Choice Reaction Time : What Role in Ability Measurement?

Clifford E. Lunneborg

Applied Psychological Measurement 1977 1: 309
DOI: 10.1177/014662167700100301

Choice Reaction Time: What Role in Ability Measurement?

Clifford E. Lunneborg
University of Washington

Three studies are described in which choice reaction time (RT) was related to such psychometric ability measures as verbal comprehension, numerical reasoning, hidden figures, and progressive matrices tests. Although fairly consistent negative correlations were found between these tests and choice RT when high school samples were used, differences from study to study highlight the need to develop more reliable measures for cognitive laboratory procedures and to study these in populations that are more broadly representative of human cognitive power.

Introduction

Two groups within psychology have been concerned with a common issue. Psychometricians have developed an array of individual difference constructs—verbal ability, perceptual speed, spatial visualization—to explain human intellectual performance. At the same time, cognitive psychologists (e.g., Hunt 1971, 1973) have elaborated models of human information processing which include such constructs as rate of transfer of information, size of information store, and strategy availability. At issue in both of these developments is understanding how humans think.

Each approach has its strengths and through them can aid the other. Although the models developed by cognitive psychology provide challenging hypotheses about underlying processes, such models tend to describe the ideal human. Psychometrics is founded on the recognition that there is no ideal intelligence and has evolved the measurement of differences between people to the status of a science. By providing a methodology for building reliable measures of individual differences in the parameters of cognitive theories, psychometrics provides cognitive psychologists with ways to test and modify their models. Such models can therefore become explanatory of the full range of human mental capacity. In return, cognitive psychology affords psychometrics a process orientation to intelligence which may be more promising than the present product orientation which equates intellectual abilities with achievements.

There should also be appreciable social gains to shifting the focus of individual differences from achievement to process. Intellectual assessment has come under considerable criticism for labeling people who may simply not have had the opportunity to achieve as lacking ability or aptitude. Such a shift would also improve the diagnostic value of psychometrics, particularly in education. For example, the search for aptitude-treatment interactions as a way of improving education has

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 1, No. 3 Summer 1977 pp. 309–330

been disappointing. The classical aptitudes have not interacted with different teaching styles to improve assignment of students to instructional techniques. However, if the information processing models identify "aptitudes" conceptually more closely tied to teaching strategies, then more efficient use of teaching and learning resources can be realized.

Psychologists have begun to study the relationships between the constructs of psychometrics and the processes of cognitive models. A series of laboratory studies at the University of Washington (UW) has provided information processing descriptions, for example, of high and low verbal subjects (Hunt, Lunneborg, & Lewis, 1975). Carroll (1974) has begun a similar line of research by describing the information processing requirements of each of the tests included in the *Kit of Reference Tests for Cognitive Factors* (French, Ekstrom, & Price, 1963). The present research was begun to relate Carroll's analytic efforts to the laboratory data being generated at UW.

Time needed to accomplish information processing tasks is a key element to the cognitive models and could be expected to show variability between people. Time is required, for example, to retrieve information from long-term memory (LTM). Similarly, time is required to apprehend a stimulus. Both times can be expected to vary across persons. The extent of such variability as well as the covariation between such times must be established if we are to know which and how many parameters account for individual differences in intellectual performance. The simplest of the processing times, reaction time, is important because it is a component of all other processing times. Its variability and covariation with other measures, therefore, need to be determined first.

Depending on whether reaction time covaries with other information processing parameters, two predictions might be made about the correlations of it with paper and pencil test performance. If simple reaction time is independent of the other processing parameters, then Carroll's (1974) analysis would predict negative correlations with some tests and zero correlations with others. Negative correlations would result with tests where performance was largely determined by quickness in recognition, and zero correlations would result with tests where performance was based on higher levels of processing. Negative correlations would result from shorter times to respond being associated with higher test scores. In contrast, the speculation of Hunt, Frost, and Lunneborg (1973) that individuals may have a characteristic "cycle" time suggests that all information processing times, including the simplest, are positively intercorrelated and would lead to negative correlations between reaction time and performance on all intellective tests. The present paper summarizes three studies designed to shed light on the role of simple choice reaction time in higher-order intellectual functioning as well as to test Carroll's hypothesized processing requirements for paper and pencil tests.

STUDY 1

The purpose of the first study was to examine the extent to which Carroll's (1974) description of the cognitive demands of several simple tests (French et al., 1963) was matched by correlations between performance on similar tests and laboratory measures of information processing. The study was based on a reanalysis of data collected prior to Carroll's paper so that not all of his cognitive tasks were included.

Method

Subjects

Sixty-four high school age participants, 33 males and 31 females from the Seattle area, were recruited during summer 1973 via a community Job Line. In groups of eight, they spent half a day of each of two consecutive days in the laboratory. During each session, they participated in several laboratory assessments and completed a battery of short paper and pencil tests. Participants were paid on an hourly basis.

Psychometric Instruments

1. *Employee Aptitude Survey (EAS) Verbal Comprehension* (Ruch & Ruch, 1963) calls for the participant to select a synonym for each of a series of words. In the Reference Kit scheme, this test measures Factor V, Verbal Comprehension.
2. *EAS Numerical Ability* requires participants to recognize the answers to simple arithmetic problems. This test measures Factor N, Number Facility.
3. *EAS Space Visualization* asks that adjacent blocks in a stack be counted. The test is considered a measure of Factor Vz, Visualization.
4. *EAS Numerical Reasoning* presents number series and asks that the next element be selected from among alternatives. This test taps Factor I, Induction, as well as Number Facility.
5. *EAS Verbal Reasoning* provides deductive reasoning tasks and is a measure of cognitive factor Rs, Syllogistic Reasoning. Proposed conclusions are reported to be true, false, or of indeterminate validity.
6. *Clerical Number:* This clerical test (Andrew & Paterson, 1946) requires the comparison of pairs of numbers and, subsequently, of pairs of personal or business names. *Clerical Number* measures Factor P, Perceptual Speed.
7. *Clerical Name:* Should also tap Factor P, Perceptual Speed, and is from the same test as 6.
8. *Hidden Figures* from the French et al., (1963) Kit is a measure of Factor Cf, Flexibility of Closure. Participants are asked to identify which polygon is hidden (included) in a series of more complex figures.

Laboratory Measures

1. *Motor Reaction Time (RT):* Median time to respond by key press to the onset of a + centered on a cathode ray tube (CRT) screen.
2. *Choice Time:* Median time to respond differentially by depressing a left or right key to a circle either in the left or right half of the CRT display was calculated. *Choice Time* was the difference between this median time and median motor RT.
3. *Prop Correct:* Proportion of choice reaction time trials on which the correct choice was made.
4. *Stroop Difference:* In a modification of the Stroop task, participants were asked first to name the colors in which a series of asterisks were printed and then to name the color of printing of an equivalent length list of contrasting color names. The task was repeated twice and the *Stroop Difference* score was the average difference in "reading times" between the name and asterisk conditions.
5. *Search Slope:* In an experiment in the Sternberg (1970) paradigm, one to six consonants were sequentially shown on a projection screen followed by a single probe consonant. Participants were asked whether the probe was in the previously exposed set. Response times were recorded and, for correctly identified instances, these times linearly regressed on the number of digits in the associates target set. *Search Slope* was the slope of this regression line.
6. *Search Intercept:* Intercept of the above regression line.
7. *Digit Span Final:* Strings of fifteen digits were presented binaurally with recall cued immediately following presentation. Recall instructions were for subjects to report in order as many digits as possible beginning with the first digit heard. Ten trials were presented with each scored for number of digits recalled in order. *Digit Span Final* is the average of performance on the last five trials.
8. *Digit Span Gain:* Digit Span Final minus average performance on the first five trials.

9. *Dichotic Category:* 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, recall was requested. On some blocks of trials, participants were to report by ear and on others by category (digit or consonant). The *Dichotic Category* score was the number of items correctly reported over all 40 category trials.
10. *Dichotic Difference:* Score based on ear trials alone minus Dichotic Category score.
11. *Clustering Base:* A final set of four scores was obtained from a semantic clustering task based on the experimental paradigm of Puff (1966). During their first laboratory session, students were shown two lists of 30 common nouns, item by item. Each list consisted of ten nouns from each of three semantic categories (fruits, occupations, animals, etc.). Each list was shown twice with recall asked for immediately following each presentation (four recalls, first day). For half the students, 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 sample, this order was reversed. When participants returned for the second day in the laboratory they were asked to recall each of the two lists (two additional recalls, second day). A semantic clustering score was computed for each of the six recalled lists. *Clustering Base* was the clustering score for the second presentation of the blocked list (first day).
12. *Clustering Difference:* Clustering Base minus clustering score for second presentation of pseudo-random lists, also on first day.
13. *Clustering Delay:* Clustering Base minus clustering score for second day recall of the blocked list.
14. *Clustering Delay Difference:* Clustering score for second day recall of blocked list minus clustering score for second day recall of pseudo-random list.

There are above three types of laboratory response measures: *time* (Motor RT, Choice Time, Stroop Difference, Search Slope, Search Intercept), response *accuracy* (Prop Correct, Digit Span Final, Digit Span Gain, Dichotic Category, Dichotic Difference) and response *clustering* (Clustering Base, Difference, Delay, and Delay Difference).

Results

Correlations between the psychometric and laboratory data are shown in Table 1. The five response time measures correlated negatively with the paper and pencil test scores. As for the accuracy measures, Prop Correct correlated weakly, though positively, with the psychometric measures. Prop Correct, however, had little variability as few errors were made (average proportion correct was .93). The negative correlations involving Digit Span Gain suggest that inasmuch as average improvement was small (from 5.75 digits to 6.23 digits), large gains were registered by students who had underperformed on the initial trials. Digit Span Final scores were not highly correlated with any of the psychometric measures.

The Dichotic Category score correlated positively with nearly all tests. Dichotic Difference (ear minus category scores), on the other hand, correlated negatively, suggesting that high test scores were earned by students with category scores much higher than ear scores. (For nearly everyone, the ear scores were smaller than the category scores.)

The clustering score, based on the number of interruptions of semantic strings, was such that lower scores indicated greater semantic clustering. Except for the reversal of signs, Clustering Base showed the same degree of correlation as the Dichotic Category scores.

Table 1

Correlations between Psychometric and Laboratory Measures, Study 1[a]

Laboratory measure	Verbal comprehension	Numerical ability	Space visualiza	Numerical reasoning	Verbal reasoning	Clerical number	Clerical name	Hidden figures
Motor RT	-40**	-38**	-42**	-40**	-27*	-17	-38**	-27*
Choice time	-40**	-45**	-55**	-50**	-30*	-28*	-37**	-40**
Prop correct	21	11	12	04	14	10	14	17
Stroop difference	04	-27*	-19	-32**	-16	-35**	-38**	-09
Search slope	-07	-10	-27*	-16	-07	-24	-26*	03
Search 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
Dichotic category	55**	46**	49**	52**	32**	02	26*	48**
Dichotic difference	-21	-29*	-23	-31*	-18	-16	-29*	-25*
Clustering base	-52**	-48**	-48**	-61**	-47**	-02	-37**	-34**
Clustering difference	06	10	10	-07	07	00	09	-14
Clustering delay	-12	-22	-21	-28*	-25*	01	-03	-35**
Clustering delay dif	-18	00	06	04	05	06	00	20

[a] Decimal points omitted.

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

Table 2

Intercorrelations among Laboratory Measures, Study 1[a]

	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Motor RT	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. Prop correct			14	25*	12	-29*	-41**	08	06	-20	17	-19	16
4. Stroop difference				27*	21	-32**	16	-11	22	14	-02	-07	-15
5. Search slope					39**	-17	-11	-21	17	15	-44**	-21	05
6. Search intercept						-21	-07	-45**	44**	21	-18	-02	-06
7. Digit span final							65**	-04	-02	14	-19	24	-28*
8. Digit span gain								-12	02	34**	-17	32**	-25*
9. Dichotic category									-75**	-55**	04	-35**	02
10. Dichotic difference										32**	08	26*	02
11. Clustering base											16	54**	09
12. Clustering difference												28*	32**
13. Clustering delay													-32**
14. Clustering delay dif													

[a]Decimal points omitted.

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

Clustering Delay showed some negative correlation whereas Clustering Difference and Clustering Delay Difference were not at all correlated with the psychometric tasks. Remembering that Clustering Delay represents first day recall minus second day recall, high scores are associated with increased reliance on semantic clustering between immediate and delayed recall. Thus, increasing reliance upon clustering was negatively related to performance on the reasoning tests and on Hidden Figures.

Table 2 illustrates how the laboratory measures were correlated among themselves. Response time measures were positively intercorrelated. Accuracy measures, however, were not consistent. Prop Correct, Digit Span Final, and Dichotic Category were not all positively intercorrelated nor were the three consistently negatively related to Clustering Base. Clustering Base and Dichotic Category were strongly related ($r = -.55$) suggesting that both provide measures of a common accuracy of immediate recall. Although the Clustering Base score is a measure of semantic clustering, it is derived from immediate recall of a blocked list. Hence, accurate serial recall of the list would yield a high degree of semantic clustering. Lastly, correlations of Motor RT, Choice Time, Search Slope, and Search Intercept with Dichotic Category and Clustering Base indicate that speed and accuracy of response are related.

Because of the overlap among the laboratory measures, a series of multiple regression analyses was conducted. Each of the psychometric tests was taken as a criterion measure with Motor RT and Choice Time forced as the first two predictors. Additional laboratory predictors were selected so long as they significantly ($p < .05$) increased R^2. This strategy was adopted to evaluate what proportion of test variance could be accounted for by the two simplest reaction time measures and also to determine which, if any, higher order information processing measures contributed as well. While most researchers would agree that the first two measures require the least intellectual activity, it was not possible to arrange the remaining measures in terms of increasing demands upon the information processer. The results of predictor selections are given in Table 3.

Motor RT was significantly related to all but one of the tests—Clerical Number. Choice Time added significant variance to five out of the eight criteria, the exceptions being Verbal Reasoning, Clerical Number, and Clerical Name. Together the variance accounted for by these two simple reaction times ranged from only 8% for Clerical Number to 34% for Space Visualization. For five of the eight measures the first freely chosen predictor was Clustering Base and in four of these five instances it terminated selection. In these four cases, the variance shared with the laboratory measures rested upon speed of response and accuracy of short-term recall. This variance ranged from 25% for Verbal Reasoning to 49% for Numerical Reasoning. The third and only added predictor for Clerical Number and Clerical Name was Stroop Difference. For these two perceptual speed tests, Stroop Difference also had the highest zero-order correlations (Table 1). Stroop Difference thus adds to the time required for a simple comparison (Choice Time), i.e., that required to look up a name. The two most complex tests in terms of the number of laboratory measures providing increments were Verbal Comprehension (vocabulary) and Hidden Figures (field independence). For both, Dichotic Category and Clustering Delay Difference were selected. Overall, the proportion of total test variance accounted for by laboratory measures ranged from 16% to 51%. The upper figure is quite high given the reliability of such short tests.

Three aspects of these results are worth emphasizing. First, Motor RT was significantly correlated with all but one of the psychometric tests. Thus, the time it takes to accomplish the simplest processing task seems to contribute more to intellectual skills than might be expected. Second, Verbal Comprehension had

Table 3

Psychometric Performance Predicted from Laboratory Measures: Increments to the Squared Multiple Correlation for Selected Predictors[a]

Verbal comprehension			Numerical ability			Space visualiza			Numerical reasoning		
Measure	R^2	Δ	Measure	R^2	Δ	Measure	R^2	Δ	Measure	R^2	Δ
Motor RT	16	16	Motor RT	14	14	Motor RT	18	18	Motor RT	16	16
Choice time	22	06	Choice time	24	10	Choice time	34	16	Choice time	28	12
Clus base	35	13	Clus base	34	10	Clus base	42	08	Clus base	49	20
Dichot cat	40	05									
Dichot dif	47	07									
Clus deldif	51	04									

Verbal reasoning			Clerical number			Clerical name			Hidden figures		
Measure	R^2	Δ	Measure	R^2	Δ	Measure	R^2	Δ	Measure	R^2	Δ
Motor RT	07	07	Motor RT	03	03[b]	Motor RT	14	14	Motor RT	07	07
Choice time	11	04[b]	Choice time	08	05[b]	Choice time	19	05[b]	Choice time	17	10
Clus base	25	14	Stroop dif	16	08	Stroop dif	26	07	Dichot cat	28	12
									Search slp	34	05
									Clus deldif	38	04

[a] Decimal points omitted. Motor RT and Choice time forced as first and second selections; additional predictors selected only where R^2 significantly increased ($p < .05$). Δ is size of increment to R^2.

[b] Motor RT or Choice time not significant.

the bulk of its reliable variance determined by these laboratory measures. This occurred despite the fact that none of the laboratory measures assessed the size of memory, e.g., extent of vocabulary. Finally, despite the variety of laboratory time measures, surprisingly little of the variance on Clerical Number and Clerical Name was accounted for. Such perceptual speed tests should have had more in common with response time measures.

STUDY 2

Certain deficits of Study 1 were considered in designing Study 2 which, again, sought to increase basic understanding of the relationships between the intellective factors of psychometrics and the processes of cognitive models. All of the psychometric tests in Study 1 were quite short. This may have contributed to their strong correlations with reaction time by placing premium on speed of answering test items. Secondly, each intellective factor except Perceptual Speed was represented by only a single test. Thirdly, many important cognitive factors were not represented at all, and finally, only single, summary motor RT and choice time scores were available, not permitting study of the differences between an initial and a learned rate of response.

Method

Subjects

First year university students were recruited fall 1974 from those who had completed the battery of tests administered to high school juniors by the Washington Pre-College (WPC) Testing Program (WPC, 1975). 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. Volunteers were paid and their participation extended over an academic quarter. Each student was involved in one psychometric and one laboratory session each week. The total group consisted of 27 men and 36 women.

Psychometric Instruments

Because of the large number of tests involved, all were first intercorrelated and then separate analyses were done by type. The resulting reduced set of measures was finally related to laboratory performance.

Verbal Ability Measures

The WPC battery provided scores for Reading Comprehension, Vocabulary, Spelling, and English Usage (Grammar). (The Verbal Composite score used in subject selection was a linear weighting of these four tests.) Other verbal tests administered following recruitment included EAS Verbal Comprehension from Study 1 and the two subparts of the Terman Concept Mastery Test (Terman, 1956), Synonyms and Antonyms and Verbal Analogies. In the French et al. (1963) scheme, all of these tests would be considered measures of Factor V: Verbal Comprehension.

A principal components analysis of the correlations among these verbal measures followed by a varimax rotation of the larger of these components resulted in a reduction to two measures:

1. *Vocabulary:* Average of standardized WPC Vocabulary and Terman Synonym-Antonym scores.
2. *Grammar:* Average of standardized WPC Spelling and English Usage scores.

Quantitative ability measures. The WPC battery provided three quantitative tests: Quantitative Skills (numeric reasoning), Applied Mathematics (word problems at the elementary algebra level) and Mathematics Achievement (achievement in high school mathematics). (The three were linearly combined to provide the Quantitative Composite measure used in subject selection.) Additional quantitative tests administered in this sample included EAS Numerical Ability and Numerical Reasoning described in Study 1. A principal components analysis was conducted and one component appeared to tap the variability central to all these five measures—Number Facility (N). A separate Reasoning (R) factor was not found and, thus only a single quantitative measure was retained:

3. *Quantitative Composite:* The WPC Quantitative Composite score.

Spatial ability measures. One spatial ability measure was included in the WPC battery, a paper folding test of the kind classified by French et al. (1963) as Visualization (Vz). Two other visualization measures were administered, the Two-Dimensional and Three-Dimensional Spatial Relations tests of the California Multiple Aptitudes Test (CMAT; Segel & Raskin, 1959). The two-dimensional test is a paper formboard and the three-dimensional test is one of surface development.

Other figural tests included Hidden Figures from Study 1 which loads on the Flexibility of Closure (Cf) factor (and, to a lesser degree, on the Spatial Orientation, S, factor), a Maze Tracing task (French *et al.* 1963) associated with a spatial Scanning (Ss) factor, and the Advanced Progressive Matrices (Set II) of Raven (1962). The Progressive Matrices test is variously considered a measure of general intelligence, an indicator of a reasoning factor, Induction (I), or a "space" measure. It was included in this set because, in the present study, it had sizeable correlations with the other figural measures.

The principal components/varimax rotation analyses of these spatial tests defined two clusters and two new composite measures were calculated.

4. *Space I:* Average of standardized Maze Tracing and paper formboard test scores.
5. *Space II:* Average of standardized Progressive Matrices, Hidden Figures, surface development, and paper folding scores.

Reasoning ability tests. All of the tests in this category were administered after subject recruitment. EAS tests of Verbal Reasoning (from Study 1) and Symbolic Reasoning (involves the concepts of equality and inequality presented symbolically) tapped Syllogistic Reasoning (Rs). The Watson-Glaser Critical Thinking Analysis (Watson & Glaser, 1964) provided scores on five subtests: Inference, Recognition of Assumptions, Deduction, Interpretation, and Evaluation of Arguments. With the possible exception of Deduction, these subtests are probably factorially complex, requiring comprehension of reading material and its integration with some already known information as well as the exercise of reasoning.

When the interrelations among all tests were examined, Symbolic Reasoning was more strongly related to quantitative measures, while EAS Verbal Reasoning and Watson-Glaser Recognition of Assumptions, Deduction, and Interpretation were moderately correlated with the verbal measures. In consequence, the two reasoning measures used here were those independent of other factors:

6. *Inference:* After reading a paragraph, a series of statements are classified as true, probably true, false, probably false or based on insufficient data.
7. *Evaluation of Arguments:* A "policy question" is followed by a number of pro and con statements to be classified as weak or strong arguments on the question.

Perceptual/motor speed tests. These tests, again all administered after subject recruitment, placed a premium on quick perception and response. The survey of Working Speed and Accuracy (Ruch, 1943) included Number Checking (comparing pairs of numbers for identity), Code Translation (substituting letters for integers), Finger Dexterity (dotting small circles), and Vowel Counting (obtaining counts for lines of test). Number Checking, Vowel Counting, and possibly Code Translation are measures of the Perceptual Speed (P) factor, whereas Finger Dexterity has been regarded as motor rather than intellective.

A Code Recognition test was also administered from the Flanagan Aptitude Classification (Flanagan, 1953) battery. Following practice in learning alphanumeric codes for common objects, participants were tested immediately on recognition of these codes. The test is a recognition form of the measures (French *et al.*, 1963) used as markers for the Associative Memory (Ma) factor.

Finally, three locally constructed tests were administered, based upon descriptions given by Rose (1974) of a battery developed at the Human Performance Center, University of Michigan. The three were a Grammatical Reasoning test (requiring comparison of a statement "A does not precede B," with a letter pair AB or BA), a Letter Search test (checking blocks of "randomly" assembled letters for a key letter or letters), and a Tapping test in which circular targets of varying size and separation are to be dotted.

Letter Search should be another Perceptual Speed test. Grammatical Reasoning was used as a potential paper and pencil surrogate for a laboratory procedure used in earlier studies (Hunt, Lunneborg & Lewis, 1975) and based on the work of Clark and Chase (1972). The tapping task was intended to provide a paper and pencil measure of a fairly unencumbered response speed. Because these three tests were so simple they were administered under highly speeded conditions on each of two occasions separated by about three weeks. Upon examination, the between sessions correlations for the same test were disappointingly low, while the between tests correlations for the same, final session were all in excess of .90. As a representative of this set of tests, one score was retained:

8. *Letter Search:* Average of standardized scores for the searches for 1, 2, and 4 letters on the final administration.

A principal components analysis of the correlations among the other perceptual/motor scores yielded three dimensions represented in turn by Number Checking, Code Translation, and Vowel Counting. As Vowel Counting was strongly correlated with already selected Letter Search, only the first two measures were retained:

9. *Number Checking*
10. *Code Translation*

Of the ten psychometric measures, six were based on essentially unspeeded performance, while four—Letter Search, Number Checking, Code Translation, and Space I—were more speed dependent.

Laboratory Measures

As with the psychometric measures, time to respond data from several laboratory procedures were first reduced by a series of principal components analyses. The resulting set of measures was then related to the ten psychometric measures described above.

Binary choice reaction time. Participants fixated a midpoint on a CRT display and responded by key press to the onset of a character to the left or right of that midpoint. Each student participated under two conditions. In the *one-finger* condition, the index finger of the preferred hand was used to depress one of two keys. In the *two-finger* condition, keys under left- or right-hand index fingers were used to signal detection of display to left or right respectively. Data were median response times, separately for presentions to the left and right, for the first and last (sixth) blocks of 100 presentations under each of the conditions. Principal components analysis yielded two components and as a result two composite measures were used:

1. *Choice Reaction Time One Finger:* Average of standardized scores for left and right presentations, first and last block.
2. *Choice Reaction Time Two Finger:* Same average based on two-finger condition.

Delayed auditory feedback times. In this procedure, participants read aloud a page length stream of nonsense syllables under three conditions: reading was taped and played back through earphones at zero, 150, and 300 milliseconds delay. Although types of errors actually motivated this task, two time scores were of interest here:

3. *No Delay Reading Time:* Time to read the passage.
4. *Delay Reading Time:* Average of two standardized times to complete reading when feedback was delayed.

During this same experimental session participants repeated to the experimenter a series of words replacing R's with L's and L's with R's. The score obtained was:

5. *R/L Latency:* Median time to transform a word.

Raven response times. Set I of the Advanced Progressive Matrices was presented individually, one problem at a time, and solution time recorded for each figure. This was done after group administration of Set II and involved 12 problems not in the larger set. Principal components analysis of these time scores produced two components based respectively on earlier and later problems in the set. Two scores were thus developed:

6. *Raven Time I:* Average of standardized times to solve problems 1 and 2.
7. *Raven Time II:* Average of standardized times to solve problems 8 and 9.

Verbal problem solving time. Another laboratory task called for learning a list of statements which associated each of a set of personal names either with occupations (Ruth is a lawyer) or locations (John is in the pool), or which associated locations with activities (If you are in the park, you can hear the chimes ring). A week later, following a recall test of mastery, a series of true or false questions was presented. The questions were such as to require the retrieval of 1, 2, 3, or 4 of the learned facts. Time to answer each was recorded and two scores of interest derived:

8. *Verbal Problem Solving Slope:* For the items answered correctly, time to respond was linearly regressed against number of facts required. The problem-solving slope is the slope of this line.
9. *Verbal Problem Solving Intercept:* Intercept of line in 8.

Results

Table 4 reports zero-order correlations between the psychometric and laboratory measures. Choice reaction times were not significantly related to the psychometric measures, not even to the speeded measures. This result contrasts sharply with the results in Table 1 for the high school group of Study 1. Reading time with delayed auditory feedback, time to exchange R's and L's, the intercept of the problem solving time line, and more weakly, solution time for Raven problems 1 and 2 yielded correlations that most closely resembled the expected pattern of negative relations. But, surprisingly, problem solving slope was positively related to certain (unspeeded) tests. Because the results of the two studies differed so greatly, the stepwise regression analyses of Study 1 were not repeated. Rather, principal components and canonical correlation analyses were done to explore the relations between psychometric performance and laboratory measures in this instance. Would even greater combining of measures into components reveal relationships more like those of Study 1?

Table 4

Correlations between Laboratory and Psychometric Measures, Study 2[a]

	Psychometric measures									
Laboratory measures	Vocab	Grammar	Quant comp	Space I(S)	Space II	Inference	Eval argument	Letter search (S)	Number check (S)	Code trans (S)
Choice RT 1 finger	10	-17	-16	-13	05	-09	-21	-06	02	-11
Choice RT 2 finger	-10	-19	-14	01	-03	00	-21	07	11	00
No delay read time	03	-10	14	00	17	-02	04	-08	11	-05
Delay read time	-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 I	-12	-18	-22	-30*	-36**	05	09	12	-10	14
Raven time II	-04	-07	02	-13	-08	01	09	10	06	-05
Verbal prob solv slope	28*	01	23	-36**	03	25*	-05	-05	-07	-02
Verbal prob solv inter	-16	-45**	-11	02	17	-26*	-06	16	-44**	-26*

[a]Decimal points omitted. (S) denotes speeded psychometric scores.

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

Table 5

Loadings of Laboratory Time Measures on Varimax Rotated Principal Components, Study 2[a]

Measure	Component 1	2	3	4
Choice RT 1 finger	71			
Choice RT 2 finger	93			
No delay read time	42			64
Delay read time			83	
R/L latency			68	
Raven time I		89		
Raven time II		81		
Verbal prob solv slope				81
Verbal prob solv inter			67	

[a]Only correlations of .40 or larger indicated. Decimal points omitted.

A principal components analysis and varimax rotation of the nine laboratory measures yielded the four components summarized in Table 5. Component 1 was loaded by time to respond measures when the processing load was minimal. In contrast, component 3 reflects time to respond with more pressure on the system (delay of auditory feedback, overcoming phonetic convention, searching LTM for learned facts). Component 4 associated time to read through an LTM store (Problem Solving Slope) with time to read through a page of nonsense syllables (No Delay Read Time). Time to solve Raven problems appears independent of the time to solve verbal problems, defining a separate, second component. The Raven items differ from the verbal problems in that they are symbolic rather than verbal, become quite difficult, and depend for solution on more than recall of a closed set of recently acquired facts.

Table 6 reports the results of a canonical correlation analysis involving the psychometric and laboratory measures. Two significant correlations were obtained: r = .86 for the first pair of canonical variates and r = .74 for the second pair. The first pair involves large positive weights for the following response time measures; Choice RT Two Finger, Delay Reading Time, Verbal Problem Solving Slope and Intercept. On the psychometric side, positive weights are assigned the Quantitative Composite and unspeeded Space II scores and negative weights to the Grammar and speeded Space I scores. The second pair weights R/L Latency positively and Verbal Problem Solving Slope negatively on the laboratory side, but weighting speeded Space I positively and Vocabulary negatively on the psychometric side.

These canonical variates are not easily interpreted. For both, a negative relationship between Space I and the Verbal Problem Solving Slope contributes. As a result, it is difficult to use this relationship to characterize either variate. Leaving it aside, the first pair of variates could be seen in terms of Grammar and Verbal Problem Solving Intercept as tapping some-

Table 6

Canonical Correlations between Psychometric and Laboratory Measures[a]

Psychometric measures	Variate 1 weights	Variate 2 weights	Laboratory measures	Variate 1 weights	Variate 2 weights
Vocab	-01	-66	Choice RT 1 finger	-34	-06
Grammar	-99	-22	Choice RT 2 finger	56	-03
Quant comp	49	-21	No delay read time	00	11
Space I (S)	-51	58	Delay read time	53	-04
Space II	53	-09	R/L latency	-08	70
Inference	02	-19	Raven time I	02	10
Eval argument	-22	23	Raven time II	23	06
Letter search (S)	33	-08	Verbal prob solv slope	52	-76
Number check (S)	-09	22	Verbal prob solv inter	68	05
Code trans (S)	-07	15			

[a]Decimal points omitted. The first canonical correlation was .86 and the second .74.

thing like speed of rule application. In both instances subjects consult a set of overlearned rules and determine whether they apply in a particular situation. The faster this is accomplished in the problem solving task (low Intercept), the higher the Grammar score. Similarly, ignoring Space I and Verbal Problem Solving Slope, the interpretation of the second variate depends on Vocabulary and R/L Latency. This suggests the second variate measures verbal fluency: slowness in converting R's to L's is associated with poor vocabulary. At best, however, these characterizations are only suggestive of the underlying bridge between intellectual performance and simple response times. Unfortunately, none of the components identified in Table 5 appeared as part of either canonical variate. Thus, whatever linked the laboratory measures with one another did not serve to relate them in turn to paper and pencil test performance.

STUDY 3

The failure in Study 2 of the choice reaction time measures to show stronger relations with the psychometric measures (in particular the failure of these laboratory scores to load together with the more speeded paper and pencil tests in a single variate) was unexpected given the results of Study 1. The differences in findings could have resulted from differences in measurement techniques, from differences in subjects, from both, or from other less obvious differences between the two studies. The most parsimonious follow-up, however, was the administration of Study 2 measures to a group similar to that of Study 1. Would the same strong relations be found if students at the high school level were the sample under investigation?

Method

Subjects

At the end of spring term 1975, high school juniors who had completed the WPC battery were recruited through their counselors. Sixty-four volunteers, 30 females and 34 males, were run in groups of eight for two days. Three-hour morning sessions were devoted to psychometric testing and slightly longer afternoon sessions involved laboratory data collection. As in earlier studies, participants were paid at an hourly student rate.

Measures

Although Study 3 involved other laboratory tasks, results will be reported here only on that subset which is in common with Study 1 and, to a greater extent, Study 2. In particular, scores were obtained on ten psychometric instruments described earlier:

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 all of the same conditions as for Study 2 and, where applicable, Study 1.

The Binary Choice Reaction Time task of Study 2 was repeated in a slightly modified format. Inspection of individual records in Study 2 suggested that stabilization of response was as likely after 300 trials as after 600. In Study 3, therefore, only 300 trials (exclusive of an initial block of 30 warm-up trials) were completed. Only the faster *Choice Reaction Time Two Finger* time was employed here.

Results

Table 7 summarizes the zero-order correlations between the binary choice reaction time

Table 7

Correlations between Laboratory Choice Reaction Times and Psychometric Performances, Studies 1, 2 and 3[a]

	Choice RT one finger			Choice RT two finger			
	High school 1973	College 1975 Block 1	College 1975 Block 6	College 1975 Block 1	College 1975 Block 6	High school 1975 Block 1	High school 1975 Block 3
Verbal comprehension	-46**	10	14	-11	04	14	09
Numerical ability	-48**	-06	-02	-19	04	-04	-19
Numerical 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		-02	-05	06	02	-10	-34**
Three dimensional space	(-55)**	11	09	-05	01	03	-28*

[a]Decimal points omitted.

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

tasks and sets of psychometric scores for Studies 1 (high school 1973) and 2 (college 1975) as well as Study 3 (high school 1975). Data from the first two studies are reported in Table 7 in slightly different form than in previous tables: actual choice reaction times from Study 1 are used rather than "Choice Time" (the difference between choice and motor reaction times) and, for Study 2, choice reaction times are reported separately for the first and last (sixth) blocks of 100 trials.

The significant relations of Study 1 were not replicated for either the one finger or two finger tasks with the college sample. For the second high school group, however, there was a tendency for this negative relation to emerge again for Block 3 data. (In connection with Tables 7, 8, and 9, it should be noted that the spatial visualization test used in Study 1 is different from those tests administered in Studies 2 and 3. Data for the Study 1 test are thus enclosed in parentheses.)

Means and standard deviations for the three groups on these variables are reported in Table 8. In terms of average psychometric performance, the Study 3 high school group was only slightly less proficient than the college group. The Study 1 high school group, however, was markedly less able, e.g., a mean of 15.8 on Numerical Ability as compared to 26.2 for the Study 3 high school students. Choice reaction times, in contrast, were not markedly different among the three groups. Of greater significance than the differences among these means are the differences in the standard deviations. While there was no consistent pattern of differences in the S.D.s on the psychometric instruments, the two high school groups were appreciably more variable on the choice reaction time tasks than the college group.

Variability is known to affect correlations greatly. Could the reduced correlations in Table 7 for the college students have resulted from their smaller RT standard deviations? This possibility can be evaluated by converting the correlations to regression coefficients. While a correlation coefficient indicates the strength of a relation in terms of the proportion of variance in one measure accounted for by a second—and is, hence, sensitive to the amount of variance present—a regression coefficient measures the same relation in terms of the amount of change in one measure that would be expected to accompany a fixed amount of change in a second.

Will the relations between Choice RT and psychometric performance appear more consistent if regression coefficients replace correlations? The regression coefficients will have the same signs as the corresponding correlations and any inconsistencies in direction of relationship will not be corrected. Thus, the positive correlations in Table 7 will remain unresolved. (None of these positive correlations, however, were significantly different from zero.) Regression coefficients were computed for the negative correlations in Table 7.

To make the regression coefficients of Table 9 easier to understand, the coefficients have been scaled: (1) Negative signs have been changed to positive so that the entries now express amount by which test scores would be expected to *decrease* with increases in Choice RT. (Remember as reaction times got larger (slower), test scores decreased in Study 1.) (2) Because the unit of measurement for Choice RT was so small (a millisecond), each regression coefficient has also been multiplied by ten. The result of the two scaling changes is that the coefficients in Table 9 are interpreted as the amount by which a test score may be expected to decrease as a result of increasing the Choice RT by one hundredth of a second (10 ms.).

The reanalysis presented in Table 9 only partly clarifies the relationship of Choice RT to paper and pencil test performance. Wholly consistent results would have been represented by a constant value for any given row (test). What resulted instead can be summarized as follows: (1) The strength of relationship between Choice RT and test performance was still higher for high

Table 8

Means and Standard Deviations for Test Scores and Choice RT, Studies 1, 2, and 3

Measures	Study 1 High school 1973		Study 2 College 1975		Study 3 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
Numerical ability	15.8	7.8	27.5	8.3	26.2	8.0
Numerical 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			21.4	7.7	26.4	7.0
Hidden figures	3.4	3.1	11.2	6.3	10.2	6.9
Two dimensional space			16.3	5.0	18.6	4.1
Three dimensional space	(19.2)	(9.8)	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

Table 9

Predicted Decrements (Regression Coefficients) in Test Score Associated with Ten Millisecond Increase in Choice Reaction Time[a]

	Choice RT one finger			Choice RT two finger			
	High school 1973	College 1975		College 1975		High school 1975	
		Block 1	Block 6	Block 1	Block 6	Block 1	Block 3
Verbal comprehension	.33	--	--	.17	--	--	--
Numerical ability	.45	.11	.04	.58	--	.04	.35
Numerical reasoning	.23	.05	.05	.20	--	.06	.08
Verbal reasoning	.23	.03	.25	--	--	--	--
Symbolic reasoning	x	.11	.29	.19	--	.08	.50
Progressive matrices	x	.05	--	.20	.03	--	.40
Maze tracing	x	.21	.39	.09	--	.28	.46
Hidden figures	.14	.04	--	.25	.29	.13	.47
Two dimensional space	x	.02	.07	--	--	.06	.33
Three dimensional space	(.64)	--	--	.08	--	--	.33

[a]x indicates no data available and -- indicates relationship reversed from expectation.

school students than for college students even after adjusting in this way for differences in variability. (2) The two high school studies provided different results. The EAS tests which were strongly correlated with Choice RT in Study 1 were not the tests which were strongly correlated with Choice RT in Study 3. Spatial tests had the bulk of the strong correlations in Study 3. (3) Which Choice RT to relate to psychometric performance remains in question. For the Study 3 group it was performance in the third block of 100 trials, while for Study 2, using the same Choice RT Two Finger task, initial performance (first block of 100 trials) was more consistent.

Discussion

The implication of these three studies would seem to be that there is no simple answer obtainable to the question of how Choice RT relates to paper and pencil test performance. The patterns of correlations between these two types of intellective measures were different among the three samples. The college sample showed the weakest relationships, which may be attributed to their lower variability on the Choice RT tasks. This group would appear to be so selective as to be a questionable population in which to explore such relationships. The differences in correlations between the high school groups, on the other hand, cannot be explained on this basis. They shared the same range of Choice RT performance. They did differ, however, in absolute level of intellective ability—the Study 3 group being much more capable than the Study 1 group. Curiously enough, while the strong correlations of Study 1 were with verbal tests, the strong correlations of Study 3 were with spatial tests.

Some part of the observed differences in the patterns of correlations are also likely to be attributed to changes in the technique of measuring Choice RT. Use of one- or two-handed tasks and basing the measure on earlier or later blocks of trials produced different results. These two aspects must be regarded as part of the definition of the Choice RT to be studied. Equally important to this, the selection of a Choice RT measurement technique can also be expected to influence the reliability (within subjects consistency) of those observations.

How should one investigate the role of Choice RT in ability measurement in light of these results? First, great care must be given to the definition of the subject population. Characteristics desired in such a population are a level and range of performance on both sides, cognitive and psychometric, which are no more restrictive than what would be found in sophomore or junior high school students. Larger samples than were used here will necessarily be required to meet this goal.

Second, equally great care must be given to the definition of the Choice RT task. Here, this should be directed towards obtaining the most reliable assessment. This quality seems more desirable than the theoretical constraints of being "initial" or "practiced." The typical subject in the laboratory performs erratically at first, then stabilizes, and finally grows erratic again as boredom and fatigue overpower the intrinsic motivation provided by the task. This probably takes place at different rates for different people. Given this situation, a tailored testing approach to Choice RT measurement seems worthy of exploration, i.e., determining for each person a Choice RT for that most stable sequence of trials wherever it occurs in the longer experimental paradigm. That median RT for left- and right-hand responses during the terminal, sixth block of 100 trials correlated only .60 in Study 2 provides testimony on this point.

Despite the inconclusive results of studies such as these, the growth of cognitive psychology coupled with disillusionment with the apparent shortcomings of classical psychometric assessment is certain to result in continuing attempts to develop measures of individual differences in intellective ability based on cognitive theory and laboratory work.

These attempts will be more effective if the two prime dicta of individual differences in psychology, just exemplified, are kept in mind: (1) defining relevant normative populations and (2) developing reliable assessments.

References

Andrew, D. M., & Paterson, D. G. *Manual for the Minnesota vocational test for clerical workers.* New York: The Psychological Corporation, 1946.

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.

Flanagan, J. C. *Flanagan aptitude classification test.* Chicago: Science Research Associates, 1953.

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.

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.

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.

Raven, J. C. *Advanced progressive matrices* (1962). London: H. K. Lewis, 1962.

Rose, A. M. *Human information processing: An assessment and research battery.* Technical Report 46. Ann Arbor: Human Performance Center, Department of Psychology, University of Michigan, 1974.

Ruch, F. L. *Manual for survey of working speed and accuracy.* Los Angeles: California Test Bureau, 1943.

Ruch, F. L., & Ruch, W. W. *Employee aptitude survey technical report.* Los Angeles: Psychological Services, Inc., 1963.

Segel, D., & Raskin, E. *Multiple aptitude tests manual.* Los Angeles: California Test Bureau, 1959.

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

Terman, L. M. *Concept mastery test manual.* New York: The Psychological Corporation, 1956.

Washington Pre-College Testing Program *Learn more about yourself—explore your future: Guide to the Washington pre-college test.* Seattle: Washington Pre-College Testing Program, 1975.

Watson, G., & Glaser, E. M. *Manual for the Watson-Glaser critical thinking appraisal.* New York: Harcourt-Brace & World, 1964.

Acknowledgements

This research was sponsored by the National Institute of Mental Health (Grant MH21795, Earl B. Hunt, Principal Investigator) and by the National Institute of Education (Grant NIE-G-74-0104, Clifford E. Lunneborg, Principal Investigator). Portions of this report were presented to the annual conventions of the Americal Psychological Association in 1974 and 1975. The assistance of Kenneth Burkhardt, Jon Palmer and Steven Poltrock in the design and conduct of the laboratory studies is gratefully acknowledged.

Author's Address

Clifford E. Lunneborg, Department of Psychology NI-25, University of Washington, Seattle, WA 98195