DOCUMENT RESUME

ED 115 686 95 TM 004 959

AUTHOR Lunneborg, Clifford E.
TITLE Choice Reaction Time and Pyshcometric Performance Revisited. Report No. 76-7.
INSTITUTION Washington Univ., Seattle. Educational Assessment Center.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.; National Inst. of Mental Health (DHEW), Rockville, Md.
REPORT NO 76-7; EAC-P-262
PUB DATE Oct 75
NOTE 36p.

EDRS PRICE MF-$0.76 HC-$1.95 Plus Postage
DESCRIPTORS Aptitude Tests; Cognitive Ability; *Cognitive Measurement; *Cognitive Processes; Comparative Analysis; *Individual Differences; Memory; *Psychometrics; *Reaction Time; Scores; Statistical Analysis

ABSTRACT

As part of a long term project relating individual differences in performance on paper and pencil tests of intellectual aptitude and achievement to differences observed in laboratory studies of cognition, data have been collected comparing test scores with simple choice reaction times. The three studies summarized here suggest that the relationship may be fairly strong except where the sampling of subjects to be studied seriously restricts variability in those measures. (Author/RC)

Educational Assessment Center

University of Washington

October 1975

76-7

Choice Reaction Time and Psychometric Performance Revisited

Clifford E. Lunneborg

As part of a long term project relating individual differences in performance on paper and pencil tests of intellectual aptitude and achievement to differences observed in laboratory studies of cognition, data have been collected comparing test scores with simple choice reaction times. The three studies summarized here suggest that the relationship may be fairly strong except where the sampling of subjects to be studied seriously restricts variability in those measures. Portions of these results were communicated in papers read by the author at the 1974 and 1975 conventions of the American Psychological Association. The research was supported by grants from the National Institute of Education to the author and from the National Institute of Mental Health to E. B. Hunt, Department of Psychology, University of Washington.

Educational Assessment Center Project: 262

Choice Reaction Time and Psychometric Performance Revisited

Clifford E. Lunneborg

Both experimental cognitive psychology and psychometrics are concerned with the development of structures which explain human information processing. To the extent that they share this area of mutual concern it ought to be possible to establish some correspondence between measures of individual performance along dimensions implicit in modern theories of human cognition, particularly theories of memorial organization and operation, and measures of intelligence or cognitive abilities. If such relations can be established they would be of interest both to the cognitive theorist and to the psychometrist. For cognitive psychology the message would be that individual differences in intelligence are well-established and need to be mirrored by individual differences in parameters postulated for any cognitive theory. For the psychometrist the expectation is that assessment of purely normative individual differences in global performance can be replaced by individual differences which can be expressed in terms of a processing metric (time, capacity, rate) required by a theory of information processing. In short, research in this area is directed at providing either a theory of cognition that takes individual differences into account or a means of assessing individual differences within some theoretical basis.

One theoretical basis for such bridge-building work is provided by the distributed memory model (Hunt 1971, 1973). Briefly characterized, the distributed memory model is concerned with the buffering of sensory information into central memory, recoding of information within and between memory stages, retrieval of information from memory, and the cross-referencing and comparison of information stored in memory. Human information processing is likened to a computing system with control processes manipulating information stored in a data structure subject to the constraints of the brain's system architecture. Between individuals, differences in cognitive performance can be contributed to by differential characteristics of the physical system, by differences in availability or efficiency of control processes, or by differences in the organization of the data structure. How might differences of these kinds relate to tested differences, for example, in verbal intelligence or spatial ability?

Differences in the speed with which individuals accomplish basic information processing tasks are a prime candidate for study and data from three studies are reported here which bear on the question of how such differences, measured in the laboratory, relate to performance on paper and pencil psychometric instruments, standard aptitude tests.

Study 1

Carroll (1974) worked with a representative sample of psychometric tests, those assembled in the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, and Price, 1963), examining each with respect to the same distributed memory, information processing model mentioned earlier. This resulted in the characterization of each test with respect to the demands that test makes on an information processing system. More specifically, those information processing characteristics were identified which would be expected to show appreciable individual differences.

Carroll's theoretical analysis suggested a framework within which to look at relations between laboratory measures of information processing and test performance.

Method

Subjects: Sixty-four high school age subjects from the Seattle area were recruited summer 1973 via a community Job Line. In groups of eight they spent half a day on each of two consecutive days in the laboratory. During each session they participated in several short laboratory assessments and completed a battery of short paper and pencil tests. Subjects were paid on an hourly basis for their participation.

Psychometric Instruments: The tests may be briefly described.

1. Verbal Comprehension, EAS test 1, calls for the subject to select a synonym from among five alternatives for each of 30 words. Five minutes are allowed to work on the task. As with the other multiple choice tests in this battery the score is the number of items answered correctly minus a fraction of those answered incorrectly. In the Reference Kit scheme this test measures Factor V, Verbal Comprehension.

2. Numerical Ability, EAS test 2. Part 1 contains 25 items of the form "16 + 18 = _____ " with the answers to be selected from five alternatives. Two minutes are allowed on this part. Part 2 consists of 25 problems of decimal fractions, for example, "1.6 x 0.3 = _____ ". Four minutes were allowed for this part. The two parts provided a single score. This test measures Factor N, Number Facility.

3. Space Visualization, EAS test 5, presents the subject with a series of ten line drawings. Each drawing depicts a stack of blocks and the subject's task is to report, by marking on an answer sheet, for each of the labelled blocks the number of other blocks in the stack that are in contact with it. Five minutes are allowed and the score is the number of correct counts completed. The test is considered a measure of Factor Vz, Visualization.

4. Numerical Reasoning, EAS test 6, presents subjects with twenty number series. The task is to select the next, eighth element from among five alternatives.

This test taps Factor I, Induction, as well as Number Facility.

5. Verbal Reasoning, EAS test 7, is a deductive reasoning test and a measure of cognitive factor Rs, Syllogistic Reasoning. Subjects are presented with a set of four facts followed by five conclusions and asked to report for each conclusion whether it is true, false, or of indeterminate validity. Six such exercises may be attempted in the five minutes testing time.

6 and 7. The Minnesota Clerical test consists of two parts each yielding a separate score. In part 1 subjects compare pairs of numbers (from 3 to 12 digits in length) and check the pairs which are the same (3 minutes). The second part of the test is identical except that the pairs consist of names of individuals or businesses, and again, this part runs for 3 minutes. These Clerical Number and Clerical Name scores should both tap Factor P, Perceptual Speed.

8. Hidden Figures is an experimental test included in the Reference Kit as a measure of Factor Cf, Flexibility of Closure. Subjects are presented with five polygons and asked to determine for each of sixteen complex line drawings which polygon is hidden in the complex figure.

Laboratory Measures: The laboratory sessions involved subjects in several settings. Only those which provided measures for the present analyses, however, will be described here.

1. Motor Reaction Time was the median time taken to respond by key press to the onset of a + on a CRT screen.

2. A one-handed choice reaction time study provided a median time to respond discriminatively by depressing a left or right key to two stimuli, a circle appearing either in the left or right half of a CRT display. This choice reaction time was not itself employed in the present analysis. Rather a variable called Choice Time was defined as the difference between the median choice and motor reaction times.

3. The choice reaction time study above also provided a second variable for this analysis: proportion of trials on which the correct choice was made.

4. In the Stroop task subjects were asked first to report orally the colors in which a series of asterisks were printed and then report the color of printing of an equivalent length list of contrasting color names. The task was repeated twice and the score used here was the average difference in "reading" times between the name and asterisk conditions

5 and 6. In an experiment in the Sternberg (1970) paradigm Ss were sequentially shown from one to six consonants and then shown a single probe consonant and asked whether the probe was in the previously exposed set. Response time was recorded and, for correctly identified instances, these times linearly regressed on the number of digits in the associated target set. For each S, then, a slope and intercept value were obtained.

7 and 8. Strings of fifteen digits were presented binaurally to Ss with recall cued immediately following presentation. In recall Ss were instructed to recall in order as many digits as possible beginning with the first digit heard. Ten trials were presented with each scored for the number of digits recalled in order. In this analysis two scores were employed--a Digit Span, Final score (average of performances on the last five trials) and Digit Span, Gain score (difference between the Final average and the average over the first five trials).

9 and 10. In a task patterned after Massaro (1972) four digits and four consonants were presented dichotically, two digits and two consonants to each ear. Following presentation Ss were cued to recall the presented material. On some blocks of trials Ss were to report by ear, i.e., report what they heard to the right ear or the left ear. On other blocks they were asked to report by category, either digits or consonants. While they knew whether they were going to have to recall by category or by ear, which category or which ear was not cued until after the presentation. Two scores were extracted for the present analysis. The Category Score gives the number of items correctly reported over all 40 category trials. An Ear Minus Category Score was obtained by subtracting the category score from a similar score obtained from the ear trials.

11 - 14. A final set of four scores was obtained from a study of clustering in recall of a list of nouns based on the experimental paradigm of Puff (1966). During their first laboratory session Ss were shown item by item, two lists of 30 common nouns. Each list was shown twice with subjects asked for recall immediately following each presentation (four recalls). Each list consisted of ten nouns each from three semantic categories (fruits, occupations, animals, etc.). For half the subjects the list presented first was blocked, all ten members of a given category appearing contiguously, and the list presented second was in pseudo-random order. For the other half of the Ss this order was reversed. When Ss returned for their second day in the laboratory they were asked to recall each of the two lists (two more recalls). A clustering score was computed for each of these six recalled lists. In the present analysis four transformed s res were employed: (11) The clustering immediately following the second presentation of the blocked list provided a base score: B1; (12) this base clustering minus the clustering for second presentation of pseudo-random list: B1-R1; and (13) base clustering minus the clustering for second day recall of blocked list: B1-B2; and (14) this second day blocked clustering score minus the second day random clustering: B2-R2.

Results

Correlations between the psychometric and laboratory data are shown in Table 1. The response time measures--Motor R T, Choice Time, the Stroop measure, and the two scores from the Sternberg task--correlated negatively with paper and pencil test performance, as might be expected. Proportion of correct responses in the choice reaction time task correlated only weakly, though positively, with the psychometric measures. This laboratory measure had little variability, however, as few errors were made (average proportion correct was .93). The negative correlations involving Digit Span, Gain suggest the possibility that, as improvement on the average was small (from 5.75 digits to 6.23 digits), big gains were registered by those who had not done well in the initial trials. The digit span score itself was not highly correlated with any of the psychometric measures.

The Category Scores from the dichotic listening task correlated positively with nearly all of the paper and pencil tests. Raw ear scores were considerably smaller than category scores (roughly 70 as opposed to 90 on the average). The pattern of correlation for the Ear Minus Category scores suggests that highest psychometric scores were earned by Ss whose category performance far outstripped their ear performance. Finally, the Clustering Score is such that low scores (actually large negative scores) indicate greater clustered recall. Except for the reversal of signs the pattern of correlations for the clustering on immediate recall of the blocked list (B1) is quite similar to that for the Category Score from the dichotic listening task. The difference in clustering scores between the immediate recall of blocked and randomized lists (B1-R1) was not well correlated with any of the psychometric measures. (Clustering was, of course, greater for the blocked than for the random list.) Clustering for the blocked list was almost exactly the same for delayed as for immediate recall (the score B1-B2 had a mean close to zero). This difference, however, did correlate with a number of psychometric measures. Increased reliance on semantic clustering between immediate and delayed recall tended to be negatively related to certain paper and pencil test performances. The last measure included

Table 1

Correlations Between Psychometric and Laboratory Measures, Study 1

(Decimal points omitted)

	Verbal Comprehension	Numerical Ability	Space Visualiza	Numerical Reasoning	Verbal Reasoning	Clerical, Number	Clerical, Name	Hidden Figures
Motor R T	-40	-38	-42	-40	-27	-17	-38	-27
Choice Time	-40	-45	-55	-50	-30	-28	-37	-40
Pr Corr CRT	21	11	12	04	14	10	14	17
Stroop (Wd-*)	04	-27	-19	-32	-16	-35	-38	-09
Sternberg Slope	-07	-10	-27	-16	-07	-24	-26	03
Sternberg Intercept	-28	-36	-49	-31	-16	-29	-39	-27
Digit Span, Final	-04	02	-03	02	-12	22	16	05
Digit Span, Gain	-10	-26	-24	-21	-29	-06	-17	-07
Category Score, D L	55	46	49	52	32	02	26	48
Ear-Category, D L	-21	-29	-23	-31	-18	-16	-29	-25
Clustering, B1	-52	-48	-48	-61	-47	-02	-37	-34
B1-R1	06	10	10	-07	07	00	09	-14
B1-B2	-12	-22	-21	-28	-25	01	-03	-35
B2-R2	-18	00	06	04	05	06	00	20

in Table 1, B2-R2, the difference in clustering of delayed recall between the blocked and randomized lists, had near zero correlations with psychometric measures. (For the randomized list, however, clustering was considerably greater on the average for delayed than for immediate recall.)

Table 1 reports zero order correlations between laboratory and psychometric measures. Because many laboratory measures are themselves not pure measures of information processing parameters--for example, clustering score B1 may be interpreted as a measure of short-term serial recall ability as well as any semantic clustering tendency--these correlations are more equivocal than desired. The laboratory performances were correlated among themselves. This is reflected in Table 2. The quickness of response measures--1, 2, 4, 5, and 6--tended to be positively correlated. Clustering on immediate recall of a blocked list had a sizeable negative correlation (-.55) with the category score in the dichotic listening task. Tables 1 and 2 provide the essential data for the generation of the final results to be reported.

The systematic stepwise multiple regression analyses completed provide a quantitative analog to Carroll's (1974) point about the information processing or memorial complexity of most paper and pencil tests. The strategy was to order the laboratory measures by level of complexity ranging from the motor reaction time measure, which places the least load on memorial operations, to recall measures in the dichotic listening and clustering studies. The multiple regression analyses assessed how much of the inter-individual variability in performance on each of the paper and pencil tests could be accounted for successively by the independent contributions of the several laboratory measures, working up from the least complex. The results are given in Table 3. Entries in the columns headed $\underline{R}^2$ are cumulative proportions of variance accounted for. Motor Reaction Time accounted for 16% of the variance in the Verbal Comprehension scores; Motor Reaction Time plus Choice Time accounted for 22% of the variance, etc. The Δ entries are the increments, the additional proportion of variance accounted for. It is important to keep in mind that these increments are for partial variables. The contribution of the Choice Time variable, for example, is the contribution of that measure after it has had partialled from it the influence of Motor Reaction Time.

Table 2

Intercorrelations among Laboratory Measures, Study 1

(Decimal points omitted)

	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Motor R T	47	-07	21	42	69	00	15	-52	38	36	-28	09	-19
2. Choice Time		14	27	43	55	-21	-11	-40	30	30	03	04	07
3. Pr Corr CRT			14	25	12	-29	-41	08	06	-20	17	-19	16
4. Stroop (Wd-*)				27	21	-32	16	-11	22	14	-02	-07	-15
5. Sternberg Slope					39	-17	-11	-21	17	15	-44	-21	05
6. Sternberg Intercept						-21	-07	-45	44	21	-13	-02	-06
7. Digit Span, Final							65	-04	-02	14	-19	24	-28
8. Digit Span, Gain								-12	02	34	-17	32	-25
9. Category Score, D L									-75	-55	04	-35	02
10. Ear-Category										32	08	26	02
11. Clustering B1											16	54	09
12. B1-R1												28	32
13. B1-B2													-32
14. B2-R2													

Table 3

Predicting Psychometric Performance from Laboratory Measures:

Increments to the Squared Multiple Correlation

(Decimal points omitted)

	Verbal Comprehension		Numerical Ability		Space Visualiza		Numerical Reasoning		Verbal Reasoning		Clerical, Number		Clerical, Name		Hidden Figures		
	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	R^2	Δ	
Motor R T	16	16	15	15	18	18	16	16	07	07	03	03	15	15	08	08	A
Choice Time	22	06	25	10	34	16	28	12	11	04	08	05	20	05	17	09	
Stroop (Wd-*)	25	03	27	02	34	00	31	03	12	01	16	08	27	07	17	00	
Sternberg Intercept	26	01	27	00	36	02	32	01	13	01	19	03	28	01	17	00	
Sternberg Slope	29	03	30	03	36	00	34	02	14	01	20	01	28	00	24	07	B
Pr Corr CRT	32	03	32	02	40	04	34	00	16	02	25	05	33	05	27	03	C
Digit Span, Final	32	00	32	00	42	02	35	01	18	02	26	01	33	00	27	00	
Digit Span, Gain	32	00	38	06	46	04	37	02	21	03	27	01	35	02	27	00	D
Category Score, D L	43	11	43	05	49	03	45	08	23	02	29	02	35	00	37	10	
Ear-Category, D L	49	06	44	01	54	05	47	02	23	00	32	03	37	02	39	02	E
Clustering, B1	53	04	46	02	55	01	55	08	29	06	32	00	41	04	39	00	F
Clustering, B1-R1	55	02	50	04	55	00	55	00	31	02	33	01	42	01	40	01	
B1-B2	59	04	50	00	55	00	55	00	31	00	34	01	43	01	43	03	
B2-R2	60	01	50	00	55	00	56	01	31	00	34	00	43	00	47	04	G

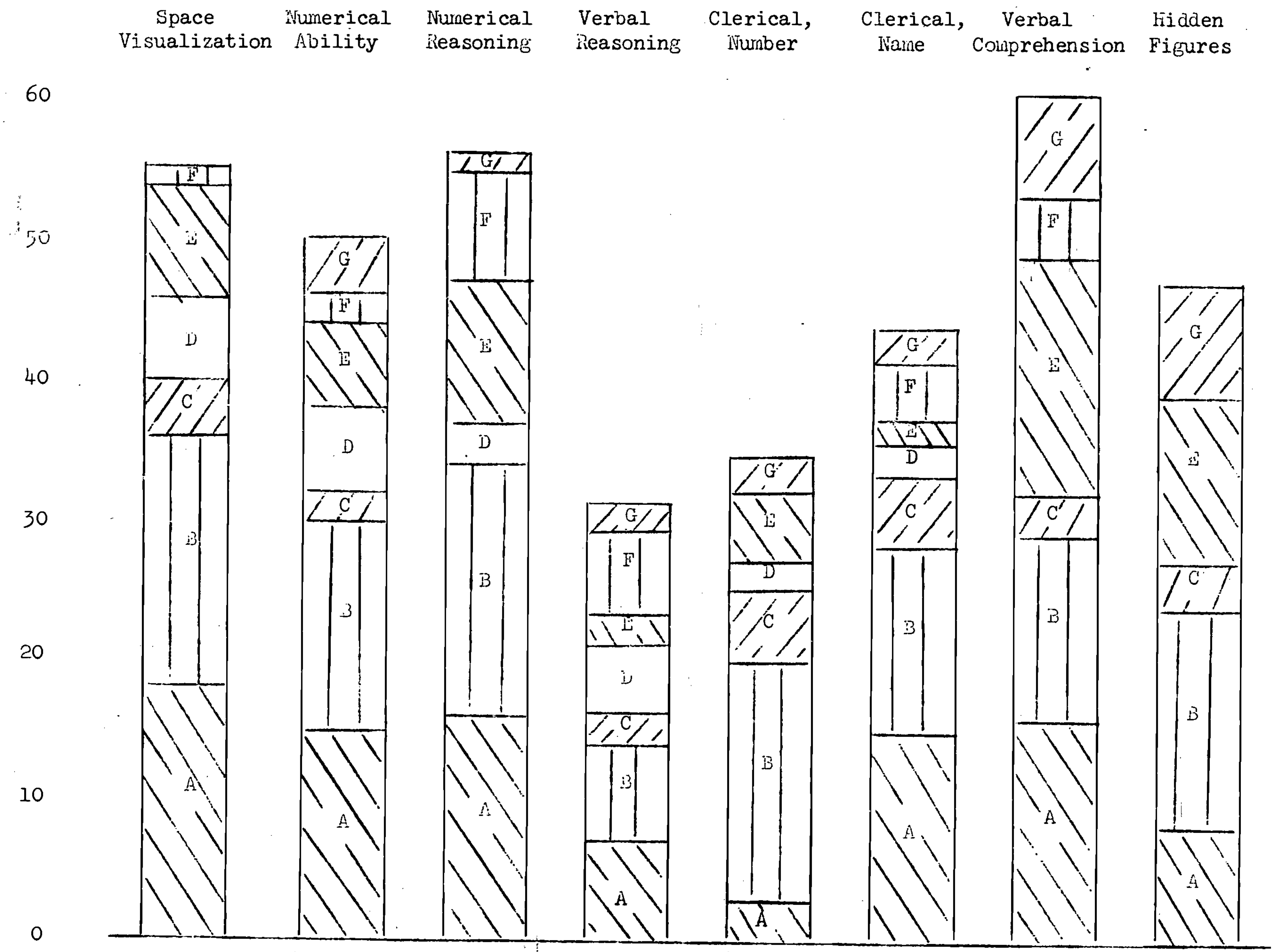

Figure 1. Cumulative Proportion of Psychometric Variance Accounted For by Laboratory Measures (see Table 3 for explanation of laboratory measures).

Figure 1 is a simplication of the data in Table 3. The measures have been grouped as shown by the letters along the right edge of Table 3. Thus, A indicates the variance accounted for by the Motor Reaction Time; B is a measure of the amount of additional variance accounted for by the remaining response time measures, C is the contribution of a carefulness-of-response measure--proportion of correct responses in the simple, two-choice reaction time study; D carries the incremental contribution of the digit span study; E does the same for the dichotic listening task, and F and G bring in, finally, any remaining contribution of the clustering on recall scores.

Discussion

Before discussing Figure 1, using Carroll's (1974) characterization of what is important to doing well on measures of the several cognitive factors, it is important to note as he did that the time limit on a test introduces a speed component to test scores of some magnitude. Because Carroll addressed himself to the question of what is important in responding to an item rather than to a speeded test as a whole, it might be expected that a generalized measure of response time would be a more significant determiner of total test performance than of success with a single item. That is, the component labelled "A" in Figure 1 may be inflated.

Space Visualization. Carroll characterized Factor Vz as dependent on a process occurring in STM whereby a spatial representation is "mentally" transformed. This process is assumed to produce both capacity and rate differences between individuals. Nothing in the present laboratory measures speaks to the question of capacity for spatial representation. However, the rate at which this representation can be manipulated may well not be unique but dependent upon more general STM processing rates. This would be consistent with the appreciable individual differences accounted for by the time components A and B.

Numerical Ability. Factor N requires retrieving number associations and arithmetic algorithms from LTM and performing serial operations on the stimulus materials using the retrieved algorithm. Individual differences, Carroll noted, appear in both content and temporal aspects of the retrieval and manipulative operations. Carroll also felt that special strategies--chunking numbers, for example--may also contribute to individual differences. Rate, either of retrieval of an algorithm and/or of application of the algorithm was apparently important and, again, could be determined by more general processing rates assessed in A and B.

Numerical Reasoning. The number series problems were most akin to Factor I, Induction, suggested by Carroll to require searching a general logic store in LTM for relevant hypotheses. The major determiner of success is whether the contents hold the solution. Some subjects, however, might construct new hypotheses by serially operating upon STM contents. A substantial amount of the variability in this test, however, was accounted for with no knowledge of the contents of an LTM logic store. Perhaps the range of inductive hypotheses likely to be employed is limited enough by the sampling of problems and subjects as to not be a source of individual differences. The importance of speed, again, plus the relatively heavy contribution of the dichotic listening and clustering scores suggests that operations on STM were indeed contributory to success.

Verbal Reasoning. Representative of Factor Rs, Syllogistic Reasoning, this test would require by Carroll's analysis retrieval of meanings and algorithms from LTM and the performance of serial operations in STM on the retrieved materials. Individual differences in content and temporal aspects of performance are postulated. There is some likelihood, as well, that Ss vary in attention paid to details of the stimulus material. Results for Verbal Reasoning are in interesting contrast to those for Numerical Reasoning. That only a relatively small proportion of the total variance was accounted for, together with a depressed contribution of the processing rate measures, suggest that indeed it is individual variability in the contents of some LTM store (rather than the speed with which it can be accessed) that accounted for individual differences in performance. An alternative explanation would be individual variability in performing transformations of statements prior to evaluating their logical consistency. If E, F, and G measure a general class of such transformations, then their small contributions to Verbal Reasoning argue against this alternative. Finally, if C, the contribution of carefulness-of-response, is interpreted as "attention to detail" (appealing in connection with the next two tests), then "attention to detail" was not an important source of variability in Verbal Reasoning.

Clerical, Number and Name. Both subtests tap Factor P, Perceptual Speed, which should be primarily sensitive to the temporal parameters of a visual search through a field for specified elements. This search, Carroll posited, occurs by addressing sensory buffers. Failure to account for more of the variance in these two tests cannot be laid to failure to tap some LTM store. It is, in fact, puzzling that there is so much left unaccounted for. The inference is that the speed of processing required in clerical work was not well explained by the speed of processing measures tapped by A and B. That C makes its greatest contribution here is suggestive of important individual differences in some kind of testing or checking loop.

Verbal Comprehension. This test and Factor V relate, roughly, to size of vocabulary and Carroll has written that performance is almost exclusively dependent upon the contents of a lexicosemantic LTM store, upon the probability that S can retrieve the correct meaning of a word. Figure 1, by contrast, reflects what has been found in other recent studies (Hunt, Lunneborg, and Lewis, 1975), that differences in verbal ability can be largely accounted for without ever inquiring about what Ss may know about the language. Being able to account for this variability for a group of adolescent native English speakers may not meet all of the goals of knowing what an individual S's vocabulary is. The Vocabulary test example, however, draws rather nicely a distinction between aptitude and achievement measures. The size of one's vocabulary is one measure of that individual's achievement. That it is so predictable from measures which are, from an information processing point of view, more basic than inventorying the contents of an LTM store suggests that vocabulary might be supplanted as an aptitude measure, however.

Hidden Figures. This last test is a measure of Factor Cf, Flexibility of Closure, and involves, by Carroll's analysis, an STM process whereby a figure is imaged in relation to a surrounding visual-representational field. Both capacity and temporal aspects may be involved. The test was, for this sample, a very difficult one--the average score was 3.5 out of a possible 16. This is reflected in the relatively small contribution of the motor reaction time measure--so few responses were made by the average S that speed of making the response was unimportant. Choice Time was important, as it was for the Spatial Visualization and Numerical Reasoning tasks, suggesting that successful test performance requires rapid evaluation of alternatives. Because of the purely figural nature of the stimulus material it is interesting to note the rather large contributions of E and G, measures grounded in symbolic and semantic manipulations. The suggestion here is that the ability to effect these manipulations was not independent of the ability to manipulate figural material.

Despite the shortcomings of the available data these analyses suggest that Carroll's thoughtful analysis of the demands psychometric tests place on human beings as information processors can be fruitfully linked to laboratory measures to gain better understanding of what the psychometrists call intelligence and what the experimental psychologist calls cognition. Interestingly, quite basic measures of individual differences in information processing--
motor and choice reaction times--appeared to be strongly predictive of individual differences in test performance across a range of content.

Study 2

The rather consistent relations found in this first study were surprisingly strong and replication, if not extension, of the results was in order. In this second study, the psychometric instruments were selected to provide a more systematic (and psychometrically reliable) coverage of the cognitive domain, including, in particular, better coverage of those cognitive factors which Carroll (1974) depicted as making the more minimal demands of the information processing system; tasks which in the distributed memory model focus on the sensory register and short term memory. At the same time, the laboratory measures of choice response time were adjusted to provide more reliable assessments of both practiced and unpracticed performance.

Method

Subjects. University freshmen were recruited fall 1974 from those who had completed the battery of tests administered high school juniors spring 1973 by the Washington Pre-College (WPC) Testing Program. Volunteers were selected so that the joint distribution of the WPC Verbal and Quantitative Composite scores in the sample reflected the distribution for the entire entering class.

Ss were paid and their participation, as regards the data reported here, extended over an academic quarter. Typically, the S would be involved in one psychometric and one laboratory session each week. The total group reported on here consisted of 63 freshmen, 27 men and 36 women.

Psychometric instruments. Those instruments contributing to the present study can be grouped under conventional psychometric headings. Because of the number of tests involved a preliminary analysis of measures under each heading was undertaken to reduce the data utilized in later analyses.

1 and 2. Verbal ability measures: The WPC battery administered in high school provided scores from tests of Reading Comprehension, Vocabulary, Spelling, and English Usage (or, Grammar). (The Verbal Composite score used in subject selection is a linear weighting of these four tests.) Verbal tests administered following subject selection included the EAS test of Verbal Comprehension used in Study 1 and the two subparts of the Terman Concept Mastery test: Synonyms and Antonyms and Verbal Analogies. In the French (1963) scheme all tests would be considered measures of Factor V: Verbal Comprehension.

To reduce the number of verbal measures to be treated here a principal components analysis of the correlations among these measures followed by a varimax rotation of the larger of these components suggested a reduction to two measures:

(1) VOCABULARY - Average of standardized WPC Vocabulary and Terman Synonym-Antonym scores;
and

(2) GRAMMAR - Average of standardized WPC Spelling and English Usage scores.

3. Quantitative ability measures. The WPC battery provided three quantitative subtests: Quantitative Skills (numeric reasoning), Applied Mathematics (word problems at the elementary algebra level) and Mathematics Achievement (achievement in high school mathematics). The three are linearly combined to provide the Quantitative Composite measure used in subject selection. Additional quantitative tests administered this sample included the EAS tests of Numerical Ability and Numerical Reasoning described for Study 1. One principal component appeared to tap the

variability central to all these measures, the Number Facility (N) factor being probably more important than the Reasoning (R) factor sometimes associated with certain of these subtests, and in the present analyses only a single quantitative measure was employed:

(3) QUANTITATIVE COMPOSITE - the WPC overall quantitative score.

4 and 5. Spatial ability measures. There was a single Spatial Ability measure included in the WPC. It is a paper folding test of the kind classified by French (1963) as a measure of the Visualization (Vz) factor. Two other Visualization measures were administered, the Two Dimensional and Three Dimensional Spatial Relations tests of the California Mental Abilities Test (CMAT). The two-dimensional test is a paper form board--determining which pieces fit together to form a desired plane figure--while the three-dimensional test is one of surface development--determining which planar figure can be folded into the desired three dimensional form.

Other figural tests included the Hidden Figures test which was used in Study 1 and loads the Flexibility of Closure (Cf) and, to a lesser degree, the Spatial Orientation (S) factor as well; a Maze Tracing task associated with a Spatial Scanning (Ss) factor; and the Advanced Progressive Matrices (Set II) of Raven. This last test is variously considered a measure of general intelligence ("g", and culture free at that), or an indicator of a reasoning factor styled Induction (I), or a "space" measure. The task is to pick a figure to substitute for a missing part of a larger figure so as to "best complete the design." It is included here because in the present study it had sizeable correlations with other space measures.

The principal components/varimax rotation analyses of these figural or spatial tests suggested two clusters and--for this report--two new composite measures were defined:

(4) SPACE ONE - Average of standardized Maze Tracing and paper form board test scores; and

(5) SPACE TWO - Average of standardized Progressive Matrices, Hidden Figures, surface development, and paper folding scores.

6 and 7. Reasoning ability tests. Two other of the short EAS tests were administered--Verbal Reasoning, used in Study 1 as a measure of the Syllogistic Reasoning (Rs) factor, and Symbolic Reasoning. The last test

involves the concepts--presented using the customary symbols--of equality and inequality and poses questions of the form "If $X \leq Y$ and $Y \geq Z$ then is it true that $X = Z$?" It appeared to also sample the R_G factor.

The Watson-Glaser Critical Thinking Analysis was also administered and scores obtained for the five subtests--Inference, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments. With the possible exception of Deduction these subtests are probably factorially complex--calling for a comprehension of material read and its integration with some already known information as well as the excercise of reasoning.

When interrelations among the tests were analyzed the Symbolic Reasoning test was more strongly related to the quantitative measures while the EAS Verbal Reasoning and three of the Watson-Glaser subtests--Recognition of Assumptions, Deduction, and Interpretation--were moderately correlated with one or another of the verbal ability measures. In consequence, only two reasoning measures were retained:

(6) INFERENCE - On the basis of a paragraph read, Ss classify a series of statements as true, probably true, false, probably false or based on insufficient data;

and

(7) EVALUATION OF ARGUMENTS - a "policy question" is followed by a number of statements pro and con to be classified as weak or strong arguments on the question.

8, 9, and 10. Perceptual/motor speed tests. A number of tests were administered Ss which were highly speeded, placing a premium on quick perception and response. The Survey of Working Speed and Accuracy, for example, has four five-minute sections: Number Checking, comparing pairs of numbers for identity; Code Translation, substituting letters for integers; Finger Dexterity, dotting small circles; and Vowel Counting, obtaining separate counts of A, E, I, O, and U occurrences in each line of text. Number Checking, Vowel Counting, and possibly the code substitution task are measures of the Perceptual Speed (P) factor. Finger Dexterity has been regarded as motor rather than cognitive.

A Code Recognition test was administered from the Flanagan Aptitude Classification battery. Ss were given practice in learning alphanumeric codes for common objects and then tested immediately on their recognition of these codes. The test is a recognition form of the measures used as markers for the Associative Memory (Ma) factor.

Finally, three locally constructed tests were administered, based upon the description given by Rose (1974) of a battery developed at the Human Performance Center, University of Michigan. The three were the Grammatical Reasoning test, requiring Ss to compare a statement "A does not precede B," for example, with a letter pair AB or BA and decide whether the statement is true or false; the Neisser Letter Search test, requiring Ss to check off blocks of "randomly" assembled letters that contain a key letter (in successive repetitions Ss search for one, two, and four letters); and the Fitts Tapping test in which Ss place pencil dots within circular targets whose size and separation distance are systematically varied.

Letter Search should be another Perceptual Speed test. Grammatical Reasoning was used as a paper and pencil surrogate for a laboratory procedure used in earlier studies (Hunt, Lunneborg, and Lewis, 1975). Similarly, the tapping task was intended to provide a paper and pencil measure of a fairly unencumbered response speed.

Because the elements of this "Michigan battery" were so simple they were administered under highly speeded conditions. The battery was administered twice on each of two occasions separated by about three weeks. (The two administrations on any one occasion included two logically parallel forms of the Grammatical Reasoning and Letter Search tests and two exact repetitions of the tapping tasks.)

Upon analysis, the between sessions correlations for the same test were disappointingly low while the between test correlations for the final session were all in excess of .90. As a representative of the "Michigan Battery" one score was retained:

(8) LETTER SEARCH - Average of standardized scores for the searches for 1, 2, and 4 letters on the fourth administration.

An analysis of the correlations among the other perceptual/motor test scores suggested three dimensions marked in turn by Number Checking, Code Translation, and Vowel Counting. As Vowel Counting was strongly correlated with Letter Search only the first two scores were kept for the present analyses:

(9) NUMBER CHECKING - Comparing pairs of numbers
and

(10) CODE SUBSTITUTION - Replacing integers with arbitrarily paired letters.

Of the ten psychometric scores, six were based on essentially unspeeded performance while four, LETTER SEARCH, NUMBER CHECKING, CODE SUBSTITUTION, and, because of the Maze Tracing test, SPACE ONE, are more heavily speed dependent.

Laboratory Measures. As with the psychometric data, clusters of scores--usually those resulting from a single experiment--were separately analyzed to reduce the number of scores to be related to paper and pencil test performance. It should also be noted that for the present report interest was centered only on speed or time data from the laboratory procedures. Much of the data provided by these Ss will not be considered here.

1 and 2. Binary choice reaction time. Ss fixated a midpoint on a CRT display and had to respond by key press to the onset of a character display to the left or right of that midpoint. Each S was run under two conditions. In the one-finger condition S used index finger of preferred hand to depress one of two keys. In the two-finger condition Ss used keys under left or right hand respectively to signal detection of display to left or right of screen. Data studied here were median response times, separately for presentations to the left and right, for the first and last (sixth) blocks of 100 presentations under each of the conditions. Correlationally, the data clustered by condition and two composite measures were developed:

(1) CHOICE R. TIME, ONE FINGER - Average of standardized scores for left and right presentations, first and last block.

(2) CHOICE R. TIME, TWO FINGER - Same average, based on two-finger condition.

3, 4, and 5. Delayed auditory feedback study. In this procedure Ss were asked to read aloud a page length stream of nonsense syllables under three conditions. S's reading was taped and played back through earphones at zero, 150, or 300 millisecond delay. Although study of types of errors motivated this study, two time scores were of interest here:

(3) NO DELAY READING TIME - Time to complete reading of passage when feedback was not delayed;

and

(4) DELAY READING TIME - Average of standardized times to complete reading when feedback was delayed.

During this same experimental session Ss were set the task of repeating back to the experimenter a word replacing R's with L's and L's with R's in so doing. For the present analyses one score was abstracted from the resulting data:

(5) R/L LATENCY - Median time to transform a word.

6 and 7. Raven response times. Set I of the Advanced Progressive Matrices described earlier was presented one problem at a time and solution time measured for each figure. This was done after group administration of Set II and involved 12 problems not contained in the larger set. Times for solving earlier and later problems in this set clustered and here two scores were employed:

(6) RAVEN TIME 1 - Average of standardized times to solve problems 1 and 2;

and

(7) RAVEN TIME 2 - Average of standardized times to solve problems 8 and 9.

8 and 9. Problem solving time. Another problem solving study called for Ss to learn a list of statements which either associated each of a set of personal names with an occupation (Ruth is a lawyer), associated these same personal names with locations (John is in the pool), or associated these locations with activities (If you are in the park you can hear the chimes ring). Ss were given a week to learn the list and, following mastery, each was asked a series of questions to be classified as True or False. The questions were such as to require the retrieval of 1, 2, 3, or 4 of the learned facts. Time to answer each question was recorded and the two scores of interest here were the parameters of the individual line best fitting response times (for correct answers) to the number of facts to be retrieved:

(8) PROBLEM SOLVING SLOPE

and

(9) PROBLEM SOLVING INTERCEPT.

Results

Table 4 reports zero order correlations between the psychometric and laboratory measures. Choice reaction times were only weakly related to the psychometric measures, and in particular to the speeded measures. This contrasts sharply with the Table 1 results for the high school group of Study 1. Because of this difference in magnitude of the relations, the stepwise regression analyses of Study 1 were not repeated. Rather, a series of principal components analyses were completed to ascertain whether the relations between psychometric performance and laboratory measures of information processing speed would be rendered clearer by the definition of broader, composite measures.

Table 5 speaks to the relations among the selected psychometric measures. The two verbal scores had different loadings as did the two space measures, the two reasoning subtests, and the three perceptual tasks. The quantitative score, by this crude index, had inter-test relations not unlike those for the second space measure.

Only one of the rotated components, 4, was very heavily identified by speeded measures and might be a perceptual speed indicator. Component 1 had the appearance of a general cognitive power dimension--high loadings by verbal, quantitative, and spatial ability measures. Component 3 suggested an accuracy in rule following and Component 2, perhaps, accuracy in comparison.

Relations among the laboratory time measures underlie the principal components analysis summarized in Table 6. Component 1 in this analysis represented speed (actually, slowness) of response when the processing load was minimal; Component 3 would seem to reflect speed of response when there was more pressure on the system--delay of auditory feedback, overcoming phonetic convention, searching LTM for learned facts; Component 4 appeared to associate speed of reading through an LTM store, with speed of reading through an external (DAF text) file.

Interestingly the speed of solving the figural Raven problems was independent of the speed of solving the verbal problems and defined a separate Component. The problems differed in other ways, of course; the matrix problems became quite difficult and depended for their solution upon something other than a closed set of recently acquired bits of information.

Table 4

Correlations Between Laboratory and Psychometric Measures, Study 2

(Decimal points omitted. (S) denotes speeded psychometric scores.)

	Vocab	Grammar	Quant Composite	Space one (S)	Space two	Inference	Eval argument	Letter search (S)	Number check (S)	Code Sub (S)
Choice R. time, 1F	10	-17	-16	-13	05	-09	-21	-06	02	-11
Choice R. time, 2F	-10	-19	-14	01	-03	00	-21	07	11	00
Read time, no delay	03	-10	14	00	17	-02	04	-08	11	-05
Read time, delay	-31	-45	-22	04	12	-22	-38	10	-11	-21
R/L latency	-54	-46	-38	-05	-12	-11	-07	11	01	-09
Raven, time 1	-12	-18	-22	-30	-36	05	09	12	-10	14
Raven, time 2	-04	-07	02	-13	-08	01	09	10	06	-05
Problem solving slope	28	01	23	-36	03	25	-05	-05	-07	-02
Problem solving inter	-16	-45	-11	02	17	-26	-06	16	-44	-26

Table 5

Loadings of Psychometric Measures on Varimax Rotated Components, Study 2

(Only correlations of .40 or larger indicated. Decimal points omitted.

(S) denotes a "speeded" measure.)

	Component			
Measure	1	2	3	4
Vocabulary	68			
Grammar	61		53	
Quant composite	81			
Space one (S)	55			54
Space two	80			
Inference		90		
Eval argument			74	
Letter search (S)				89
Number checking (S)		45		
Code substitution			62	

Table 6

Loadings of Laboratory Time Measures on Varimax Rotated

Principal Components, Study 2

(Only correlations of .40 or larger indicated.

Decimal points omitted.)

Measure	Component 1	2	3	4
Choice time, 1F	71			
Choice time, 2F	93			
Read time, no delay	42			64
Read time, delay			83	
R/L latency			68	
Raven time 1		89		
Raven time 2		81		
Problem solving slope				81
Problem solving intercept			67	

Table 7, finally, reports the results of a rotated components analysis involving both psychometric and laboratory measures. That the number of components, 8, was equal to the total of the number found in the separate analyses suggests the two sets did not fit together very neatly. Indeed, four of these eight final components were loaded from only one of the sources. Following the notation, "Possible Source," at the bottom of Table 7, these corresponded, roughly, to laboratory components 1 (L1, response slowness under minimal load) and 2 (L2, time for solving the Raven figural problems) and psychometric components 3 (P3, rule following accuracy was the tentative name) and 4 (P4, perceptual speed).

The two remaining components from each source appeared to cross with each other in this final analysis. Thus the quantitative and space measures broke away from the verbal in the first psychometric component to associate in the fifth component of this second analysis with a reduced reading speed (perhaps symptomatic of care in reading). The verbal tests on the other hand clustered more closely with speed of read-back when an interfering task is surimposed on that read-back, Component 1. Similarly, the third laboratory component split into this read-back under interference and an association of overall problem solving time (difference in intercept) with number checking or attention to detail, Component 3. The final, shared component, Component 6, was somewhat puzzling. It suggested that greater increases in problem solving time associated with retrieving an additional bit of information were related to higher scores on the inference test. Although not intuitively appealing, one possibility is that this component tapped an accuracy (vs. speed) dimension.

Discussion

Two of the final components, 5 and 6, were suggestive of an association between facility in psychometric problem solving where speed is not a critical factor and less rapid information processing in laboratory problem solving studies. In the other two components of interest, 1 and 3, other processing speeds in the laboratory (DAF, R/L, Intercept) were positively associated with accuracy on easier, or better learned, psychometric tasks.

Table 7

Loadings of Psychometric and Laboratory Measures on Varimax Rotated Principal Components, Study 2

(Only correlations of .40 or greater indicated. Decimal points omitted.)

Measure	Component 1	2	3	4	5	6	7	8
Vocabulary	79							
Grammar	56		41					
Quant composite	47				68			
Space one (S)							70	
Space two					75			
Inference			46			60		
Eval argument					48			
Letter search (S)							77	
Number checking (S)			87					
Code substitution (S)								90
Choice time, 1F		81						
Choice time, 2F		88						
Read time, no delay		52			50			
Read time, delay	-49			-42				
R/L latency	-86							
Raven, time 1				81				
Raven, time 2				81				
Problem solving, slope						86		
Problem solving, inter			-73					
Possible Source:	L3	L1	L3	L2	L4	L4		
	P1		P2		P1	P2	P4	P3

The picture appears somewhat different than in Study 1. There, directly measured reaction time to simple stimuli was strongly correlated with a host of apparently diverse psychometric measures. Here, response times--from the binary choice study--were lowly correlated with psychometric performance and appeared to define a component in isolation in the final analysis.

Study 3

The differences in findings between the first and second studies could have resulted from differences in measurement techniques, from differences in subject sampling, or from other, non-obvious differences between studies. The most obvious follow-up was to repeat the measures developed in the second study with a subject sample more nearly like those who participated in the first study. Would stronger relations be found if students at the high school level were used as subjects?

Method

Subjects. At the end of spring term 1975 high school juniors who had completed the WPC battery were recruited through high school counselors. Sixty-four volunteers were run in groups of eight for two days apiece. Three hour morning sessions were devoted to psychometric testing and slightly longer afternoon sessions involved Ss is laboratory data collection. Ss were paid at an hourly student rate for participation.

Measures. Although this more recent high school study involved other tasks, results will be reported here only on a subset of tasks that were in common with Study 1 and, to a greater extent, Study 2. In particular, scores were obtained on ten earlier described psychometric instruments:

1. EAS Verbal Comprehension
2. EAS Numerical Ability
3. EAS Numerical Reasoning
4. EAS Verbal Reasoning
5. EAS Symbolic Reasoning
6. Advanced Progressive Matrices (Set II)
7. Maze Tracing
8. Hidden Figures
9. Two Dimensional Spatial Relations (CMAT)
10. Three Dimensional Spatial Relations (CMAT)

Tests were administered under the same conditions as for Study 2 and, where applicable, Study 1.

The Binary Choice Reaction Time task of Study 2 was repeated with the high school groups, but in a slightly modified format. Inspection of individual records in Study 2 suggested that stabilization was as likely after 300 trials as after 600. Ss is this third group, therefore, completed only 300 trials (exclusive of an initial block of 30 warm-up trials). Also, only the faster two-finger version of the task was employed.

Results

Table 8 summarizes the zero order correlations between the binary choice reaction time tasks and a set of psychometric scores for this study (High School 1975) and for Studies 1 (High School 1973) and 2 (College 1975). (Data from the two earlier studies are reported here in slightly different form: actual choice response times are used from Study 1 rather than the difference between choice and motor response times used in Tables 1 - 3 and, for Study 2, choice response times are here reported separately for the first and last blocks of trials.) The significant relations of Study 1 were not replicated, for either task, with the college sample. For the second high school group, however, there is a tendency for these relations to emerge again, particularly for the Block 3 response times.

Means and standard deviations for the three groups on these variables are reported in Table 9. In terms of average psychometric performance, the second high school group was only slightly less proficient than the college group. The initial high school group, however, was markedly less able. Choice response times are not markedly different between the groups. Of greater potential significance, however, are the differences in variability between the groups. While there is no consistent pattern of differences on the psychometric instruments the two high school groups are appreciably more variable on the choice response time tasks.

Discussion

The data for the three groups of subjects taken together provide rather clear evidence that the answer to the question of what relations exist between choice reaction time and performance on psychometric

Table 8

Correlations between Laboratory Choice Response Times and Psychometric

Performances, Studies 1, 2 and 3

	One finger choice time			Two finger choice reaction time			
	H.S. 1973	College 1975		College 1975		High school 1975	
		Block 1	Block 6	Block 1	Block 6	Block 1	Block 3
Verbal comprehension	-.46**	.10	.14	-.11	.04	.14	.09
Numeric ability	-.48**	-.06	-.02	-.19	.04	-.04	-.19
Numeric reasoning	-.51**	-.10	-.08	-.22	.04	-.15	-.12
Verbal reasoning	-.33**	-.04	-.24*	.02	.17	-.00	.05
Symbolic reasoning		-.08	-.17	-.08	.05	-.08	-.31**
Progressive matrices		-.05	-.00	-.11	-.02	.11	-.27*
Maze tracing		-.12	-.19	-.03	.00	-.29*	-.28*
Hidden figures	-.38**	-.03	.01	-.11	-.13	-.13	-.29*
Two dimensional space rel'n.		-.02	-.05	.06	.02	-.10	-.34**
Three dimensional space rel'n.	(-.55)**	.11	.09	-.05	.01	.03	-.28*

*p < .05, **p < .01, directional hypothesis.

Table 9

Means and Standard Deviations, Choice Times and Test Scores,

Studies 1, 2, and 3

	High school 1973		College 1975		High school 1975	
	$\overline{X}$	S.D.	$\overline{X}$	S.D.	$\overline{X}$	S.D.
Verbal comprehension	13.1	6.0	19.4	4.2	18.6	4.0
Numeric ability	15.8	7.8	27.5	8.3	26.2	8.0
Numeric reasoning	8.2	3.8	13.8	2.4	12.5	2.9
Verbal reasoning	13.0	5.9	18.6	3.8	17.4	4.8
Symbolic reasoning			16.0	6.3	14.0	6.9
Progressive matrices			26.5	4.8	24.0	6.4
Maze tracing			24.1	7.7	26.4	7.0
Hidden figures	3.4	3.1	11.2	6.3	10.2	6.9
Two dimensional space rel'n.			16.3	5.0	18.6	4.1
Three dimensional space rel'n.			15.3	4.4	15.7	5.0
One finger CRT, initial	338.0	83.7	331.2	43.8		
One finger CRT, terminal			321.3	37.1		
Two finger CRT, initial			285.4	27.0	303.8	71.2
Two finger CRT, terminal			270.3	28.6	288.2	42.8

instruments can only be attempted with a particular population of Ss specified. The spread of choice times between individuals seems particularly influential and it is variability in this performance, not directly a basis of selection in college recruitment, that is severely attenuated in the entering population of, at least one, state university. It would appear that the risk of attenuating information processing relations of interest, indeed, of failing to detect them, is high unless the researcher taps a range of performance that is typical of the unselected high school population. Narrowing the range of cognitive ability below that may be expected to result in studies yielding relations which are complex, and difficult to replicate. Using the fuller range, however, holds the prospect of revealing stronger, more general links between laboratory based cognitive performance and mental tests.

References

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. Spence and J. Spence (Eds.), The psychology of learning and motivation II. New York: Academic Press, 1968.

Atkinson, R. C., & Shiffrin, R. M. The control of short term memory. Scientific American, 1971, 225, 82-90.

Carroll, J. B. Psychometric tests as cognitive tasks: A new "structure of intellect." Paper presented at the LDRC Conference on the Nature of Intelligence, Pittsburgh, March 1974.

Clark, H., & Chase, W. On the process of comparing sentences against pictures. Cognitive Psychology, 1972, 3, 472-517.

Day, R. S., Cutting, J., & Copeland, P. Perception of linguistic and nonlinguistic dimensions of dichotic stimuli. Paper presented at 12th annual meeting of the Psychometric Society, November 1971.

French, J. W., Ekstrom, R. B., & Price, L. A. Manual for kit of reference tests for cognitive factors. (Revised 1963) Princeton, N.J.: Educational Testing Service, 1963.

Green, D. R. (Ed.) The aptitude-achievement distinction. Monterey, Cal.: CTB/McGraw-Hill, 1974.

Hunt, E. B. What kind of computer is man? Cognitive Psychology, 1971, 2, 57-98.

Hunt, E. B. The memory we must have. In R. Schank & K. Colby (Eds.), Computer simulation of information processes in man. New York: W. H. Freeman, 1973.

Hunt, E. B., Frost, N., & Lunneborg, C. E. Individual differences in cognition. In G. Bower (Ed.), Advances in learning and memory, Vol. 7. New York: Academic Press, 1973.

Hunt, E. B., Lunneborg, C. E., & Lewis, J. What does it mean to be high verbal? Cognitive Psychology, 1975, 7, 194-227.

Massaro, D. W. Preperceptual and synthesized auditory storage. Technical report 72-1. Madison: University of Wisconsin Psychology Department, 1972.

Nelson, T. O. Response probability and response latency in long-term retention. Presented at American Psychological Association, Montreal, 1973.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual items. Journal of Experimental Psychology, 1959, 58, 193-198.

Posner, M. I., & Boies, S. Components of attention. Psychological Review, 1971, 78, 391-408.

Puff, C. R. Clustering as a function of sequential organization of stimulus word lists. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 503-506.

Sternberg, S. Memory scanning: Mental processes revealed by reaction time experiments. In J. S. Antrobus (Ed.), Cognition and affect. Boston: Little-Brown, 1970.

Trabasso, T. Mental operations in language comprehension. In J. B. Carroll and R. V. Freedle (Eds.), Language comprehension and the acquisition of knowledge. Washington: Winston-Wiley, 1972.

Wickens, D. Encoding categories of words: An empirical approach to meaning. Psychological Review, 1970, 77, 1-15.