	.55**	1.0													
WK	.74**	.46**	1.0												
PC	.46**	.47**	.53**	1.0											
NO	−.04	.22**	−.10*	.05	1.0										
CS	−.00	.21**	−.01	.20**	.52**	1.0									
AS	.49**	.42**	.44**	.30**	−.08	.07	1.0								
MK	.54**	.68**	.41**	.41**	.28**	.23**	.35**	1.0							
MC	.62**	.56**	.56**	.39**	−.03	.10*	.63**	.53**	1.0						
EI	.61**	.46**	.53**	.39**	−.09	.06	.59**	.44**	.66**	1.0					
RAVEN	.47**	.56**	.35**	.25**	.03	.11*	.33**	.58**	.53**	.39**	1.0				
SEQUENC	.29**	.52**	.26**	.27**	.16*	.20**	.16**	.45**	.32**	.24**	.51**	1.0			
COUNTERS	.24**	.45**	.17**	.19**	.19**	.24**	.24**	.43**	.27**	.24**	.39**	.48**	1.0		
SD	−.25**	−.26**	−.18**	−.27**	−.04	−.25**	−.26**	−.22**	−.35**	−.23**	−.30**	−.33**	−.29**	1.0	
RT	−.21**	−.22**	−.04	−.13**	−.11*	−.22**	−.22**	−.23**	−.29**	−.17**	−.25**	−.27**	−.27**	.68**	1.0

* $-p < .05$
** $-p < .01$